USER PREFERENCE AND EMBODIED CONVERSATIONAL AGENT GROUPS

NEIL BARRETT

User Preference and Embodied Conversational Agent Groups

by

A thesis submitted to the
School of Graduate Studies
in partial fulfilment of the
requirements for the degree of
Master of Science

Department of Computer Science
Memorial University of Newfoundland

May 2007

St. John's Newfoundland

Abstract

The interface used to interact with the everyday computer is evolving, possibly towards embodied conversational agents (ECA/agent). ECAs are virtual characters with a body and persona that typically communicate with users via natural language speech. ECAs are successful in many ways and multi-ECA environments exist; yet there is a lack of results, especially pragmatic results, pertaining to a user's experience in multi-ECA environments.

An experiment was realized, using a general framework built for simulating ECA groups, to understand user response to such groups. In particular, this experiment studies the difference in user response to a second additional agent in a single agent "group".

WOZECA, a framework for simulating ECA groups, was built to accommodate ECA characteristics researched to date and facilitate ECA experiments. The framework presents several ECA group interface characteristics using a behind-the-scenes operator and configurable video clips. This flexibility permits multiple different experiments such as the one presented in this thesis.

The experiment manipulates the *number of embodied conversational agents* (levels: 1 and 2 ECAs) and *agent behaviour* (levels: neutral, positive) in a two factor, two level design. The experiment manipulated ECA behaviour. The literature suggests users will respond socially and naturally to the interface. This prompted the use of the following constructs to quantify user response: group cohesiveness, the user's emotional state (valence & arousal) and the group's impact on user self-efficacy.

The results suggest that a second additional agent influences user response. Specif-

ically, ECA behaviour has more influence on group cohesion and valence than does the number of agents. Furthermore, group cohesion depends on the user's personal style, age and the ECA group's behaviour. WOZECA successfully simulated ECAs during the experiment and is capable of a variety of future experiments.

Acknowledgments

It is a terrible thing, this kindness that human beings do not lose. Terrible, because when we are finally naked in the dark and cold, it is all we have. We who are so rich, so full of strength, we end up with that small change. We have nothing else to give.
– Ursula K. Le Guin, The Left Hand of Darkness

So long and thanks for all the fish
– Douglas Adams

Partial funding provided by Dr. Edward Brown through the National Science and Engineering Research Council of Canada (Discovery Grant no. 138209-03).

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The everyday computer interface has evolved. It has progressed from the command line to the graphic user interface (GUI), becoming richer and more complex at each step. Computers communicate in ways that are more familiar, more human, than before. One possible destination for this evolution of computer communication is the embodied conversational agent.

According to Wooldridge [92, page 15], a computer agent is an autonomous computer system in some environment that is capable of taking actions in the environment to further its design objectives. For example, satellite software that acts autonomously to diagnose and repair itself is an agent. A very different example of an agent is an on-line web-application that autonomously decides which items are of interest to the shopper. Wooldridge's definition of agent is broad, encompassing many dissimilar computer systems. This thesis focuses on a particular type of agent called embodied conversational agents.

Agents comprised of a physical appearance and a persona that maintain a dialogue

with human individuals are embodied conversational agents (ECA/agent). For example, the real estate agent named Rea appears to the user on a large projection screen [19] and maintains a dialog with the user while selling virtual real estate; whereas MACK answers questions at an information kiosk [80]. MACK appears to the user of the kiosk as a life-sized blue robot able to converse using natural (English) language and to interact with the kiosk user via a paper map. Embodied conversational agents are stylized virtual people.

Researchers are building the pieces that will make embodied conversational agents a reality. Some researchers work towards visual or gestural realism, others investigate the user-agent relationship and its interaction. The use of embodied conversational agents as computer system interfaces is motivated by a human desire for intuitive and social relationships built on the strength of cognitive accessibility and natural communication style [41, page 356]. Embodied conversational agents promise to provide a level of interaction familiar to the human individual. As Laurel points out:

> "*...technologies offer new opportunities for creative, interactive experiences and, in particular, for new forms of drama. But these new opportunities will come to pass only if control of the technology is taken from the technologist and given to those who understand human beings, human interaction, communication, pleasure, and pain.*" [42]

1.1 ECA Groups

An area with many questions and few answers relates to how users respond and interact to groups of embodied conversational agents. Social groups such as academic groups, families, and sports teams are prevalent within society. These groups are a staple of human existence. One would expect ECA groups to benefit users in a

similar manner that traditional human-only groups benefit members. For instance, interacting with a pedagogical ECA group has been shown to have benefits over interacting with a single pedagogical agent (e.g. [53, 33] Section 2.1). As Rieken, Maes, Shneiderman and Smith remarked:

> "*the human perspective is a crucial one in understanding the potential of [user interface] software systems which are both usable and useful for user tasks.*" [73]

In other words, the success of ECA groups as an interface requires an understanding of user response to groups.

1.2 Research Focus

User response refers to the act of responding and replying to the ECA interfaces multi modal output as well as the entire interaction experience (e.g. emotional, analytical or spontaneous response).

Little is known about user response to embodied conversational agent groups (Chapter 2). It is unclear how users will respond to the addition of a second agent to a single agent group. For example, a second agent may only be effective under specific conditions; ECA groups (with two ECAs or more) may only be desirable for certain tasks; and certain ECA attributes may be crucial in affecting user response.

To date, it appears that users respond socially and naturally to media, including the (single) ECA interface. Since previous research focuses on the single ECA interface, it does not clearly indicate which novel aspects of the ECA group interface affect user response and what group aspects are desirable. With multiple agents in

the ECA interface, user response to agent-agent interaction can be examined. Unlike previous research, this thesis examines user response to agent-agent interaction within an ECA group and attempts to identify at least some aspects of a two-agent ECA group interface that are desirable. More specifically, this thesis presents an experiment, implemented on original multi-agent simulation software, that compares user response to a single ECA with user response to a two-agent ECA group, by examining specific psychological constructs that associated with positive group experiences. Thus, the research gap of interest is the question of how users respond to the interaction between agents in a two-agent ECA group - whether their behaviour (i.e. agent-agent interaction) is more important than the group's additional agent.

1.3 Thesis Overview

To understand how users respond to an ECA group, a framework for simulating ECA groups was built (Chapter 3) and an experiment was realized using this general framework (Chapter 4).

Drawing mostly on the *one-user-one-agent* literature, the review in Section 2.2 suggests that characteristics such as gestures and learning are important to ECA group interfaces. A framework named WOZECA (Chapter 3) was built to accommodate these characteristics, to simulate ECA groups, and to facilitate ECA experiments. WOZECA's flexibility permits a large variety of experiments such as the experiment presented in this thesis. WOZECA presents characteristics of ECA group interfaces using a behind-the-scenes operator and configurable video clips. The suggested characteristics provide a means for evaluating WOZECA as well as its

requirements.

The experiment (Chapter 4) investigates the difference in user response in ECA groups due to the addition of a second agent to a single agent group. Since the literature (Section 2.3) indicates user response is affected by ECA behaviour more than their appearance, the experiment manipulated the ECA group's behaviour as well. The experiment used two groups with two behaviours and measured user response to these four scenarios. The concepts, anthropomorphism (Section 2.3.1.1) and the Media Equation (Section 2.3.1.2), suggest that users will respond socially and naturally to the ECA interface. Theses prompted the use of constructs from psychology in order to assess user response. The constructs chosen include group cohesiveness, the user's emotional state (as valence & arousal) and user self-efficacy.

The experiment results (Chapter 4, Section 4.5) suggest that the addition of a second agent to a one agent group influences user response. ECA behaviour was found to influence group cohesion and valence more than the number of agents, whereas arousal and self-efficacy changed over time. The results also suggest that the user's rating of group cohesion depends on their age and personal style as well as the ECA group's behaviour.

The thesis concludes (Chapter 5) by suggesting improvements to WOZECA, highlighting the experiment's main results and proposing future work. Improving WOZECA's audio and video devices would improve the behind-the-scene operator's ability to respond to user statements. Two possible future directions are investigating ECA groups where the ECA members have conflicting views and building a better understanding of ECA meta-conversational requirements.

Chapter 2

Background

This chapter presents three different perspectives on embodied conversational agents (ECA) in interfaces and user response to these interfaces. The first perspective (Section 2.1) discusses the literature on ECA groups. It demonstrates that users can respond differently to a single agent and a multi-agent group, but provides little advice about this difference. The second perspective (Section 2.2) examines the available, mostly one-user-one-agent, ECAs and proposes several characteristics likely to cause a user response. These characteristics are derived from important ECA features as reported by researchers. The final perspective (Section 2.3) discusses user response to interfaces and ECAs. It proposes that users respond socially to the interface and that an ECA's behaviour is more important than its appearance. Section 2.4 summarizes this chapter.

2.1 ECA Groups in User Interfaces

Few researchers report on ECA groups in interfaces. However, the available literature suggests users can respond differently to a single ECA and a multi-ECA group.

Reporting on early (1994) ECA research, Cassell et al. discuss a two ECA conversational scenario [21] where two agents hold a multi modal conversation. The paper concludes that it is possible to automatically generate multi modal information to animate interactive dialogue. Unfortunately, user response to such an agent-agent conversation and relationship are not within the scope of Cassell et al. 's paper.

Elva [95, 94] and Gamble [71] solicit a positive user response when many users simultaneously interact with one agent. Elva, built for autonomy and believability, responds intelligently and contextually to interactions. Gamble, an accepted and competent game player, is incapable of understanding off-topic discussions by the human players. The papers do not discuss user response in depth, nor do they address one user to many agent relationships.

The Steve framework is pedagogical in nature, simulating military situations for training purposes [87, 82]. Rickel et al. conclude that virtual humans are limited by their spoken dialogue and domain task models; and that human-level intellect demands abilities such as planning, belief representation, communication, emotional reasoning and, most predominantly, integration of the latter. Although the framework demonstrates the one user to many agent relationship and Rickel et al. discuss current ECA limitations, the research does not report on user response to multiple ECA soldiers.

Morishima et al. investigate a single user interacting with multiple ECAs from a

pedagogical perspective [53]. A user interacts with two ECAs within a collaborative learning system where one ECA fulfills the traditional role of pedagogical agent and a secondary ECA, called a co-learner, participates by learning alongside the user. The results show that the co-learner agent adds to the effectiveness of the system and that the agent increases the positive impression of the application, a result supported by Ju et al. [33].

2.1.1 ECA Groups Summary

Pedagogical results show multiple agents help a user learn more effectively and suggest user response to multi-agent groups differs from a single agent. Use of multi-agent groups may improve ECA interfaces when users respond favourably to these multi-agent interfaces.

2.2 Characteristics of ECAs

To gain a greater understanding of user response to ECA groups, this thesis moves from the multi-agent literature to the more extensive one-user-one-agent literature. The one user one agent ECAs often exist in some form whether robotic or virtual, operate within an environment such a hallway or laboratory, and typically claim some form of autonomy to their being. This perspective examines the ECA's characteristics expected to affect user response. The underlying assumption is: if researchers focus on particular ECA characteristics, such as hand gestures, then these researchers attribute importance to said characteristics. Since a comprehensive model of ECA characteristics is nonexistent, the following original categories (suggested by the au-

thor) of characteristics are offered to assist in reviewing these ECA systems.

- *Physical awareness* is an agent's ability to track or manipulate physical objects in the user's space and hold communications with respect to these objects.
- *Communicational awareness* is an agent's ability to understand varying methods of conversation (speech or text) and conversation subtleties.
- *Learning* is characterized by the use of multi modal interaction with human users as ECA learning input. That is, the user trains the ECA via speech and gestures.
- *Gesture coordination and understanding* is an ECA's ability to communicate to their user's without speech or text.
- An ECA accounting for (e.g. changing behaviour) and updating information representing the user's emotional state is considered to be modeling *emotional state.*
- An ECA's *persona* consists of their appearance (visual manifestion), role (purpose), back-story, behaviour patterns and emotions. The ECA persona is one of the least developed aspects of the ECA and, in many cases, their persona is restricted to their function.

For each embodied conversational agent surveyed in the remainder of this subsection, characteristics addressed in the associated research will be indicated in italics in the text describing that agent.

2.2.1 ARMAR II

Collaborative Research Center 588 located in Karlsruhe (Germany) focuses on teaching cooperative multi modal robots adaptation to temporal commonplace environments and creating humanoid robot systems capable of interacting with humans in a human-centered environment [27]. ARMAR II, a rather sophisticated human-like robot, localizes and tracks users both visually and audio-visually (*physical awareness*); once localized ARMAR communicates with users in a natural fashion using speech (*communicational awareness*) and gestures (*gesture coordination & understanding*).

2.2.2 AutoTutor

Graesser et al. of the University of Menphis, in collaboration with the Massachusetts Institute of Technology (MIT), developed AutoTutor (AutoTutor 3D) a pedagogical (*persona*) embodied conversational agent [30, 1]. AutoTutor, located on the left half of the computer monitor, helps a student in subjects such as physics using natural discourse (*communicational awareness*) and an interactive three dimensional simulation of the physical environment. Graesser et al. wish to extend AutoTutor's capabilities by incorporating a model of the student's emotional state (*emotional state*) derived from dialogue patterns (*learning*) that occur while tutoring. This would result, for example, in AutoTutor leading a frustrated student on a positive learning trajectory with quality hints.

2.2.3 Elva

Elva, an embodied tour guide (*persona*) created at the National University of Singapore by Yuan, Chee et al., provides tours of a virtual gallery [95, 94]. Unlike most embodied conversational agents, Elva interacts with a virtualized human group rather than a single individual. Gallery visitors communicate with her by typing text into a chatter box (text field) and she replies with text messages (*communicational awareness*). This conversational interaction is similar to Internet Relay Chat (IRC). Visually, Elva is a full-bodied on screen female presence (*persona*) sharing the visitor's virtual space. The research focuses on autonomy and believability using a four layer software design to handle group interaction. Underlying Elva is a knowledge base allowing Elva intelligent and contextually relevant responses to interactions (*communicational awareness*).

2.2.4 Gamble

The Gamble project, by Rehm et al. located at the University of Augsburg (Germany), studies multi party interactions where one member of the group is a Greta ECA [71]. A prior study saw the Gamble agent as an accepted and competent game player (*persona*). However, the social interactions that occurred, such as off-topic discussions by the human players, were beyond Gamble's capabilities. Rehm et al. 's future work will focus on Gamble perceiving off-topic discussions and comments referring to the agent (*communicational awareness*). Gamble will also integrate an assessment of the user's emotional state during game play (*emotional state*).

2.2.5 Gandalf and Mirage

While at MIT, Kristinn et al. realized Gandalf, an agent capable of perceiving multi modal acts and responding in a comparable manner [85] (*gesture coordination & understanding*). Gandalf draws upon its knowledge of the solar system to explain and converse with those interested (*communicational awareness*). Dialogue between Gandalf and the user is not limited to speech but, may include, gestures such as pointing. In other words, Gandalf responds with speech, gaze, facial and manual gestures, and head movement.

The ECA Mirage, also created by Kristinn and students at Comlumbia University, improves over Gandalf by occupying the same physical space as the individual with whom Mirage interacts [86] (*physical awareness*). The space sharing is accomplished with special glasses (Sony LDI-D100B) that superimpose Mirage on the user's visual world. Furthermore, Mirage appears to move or remain stationary independently of the user via orientation and position tracking of the user's hand and head, providing a unique interaction experience compared to other embodied conversational agents.

2.2.6 Greta

Greta is an embodied conversational agent conceived by Pelachaud et al. with preliminary work in Rome and, subsequently, at the University of Paris 8 [67]. Using a three dimensional virtual face, Greta visually expresses several human emotions (*emotional state*).

A more recent Greta [68] is endowed with a body able to express emotions in a similar manner as her face. A six dimensional expressivity model coordinates the

behaviours resulting in temporal emotional expressions such as surprise. According to Pelachaud et al. user perception of the Greta's expressivity model is relatively positive - the model successfully expresses emotion some of the time - yet, more work is required to perfect that which is already in place.

2.2.7 iCat

Breemen et al. , working at Philips Research, devised a robotic research platform called iCat[1] [17]. iCat is a 38 centimeter tall immobile yellow cat-like robot[2]; its appearance is crafted to expedite a social relationship (*persona*) between it and its user. People are motivated to interact with iCat because the robot performs personalized domestic functions such as altering a room's ambiance (music and lighting) (*persona*).

2.2.8 Laura

Laura is a component of the FitTrack program, conceived by Bickmore (Northeastern University) and Picard (MIT Media Laboratory), that targets or motivates exercise adoption (*persona*) and, in particular, permits study of relational behaviour in long term human-agent relationships [8, 10]. Participants desiring health behavior change interact with Laura (approximately ten minutes daily) via inputs consisting of "text-phrase" menus and a range of outputs such as synthesized voice, hand gestures, walking on and off screen, four different facial expressions, wide to close-up camera shots

[1]An interesting aspect of iCat is that Philips Research has chosen to make the iCat platform available to universities and research laboratories. The platform contains a software component called Open Platform for Personal Robotics (OPPR) which enables the programmer to probe the system during runtime as well as add and remove modules. Furthermore, graphical tools exist facilitating creation of animations for the robot.

[2]The iCat robot resembles a child's plastic toy; one might suspect that the look of future version will need to match the owner's personality a little more.

and several idle-time behaviors (*communicational awareness* and *gesture coordination & understanding*).

2.2.9 Leo

Breazeal at al. research socially intelligent robots [15] and embodied conversational agents in the form of Leo [16]. Leo is not fully conversational for he can not speak, however communication between Leo and his user (teacher) is bidirectional. Leo uses multi modal output such as gestures and facial expressions to communicate his desires whereas users may speak and gesture such as pointing (*communicational awareness* and *gesture coordination & understanding*). The research objective is to render robots more efficient, enjoyable and intuitive during human-robot interaction with a special focus on human-robot interaction where humans act as teachers (*learning*).

2.2.10 MACK

MACK is an embodied conversational agent who answers questions at an information kiosk (*persona*) situated in MIT's media lab [80]. MACK appears to the user on a screen as a life-sized blue robot who understands (English) speech, senses the user with a pressure-sensing chair mat (*physical awareness*) and perceives user actions that occur on a paper map placed a top a Wacom tablet. With MACK's use of speech, his ability to highlight areas of the map using an LCD projector and his synchronized head, eye and arm movement, questions posed by a user such as "*Tell me about this*" (user pointing to a specific research group on the map) are answered (*communicational awareness* and *gesture coordination & understanding*).

An enhanced MACK recently tested a model of Face-to-Face grounding - a mechanism by which conversational participants, verbally or non-verbally, confirm an understanding (*gesture coordination & understanding*). When a question is asked of MACK, he responds and awaits grounding behaviour such as a verbal "*OK*" or a non-verbal blank-stare. Nakano et al. [57] conclude that a more sophisticated grounding strategy than the one tested is warranted to distinguish sounds that may or may not be speech or varying levels of miscommunication.

2.2.11 Mel

Sidner et al. research using a robotic penguin, Mel, that opens and closes his beak, flaps his wings, didactically points at objects with a laser light attached to his beak and outputs speech[3] [77] (*communicational awareness* and *gesture coordination & understanding*). Enhanced with a camera, Mel employs contextual dialogue information to predict and discern head gestures such as a horizontal head shake [52]. Lexical features derived from spoken words, punctuation features such as a comma, timing features such as the end of a sentence, and gesture features such as pointing may influence the listener's feedback. Sidner et al. conclude that context derived from features improves Mel's gesture recognition.

2.2.12 NUMACK

At Northwestern University, NUMACK (similar to MACK Section 2.2.10), an embodied conversational agent, answers questions related to locations and buildings on

[3]Speech is generated and recognized with IBM's (java based) ViaVoice JSAPI.

campus [38] (*persona*). Kopp et al. discuss the integration of arm and hand motion to convey spatial and visual information related to an object or event relevant to current dialogue. NUMACK can associate image features with their discrete gestural forms (iconic gestures) using an integrated and real-time micro-planning software process (*gesture coordination & understanding*).

2.2.13 Rea

Rea, an ECA that sells real estate (*persona*) in the Boston area, is likely the most well-known embodied conversational agent. Rea, developed by Cassell et al. [19, 9], appears life-sized on a large projection screen from which she is capable of maintaining a multi modal interaction with the prospective real estate buyer - the user. To support a natural interaction, Rea employs two cameras that track head and hand position as well as movement (*physical awareness*), and a microphone for audio input. Rea's multi modal responses, managed by several computers, involve speech (with intonation), facial expressions and hand gestures (*communicational awareness* and *gesture coordination & understanding*). In order to augment a user's trust (in Rea), she performs small talk - casual or trivial conversation - and task talk - purposeful conversation. Rea's discourse planner can interleave small talk and task talk smoothly transitioning between both modes of discourse (*communicational awareness*).

2.2.14 Steve

Rickel et al. created the pedagogical ECA system Steve, located at the University of Southern California Information Sciences Institute [87, 82]. Multiple Steve agents

are projected on an eight foot tall screen that encompasses the viewer (150 degree arc 12 foot radius) and a comparable audio experience originates from ten speakers and two sub-woofers providing spatialized sound. The totality is an immerse virtual experience, called the Mission Rehearsal Exercise (MRE) system, designed to teach leadership skills in high risk social scenarios (*persona*) and exhibit the implementing technologies (ECA/virtual human).

In order to construct a successful three dimensional virtual world, the Steve agents require planning algorithm for movement and displacement (*physical awareness*), sensitivity to human gaze, the use of gaze for referencing objects and regulating turn-taking in team dialogue (*communicational awareness*). Furthermore, the MRE system includes natural language comprehension and understanding, speech recognition, computation of emotional states (*emotional state*) and dialogue.

2.2.15 Valerie

At Carnegie Mellon, Valerie, a roboreceptionist and embodied conversational agent, is a permanent installation in Newell-Simon Hall as part of the Social Robots Project (SRP) [28]. The goal of the SRP is to study long-term human-robot social interactions. Valerie's interactions are multi modal; gathering information via a keyboard, speech recognition software (*communicational awareness*), a range finder (for tracking users) (*physical awareness*) and a card reader. The card reader allows Valerie to remember relevant user information over multiple separate interactions. To express herself, Valerie speaks (text-to-speech) and produces facial expressions (*gesture coordination & understanding*) on a flat screen positioned on top of her body (B21r

mobile robot created by iRobot).

In collaboration with the Carnegie Mellon School of Drama, an elaborate back-story was imagined including a singing career and a love life (*persona*). Her life-story evolves as time progresses in a similar fashion to reality TV and soap operas. These new pieces of Valerie's life, conveyed through monologues styled as telephone conversations, exist to keep users interested.

The Social Robots Project has seen Valerie replaced by Tank [3] in an effort to improve and better understand social human-robot interaction.

2.2.16 ECA Characteristics Summary

Embodied conversational agents are truly crippled by their limited *persona* (see summary table 2.1). The most diverse aspect of ECA persona is appearance. Very few agents have back-stories, behaviour patterns (personality) and emotions. Perhaps this shortcoming relinquishes ECAs to pedagogical or subservient roles (information dispensers). There appears to be more interest in *physical awareness*, *learning*, and modeling the user's *emotional state*. For example, physical awareness, for many ECAs, involves tracking the user but little beyond this action.

On the other hand, many ECAs converse using "direct statement based" text or speech (*communication awareness*) and others concentrate on communication using gestures (*gesture coordination and understanding*). Gesture communication is typically restricted to arm movement (or something equivalent such as a laser pointer) and facial expressions, although several ECAs understand speech with matching simplistic gestures. These ECAs have difficulty with (complex) conversation and conversation

Table 2.1: Summary of ECA Characteristics. Physical Awareness (PAW), Communicational Awareness (CAW), Learning (LEA), Gesture Coordination & Understanding (GCU), Emotional State (EMO), Persona (PER) (* indicates role/appearance aspect of persona only, for definitions of these terms see Section 2.2).

ECA	PAW	CAW	LEA	GCU	EMO	PER
ARMAR II	√	√		√		
AutoTutor		√	√		√	*
Elva		√				*
Gamble		√			√	*
Gandalf/Mirage	√	√		√		
Greta					√	
iCat						*
Laura		√		√		*
Leo		√	√	√		
MACK/NUMACK	√	√		√		*
Mel		√		√		
Rea	√	√		√		*
Steve	√	√			√	
Valerie	√	√		√		√

that differs from gesture. For example, a user speaking "right" yet pointing "left", or communication not directed at the ECA but making reference to it such as "*I think the ECA, Ace, is unintelligent*". Coordinating both forms of communication into one that is seamless and comprehensive is an ongoing challenge.

Given the above, these characteristics (physical awareness, communicational awareness, learning, gesture coordination and understanding, modeling the user's emotional state and persona) are expected to affect user response to ECAs and ECA groups. Hence, they are considered in the design, creation and evaluation of the ECA group simulation framework described in Chapter 3.

2.3 User Response to ECAs and the Interface

This perspective extends the understanding of user response developed in the previous perspective (Section 2.2) by concentrating on the literature that directly addresses user response to interfaces and ECAs. Knowledge gained in this perspective is applied to the ECA group simulation framework as well as the design of the experiment.

This perspective begins by discussing two prominant concepts in the literature: anthropomorphism (Section 2.3.1.1) and the *Media Equation* (2.3.1.2). Both of which suggest that users will respond socially to the ECA interface. In other words, users are expected to respond socially to ECA appearance and behaviour.

Subsequently, the thesis examines user response to appearance (Section 2.3.2) and behaviour (Section 2.3.3). These sections expose the user's need for consistency in appearance and behaviour as well as suggesting behaviour's effect on user response is stronger than appearance.

2.3.1 User Response is Social and Natural

2.3.1.1 Anthropomorphism

Anthropomorphism is a behaviour in which humans attribute human-like features to non-human entities such as chairs, animals or computers [25]. For example, a car experiencing mechanical or electrical failure is referred to as dying or dead even though the car is never alive. In the domain of human-computer interaction (HCI[4]) beginners anthropomorphize more than experts [25]. A beginner will describe a computer application's behaviour as uncooperative when the application does not perform as expected whereas an expert is more likely to attribute the application's performance (or usability) to poor programming and design. A novice, given any new situation, lacks a conceptual understanding of the mechanics underlying the new situation and a precise vocabulary to express these mechanics. Novice users employ common-place social metaphors in order to communicate their situation to their peers. Anthorpomorphism considers human-computer interfaces as social in nature.

2.3.1.2 The Media Equation

According to Reeves and Nass the "*interactions with computers, television, and new media are fundamentally social and natural*" [70]. This tendency for media to "equal" real life is Reeves and Nass' *Media Equation*. Although the Media Equation is similar to anthropomorphism, they are distinct and distinguishable concepts.

Users of varying abilities, from novices to experts, routinely deny their application of social rules to computer interaction [70]. Anthropomorphic behaviour fades

[4]The study of design, evaluation and implementation of computing system interfaces.

as users acquire the vocabulary and concepts required to communicate their situation whereas the Media Equation survives a user's interaction with media [70]. Even simplistic (poor, crude) technology can provoke and manipulate a user's social behaviour [61, 60]. Consider the example where Lee et al. created social presence with a computer voice [45]. A users' feeling of social presence, "*the sense that other intelligent beings co-exist and interact with you*" [45], was affected by computer voices including synthesized voices. In particular, users preferred written text or speech, either extrovert or introvert, matching their own personalities. The Media Equation casts human-computer interaction as similar to human-human, and group, interaction.

2.3.2 User Response to Appearance

Common wisdom warns against deriving a conclusion from an appearance but inevitably people do. User response to embodied conversational agents is no different; users make superficial judgments [93].

In some cases, users have identified with the physical components of the computer rather than the on screen windows [83]. Likewise, the physical or virtual elements constituting an embodied conversational agent depend on circumstance. Users may create the head and body dichotomy using the components that appear most suitable. For example, users are likely to perceive Valerie's [28] robotic parts as her body whereas Rea [19] exists entirely on screen.

Embodied conversational agents are subject to anthropomorphic expectations and acceptance by commonality [25, 24]. That is, the embodied conversational agent's appearance implies its functionality and character. A simple example is an agent's

ethnicity affecting the user's attitudes. When the ECA and user have similar ethnicity, the embodied conversational agent is perceived as more socially attractive, similar and trustworthy [43, 22]. In an experiment where user-ECA similarity is brought to an extreme and users interact with their own face the computer is perceived as more fair and trusting, and comments made by the computer are remembered more positively [58]. Computer users are affected not only by the appearance of the ECA but also by the quality of the visual image. A more realistic embodied conversational agent is thought to be more useful, friendly and intelligent, and to understand users better [18, 93].

Superficial judgment of embodied conversational agents is not restricted to their visual appearance. Nass et al. [59] demonstrated that users perceive computer generated voices as having personality. In the study, students listened to book descriptions after which they evaluated their experience. The students were influenced by the reviewer's voice, perceiving the personality of the reviewer differently depending on the voice.

Users interacting with ECAs derive conclusions from the embodied conversational agent's multi modal output. Multi modal output consistency is extremely important; inconsistencies in output modalities degrade the user's experience and efficacy of the ECA [18]. For example, if an ECA's appearance is very realistic and its voice is clearly synthetic then users will react negatively to this inconsistency in output. The opposite is also true, a realistic voice paired with a computer generated image has a negative impact on user experience. The computer appearance and voice must match [22].

The user imposes their desire for consistency on the interaction scenario preferring

embodied conversational agents that reflect the context of the interaction [91, 81]. For example, medical advice is better accepted by an embodied conversational agent that acts and looks like a doctor, or advice on the purchase of a car would be better accepted from an ECA that acts and looks like a mechanic.

2.3.3 User Response to Behaviour

An ECA's social intelligence is its ability to maintain a dialogue with a user such that the ECA appears to behave as a normal individual (e.g. the ability to carry a regular conversation, possibly over multiple separate interactions). An embodied conversational agent's social intelligence has a significant effect on the user-ECA interaction. Users judge ECA behaviour as more important than appearance [24]. For example, an ECA's intelligence is established by its ability to interact socially not by its image [37]. An embodied conversational agent incapable of social intelligence quickly becomes uninteresting and, perhaps, even annoying. Users engaged in long-term interaction with Laura found her repetitive, diminishing the user experience [10]. Likewise, when users obtained text editing help from an embodied conversational agent their perception of the agent substantially improved if irrelevant suggestions were avoided [93].

Successful embodied conversational agents will likely engage and captivate their users intellectually through similar psychological phenomena present in human relationships. Very little is required to socially engage users with even crude computer applications manipulating social behaviour [14] and having (in some sense) personalities [60]. Minimal social triggers create computers with personality and these triggers

solicit a response comparable to human-human interaction. For example, Rickenberg et al. [72] confirmed, whether electronically or in person, the notion of being observed by an animated character impacted anxiety and performance[5] similarly to when observed by a person. Consider another example where a basic language with certain characteristics (a infinitive tenses, elementary utterances composed by co-ordinated clauses, a limited vocabulary and highly restricted dialogue) generated the notion of anthropomorphism greater than that attributed to a computer without this basic language [25].

Users require queues such as distinct voices to consider an embodied conversational agent as human [62]. The question remains as to which cues create socially interesting human-like ECAs. One possibility is to imbue agents with the same characteristics that captivate viewers when watching cartoons; that is, uniqueness will make ECA characters more interesting and consequently believable [47]. A more general proposition by Hayes-Roth and Doyle (1998) is to construct (ECA) characters with ten specific qualities: back-story, appearance, content of speech, manner of speaking, manner of gesturing, emotional dynamics, social interaction patterns, role and role dynamics [47]. These characteristics are meant to provide synthetic characters such as ECAs with lively autonomy and individual personas. For example, appearance conveys "*race, size, build, hair color and style, weight, age, and gender - and certain elements of the characters personal history and temperament such as socioeconomic background and style*" [47]. These attributes distinguish one ECA from another, building individuality between ECAs; similarly, each of the ten specific qualities help create distinct ECAs.

[5]Users performed not as well and were more anxious.

2.3.3.1 Appearance and Behaviour Consistency

Agents consistent in appearance and behaviour are likable and believable (examples of research supporting consistency [91, 81, 4, 69, 26]). The multi modal output, social intelligence, emotions, appearance, and other aspects of the embodied conversational agent must be synchronized.

2.3.3.2 An Alternate Perspective and Its Consequences

There is fairly strong evidence (e.g. [70] and its backing research, [43], [22]) supporting the Media Equation and, in context of computers, the computer as a social actor (CASA). Although this thesis subscribes to CASA, there exist several instances where users have reacted contrary to CASA [29, 76, 11]. That is, users react socially but they do not react as they would towards other people - their human-computer interaction does not mirror their human-human reaction.

Goldstein et al. [29] demonstrated that people are not always polite towards small computers such as personal digital assistants, challenging a previous results by Reeves and Nass. Shechtman et al. [76] constructed an experiment where users were led to believe that they interacted with either a computer or a human (via a computer interface); when, in fact, both groups interacted with a single computer program. Their results describe a "*media inequality*" [76] where the groups respond differently to identical cues. Similarly, Bonito et al. [11] did not find a linear relation between human-like features in the interface and a positive evaluation of the interface (more features implies a more positive evaluation), though they expected one.

Much of the research supporting how users respond to ECA appearance and be-

haviour presents hypotheses, discussions and conclusions in context of CASA - users reacting socially towards computers as they would towards other people. In absence or refutation of CASA, the explanation of research results (why users responded as they did) may require the introduction of a new paradigm. In other words, if user response is not captured by CASA, then there may exist another overarching paradigm typifying user response. Futhermore, the logical arguments that assume CASA become more tenuous. For example, consider the experiment presented in Chapter 4. The experiment's hypotheses are based on the expectation that CASA applies to the experiment's context. The experiment's hypotheses may still be supported by results however the logical arguments applying a social (psychological, etc.) model to user response may require re-examination. In absence of CASA, it is likely that user response is due to other factors present in the interface, and not due to those factors, injected into the interface, meant to build social interaction. Without CASA, the observed response remains but the suggested reason behind such an observed response becomes questionable.

2.3.4 User Response to ECAs Summary

Not only do users engage in anthropomorphic descriptions applying social terminology to their computer interactions, the Media Equation proposes human-computer interaction as fundamentally social in nature. In this light, embodied conversational agents are social actors[6] performing on the social stage and being judged accordingly. Thus, user response to ECAs and ECA groups is expected to parallel an individual's

[6]Embodied conversational agents as social actors is similar to computers as social actors (CASA); terminology derived from a publication authored by Nass and colleagues [61]

response to social scenarios.

Superficial judgment and expectations dominate user response to embodied conversational agent appearance, however, entertaining the user intellectually is more important. That is, behaviour is more effective in user response than appearance. Consider examples drawn from past and present communication: written communication such as the letter, electronic mail and cell phone text-messaging provide social interaction without visual stimulus.

This section reinforces elements of *persona* (Section 2.2) as an important ECA characteristic in effecting user response. During the design, creation and evaluation of an ECA group simulation framework (Chapter 3), special consideration is given to appearance and behaviour.

This section suggests that ECA behaviour appears to have a greater effect than appearance and is manipulated during the experiment (Chapter 4). This includes modifying an ECA's behaviour in an attempt to increase the user's response to ECA groups.

2.4 Summary

Users can respond differently to a single agent versus multiple agents. However, understanding of this difference in response remains poor. A survey of ECAs suggest certain characteristics (physical awareness, communicational awareness, learning, modeling the user's emotional state, gesture coordination & understanding, and persona) are candidates for affecting user response. Furthermore, literature reports that the user responds socially (naturally) to the interface, that ECA behaviour and ap-

pearance affect user response and that ECA behaviour has a greater influence than appearance. These characteristics and aspects of user response are considered in a system designed to simulate ECA groups (Chapter 3) and the user's social approach to the interface and their strong response to ECA behaviour impact the design of the experiment (Chapter 4).

Chapter 3

WOZECA - A Framework for ECA Experiments

This chapter introduces WOZECA and the motivation behind its construction in Section 3.1. The remainder of this chapter describes WOZECA starting with an overview (sections 3.2 and 3.3) and moving to implementation details such as protocols (Section 3.4) and movie clip production (Section 3.5). It concludes with an evaluation of WOZECA (Section 3.6).

3.1 Motivation and Previous Work

An environment is required to examine user response to embodied conversational agent groups. Implementing a group of ECAs similar to those seen in Section 2.2 is prohibitively time consuming and expensive, and any such an ECA group would be so specific it would constrain experiments to only a few features. Experiments based on current ECAs contend with ECA limitations making testing of interaction alternatives

outside of current implementation constraints difficult. Rather than implementing an ECA and ECA group, embodied conversational agents can be simulated, allowing various combinations of characteristics.

The Wizard-of-Oz (WOZ) terminology describes a system where a facade, seen and manipulated by a user, is controlled and simulated by a hidden individual called a Wizard [64]. In the case of computer-human interaction experiments (Chapter 4), a WOZ setup is an effective means of testing[1] interfaces and interaction concepts prior to the time consuming and error prone task of implementing a complete system.

A WOZ system for research on user response to ECA groups should be capable of simulating many of the characteristics available in implemented ECAs (Section 2.2). Furthermore, it should allow for the customization of the ECA's appearance and behaviour since these affect user response. This flexibility and reconfigurability facilitates different kinds of experiments and the expression of the associated characteristics such as conversational or physical awareness.

Suede [36], CrossWeaver [78] and DEMAIS [6] are popular Wizard-Of-Oz applications that do not focus on embodied conversational agents but are, nonetheless, ECA simulation candidates. The WOZ systems developed by NICE project [48] and Cavalluzzi et al. [23] are ECA focused and, consequently, more likely candidates. Below, these candiates are evaluated as to their appropriateness for simulating ECA groups and supporting a variety of ECA group experiments.

Suede facilities the prototyping of speech based interfaces whereas CrossWeaver and DEMAIS, although not limited to prototyping of conventional interfaces, cater

[1] A properly developed experiment will enforce the appropriate amount of consistency across all Wizard performances. For example, the experiment described in Chapter 4 uses specific rules (see Section 4.4.2.2) to enforce consistency across Wizard performances.

to windowed GUIs. CrossWeaver and DEMAIS function without a behind-the-scenes Wizard. Pre-scripted inputs are used to prescribe the interaction response instead of a Wizard. For example, CrossWeaver allows speech input and pen gestures to transition a prototype application between states. Suede differs, relying on a behind-the-scenes Wizard to activate a response to user input.

This thesis' research includes unstructured conversation (within context of ECA groups) that may not follow a predetermined path, even when there are repeated and observable tendencies [65]. Consequently, the system was built to help a Wizard organize the available responses to user statements without a prescribed structure. It allows the Wizard to adopt different conversational structures such as the speech act model [90], as depicted in Figure 3.1, or to proceed without any preconception of structured interaction. More concretely, the WOZ system furnishes mechanisms such as a drawing palette which help a Wizard organize the available ECA responses into any discovered structure.

Suede, CrossWeaver and DEMAIS provide a notation for structuring the interaction between the user and the prototype similar to a finite state machine. Suede uses a flowchart whereas CrossWeaver employs a storyboard.

The NICE project investigates ECA interaction in a two dimensional fairy-tale game environment [48] with a focus on multi-modal interaction (pen and speech), stylized 2D characters for child users. Such specific aims are too constrained for a general ECA research making the NICE WOZ system inappropriate.

Cavalluzzi et et al. [23] discuss a WOZ platform based on Haptec agents [2]. Using various configuration mechanisms such as extended markup language files, Cavalluzzi et et al. manipulate every aspect provided by the commercial Haptec software agent

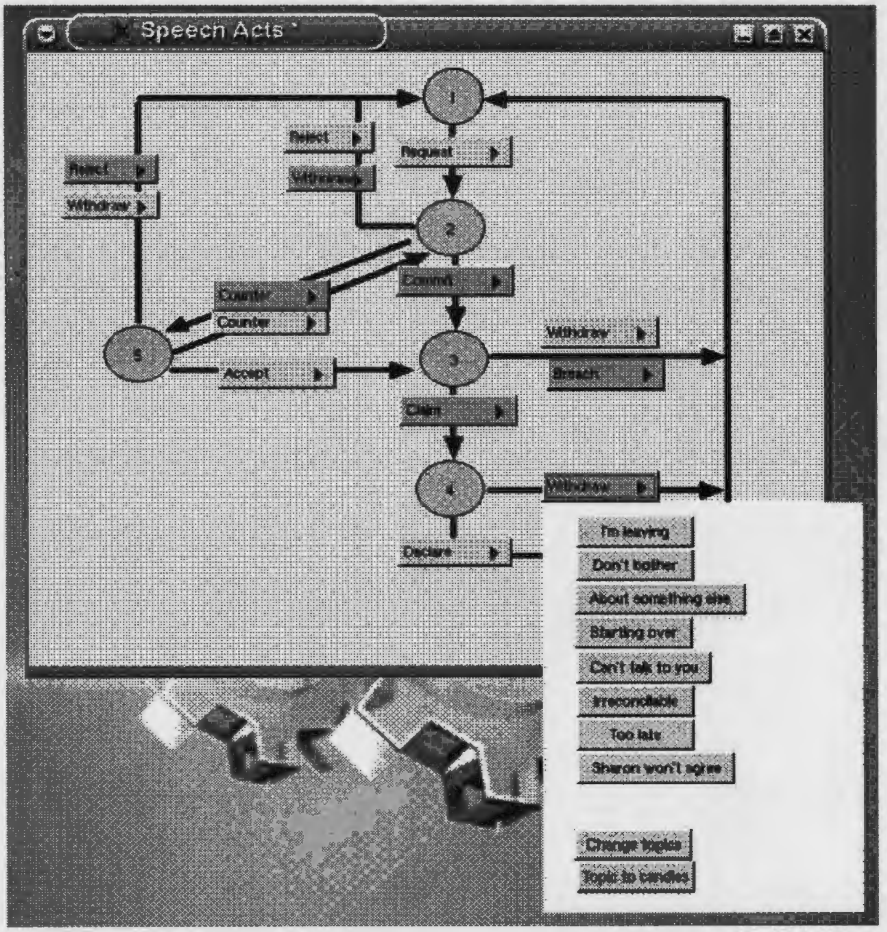

Figure 3.1: A notation in WOZECA based on speech act theory.

(e.g. appearance). Their WOZ platform and research objectives are, of the studied Wizard-of-Oz systems, the most similar to this thesis. Their particular focus is *one user one agent* interaction (as opposed to group interaction) with the Wizard interface designed accordingly. Similar to Suede, CrossWeaver and DEMAIS, their Wizard interface prescribes a conversational structure through programming and configuration. This prescribed structure impedes system flexibility - a flexibility that facilitates different kinds of experiments.

Given the above, the development of new WOZ software was required because 1) few ECAs are capable of group interaction (Section 2.2), 2) implementing ECA groups would be time consuming, 3) applying a Wizard-of-Oz framework to ECA research has found success elsewhere, and 4) the available ECA WOZ systems are inadequate for this research. WOZECA was designed to facilitate embodied conversational agent research such as the experiment presented in the next chapter. WOZECA differentiates itself from other WOZ tools with a malleable Wizard environment, low experiment preparation overhead, a focus on ECA group interfaces and an ability to present various ECA characteristics in the interface.

3.2 Brief Implementation Description

WOZECA currently runs on a Gentoo Linux (2.6.14-r5) box containing a 64-bit AMD Athlon 3000+ processor, an io Vibe ieee1394 card (NEC uPD72874 chip) connected to a Canon Elura 85 and a Nvidia NV43 GeForce 6200 graphics card. The hardware was chosen for cost, compatibility and performance. The graphics card is well supported by both Linux and MPlayer (a video player), and the digital video camera

performs well in low light environments. The camera includes a microphone-in jack for sound recording which should improve the recording quality. The Revolution 2.6 programming environment expedites Wizard GUI development with built-in palette tools and an object editor. Kino[2] (used version 0.7.5 - latest version 0.8.0) provides simple video editing.

WOZECA's primary implementation weakness is the Revolution/Transcript component. The Transcript interpreter treats all strings as script; even user dialogue messages arriving over the network. Consequently, the Wizard occasionally sees the evaluation of the user's message (e.g. the word False) rather than the dialogue message itself, and certain characters, such as the comma, are removed from dialogue messages to avoid crashing the Wizard's interface. (From experience, this issue affects the Wizard's response. However, string evaluation of the user's message occurs infrequently, rarely causing problems. In the case of the experiment presented in Chapter 4, string evaluation did not affect the Wizard's response.)

3.3 Overview

In WOZECA (Figure 3.2), a Wizard initiates the user-agent conversation through an ECA appearing on the user's interface, and the user responds to this ECA with text chat. In turn, the Wizard reads the user's statement and selects the appropriate ECA response by clicking on a button representing the selected response.

The interface agents manifest themselves in windows on the user's computer and their related audio track (typically speech) is heard via attached speakers. The agent

[2]Kino's instability is a reflection of the immaturity of multi-media software on Linux systems.

window's size and location is configurable permitting various agent window layouts on the user's screen. For instance, users can experience two agents side-by-side (Figure 3.3), one centered agent (Figure 3.4) or three agents in the shape of a pyramid. Users communicate to the agents by typing in a chatter box window, typically located in the bottom portion of the screen, that resembles Instant Messaging or Internet Relay Chat in functionality. When agents react, the chatter box window remains in focus to accept key strokes. Since users require (only) the keyboard to interact with the on screen agents, no mouse is needed. Under these conditions, users are unaware of the behind-the-scenes Wizard and believe they are interacting with the agents.

The Wizard's interface to WOZECA is considerably more complex (Figure 3.5 and 3.6). His main window, called the Manager (Figure 3.5 item 1), provides essential functionality such as loading the ECA responses, access to the tools used to annotate a Wizard's work area and control of the network connection. The Wizard creates a work area by loading a movie clip inventory into a new empty window. The movie clip inventory (Figure 3.5 item 2) appears as several lists of buttons where each list is preconfigured grouping. Once an inventory is present, either buttons or groups of buttons may be copied or re-arranged within the inventory window or to a new empty work area window (not shown). A button's main function is to represent and play, when pressed, a movie clip. Buttons may also activate quick-windows (Figure 3.7 item 3) that are similar in functionality to drop-down (pull-down) menus. Instead of menu items, a quick-window appears with the same capabilities and semantics as all work area windows. The Wizard may annotate windows using drawing and other tools (Figure 3.5 and 3.7 item 4) to create graphics such as circles, squares or background colours.

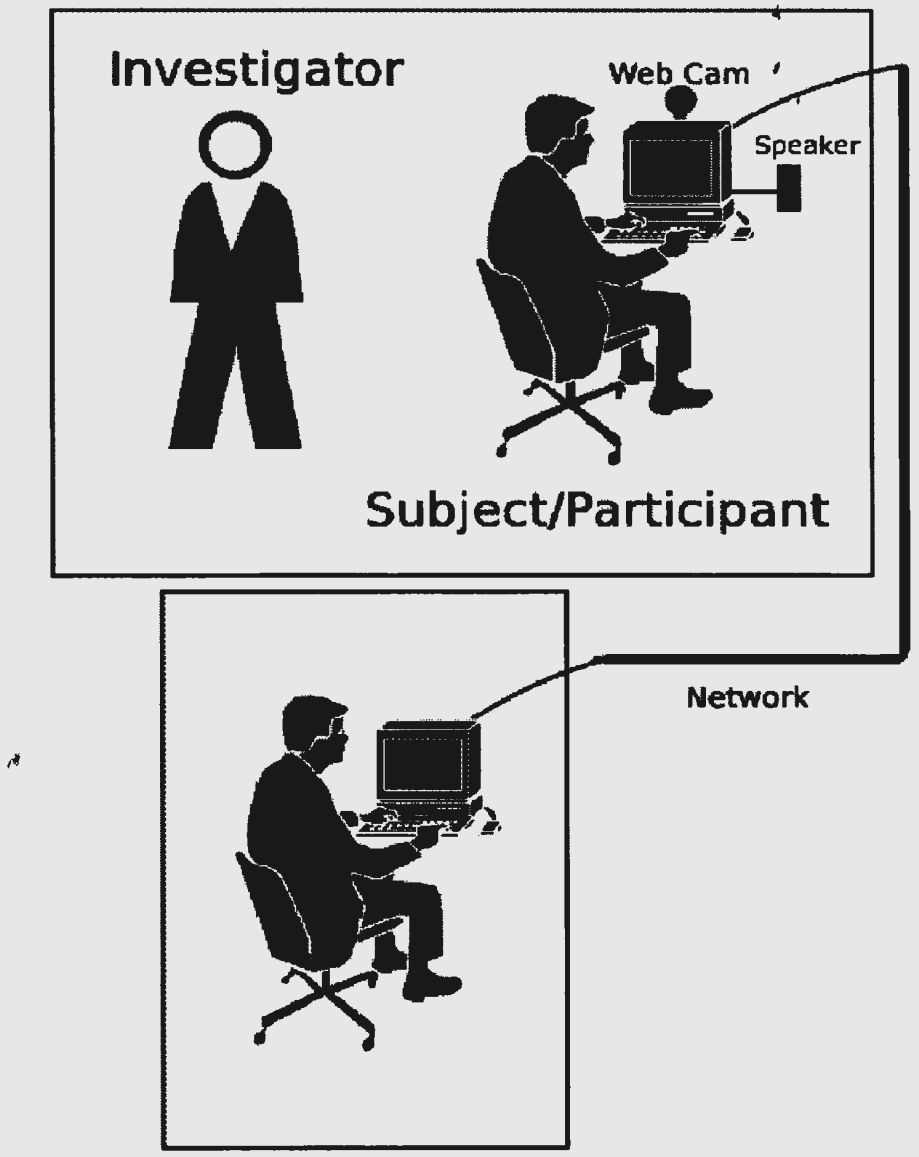

Figure 3.2: A graphical depiction of the WOZECA system.

Figure 3.3: User's interface when interacting with two agents.

Figure 3.4: User's interface when interacting with one agents.

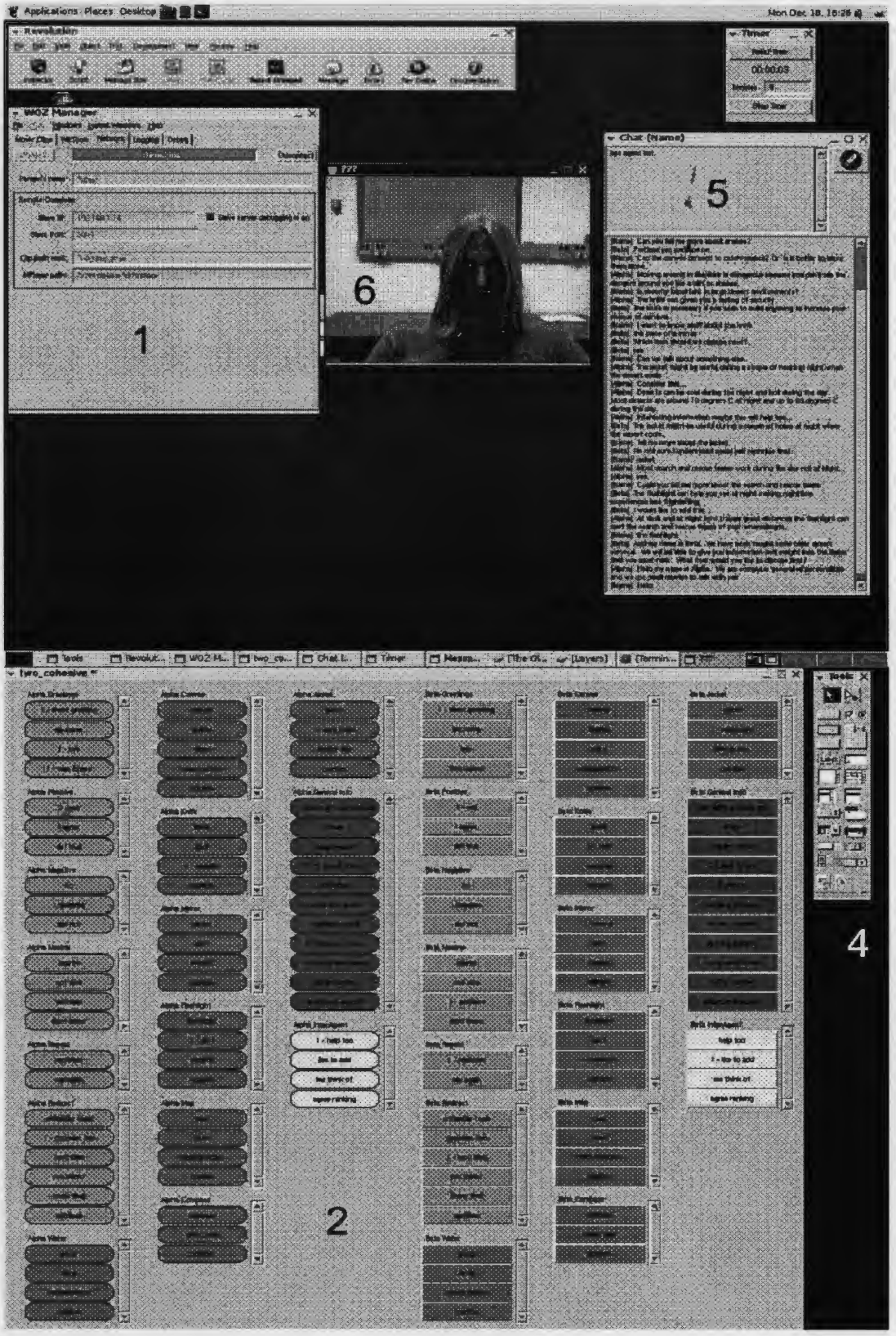

Figure 3.5: The wizard's interface during a test phase (two monitors shown vertically).

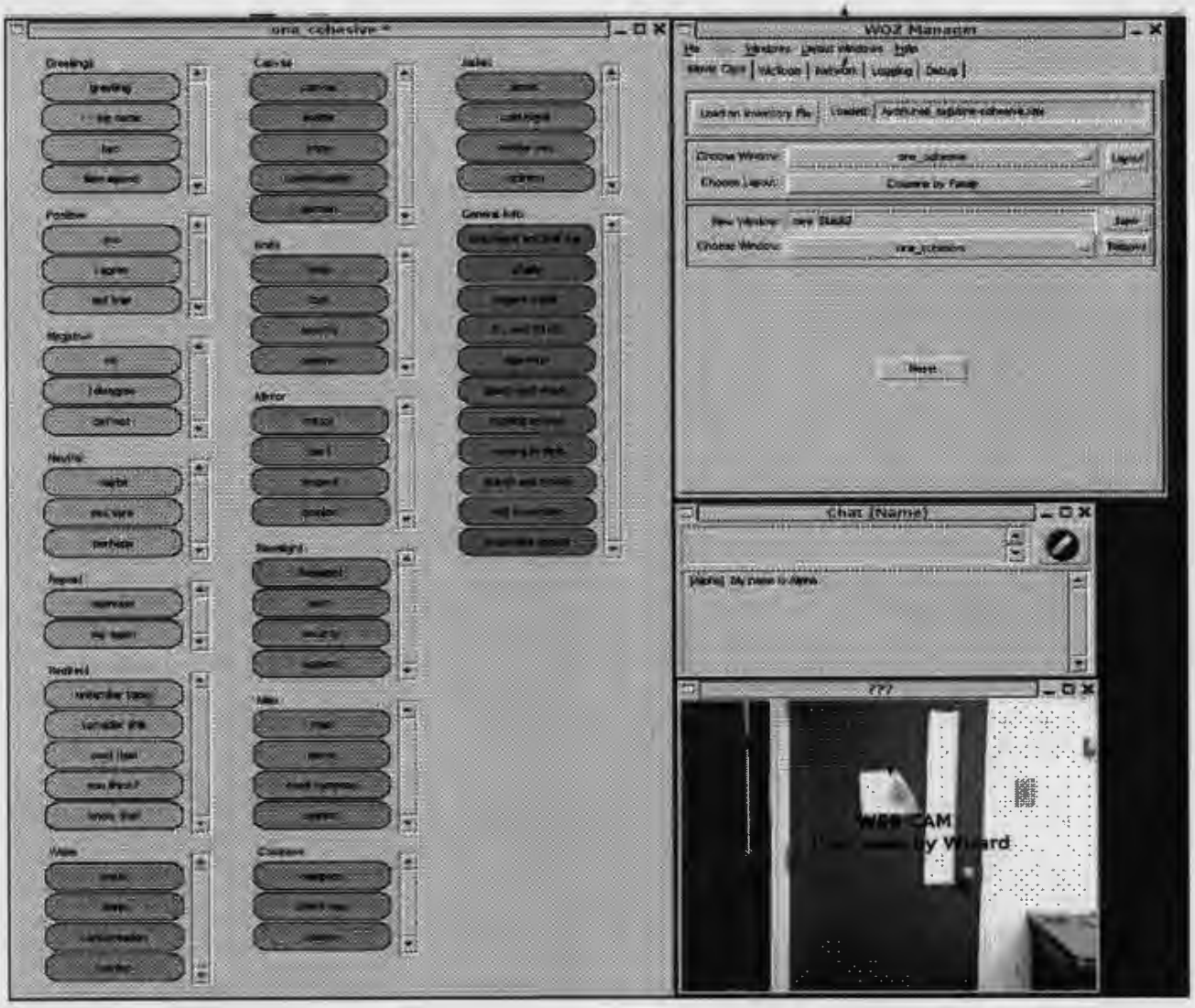

Figure 3.6: A Zoom-in of the wizard's interface during a test of the system.

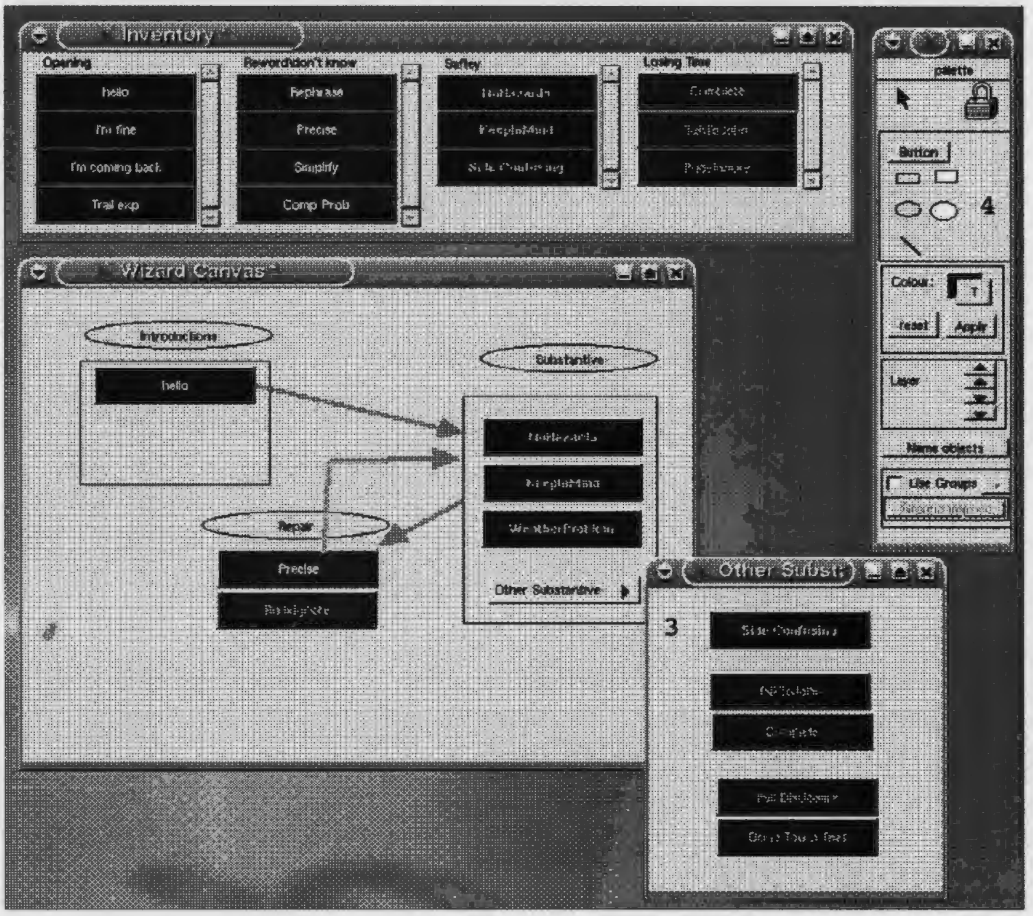

Figure 3.7: Inventory layout using the Wizard's tools.

A "chat with history" (Figure 3.5 item 5) window displays the user's typed messages in real-time as well as the user's final message interleaved with text spoken by the agents. The dialogue between the agents and the user is constantly available to the Wizard including the message being composed by the user. Furthermore, a web cam (Figure 3.5 item 6) placed on the user's computer captures the user's facial expressions and body language displaying these captured images, in real-time, in a window accessible to the Wizard. This allows the Wizard to use the ECA's movie clips more effectively when communicating with the user.

Some of WOZECA's important features are: simulation of multiple agents, remote-control of agents via a network, a configurable Wizard work area that permits the Wizard to optimize the Wizard environment, and the ability to save this work area.

3.4 Design Details

Underlying the user and Wizard's interfaces is cross-platform client-server architecture software. The Wizard's interface, a client, was written using Revolution 2.6/Transcript (www.runrev.com), chosen for its rapidity in developing graphic user interfaces. The server, located on the user's computer, was written in Python. The server and client-server protocol act as an abstraction layer between the Wizard and embodied conversational agents. For this thesis, the server hid the evolving MPlayer movie control protocol. The abstraction layer supplies WOZECA with flexibility and extensibility such that the Wizard's interface and the ECAs are not tightly coupled (i.e. one can change without modification of the other).

Upon receiving a TCP connection, the server creates a thread to handle the sim-

ple low-bandwidth asynchronous protocol between itself and the Wizard's interface. Thus, each TCP connection receives a dedicated thread managing communication between the server and the Wizard's interface. Another thread is created for the sole purpose of interpreting MPlayer feedback during movie clip play. When MPlayer, as a child process of the server, plays a movie clip, a thread connects to MPlayer's standard input and standard output streams; controlling MPlayer through its standard input and reading MPlayer's standard output to relay movie progress to the Wizard interface (Figure 3.8).

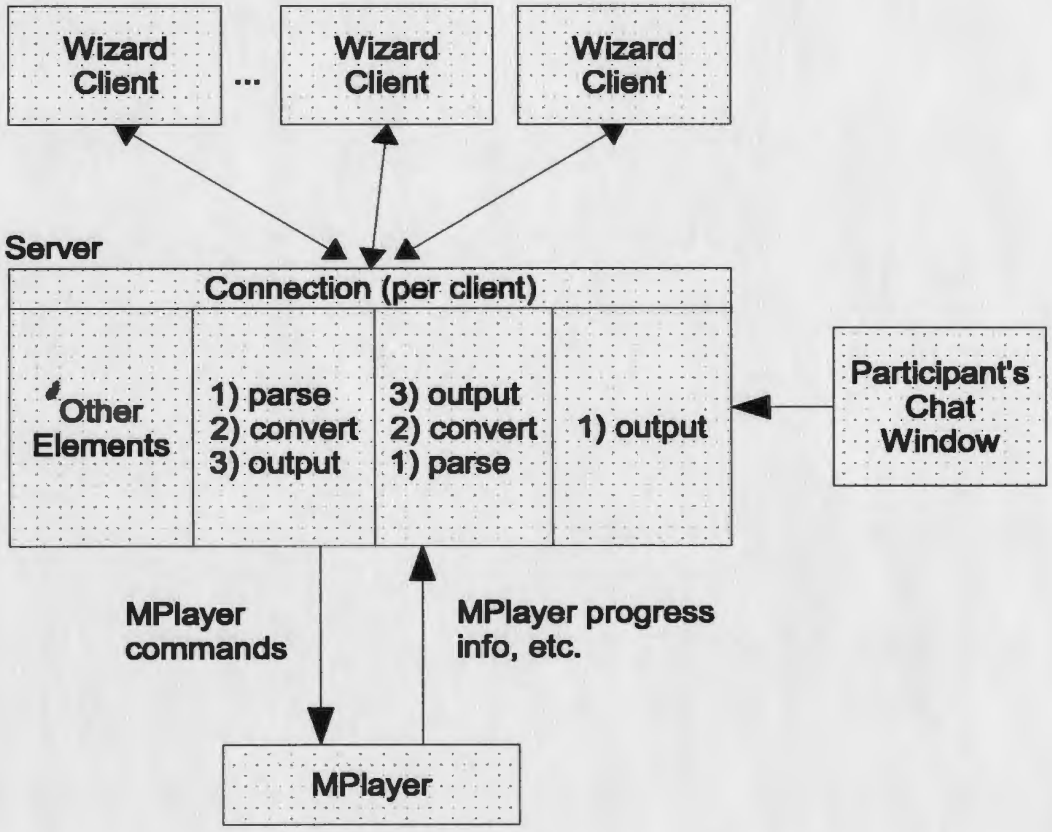

Figure 3.8: An Architecture Diagram of WOZECA with a Focus on MPlayer.

The server protocol uses messages where the general message format is

$$< scope >:< command >< subscope > [=< values >]$$

This message format allows for the addition of future messages without substantial code modifications. Additional messages are expected to reuse the network input and output routines as well as the message parsing and composition functions.

The element $< scope >$ denotes which server side agent should fulfill the embedded request. For example, if the server contains two agents named *Foo* and *Bar* setting the scope to *Foo* directs the message command to agent *Foo* rather than *Bar*. The element $< command >$ designates the action to take or the request to fulfill. For example, the "play" command tells an agent to play a specific movie clip. For a complete listing of commands see table 3.1. The commands *set* and *get* read and change $< scope >$ properties. For a listing of the properties manipulated by *set* and *get* see table 3.2. The element $< subscope >$ narrows the command's effect and the optional element $< values >$ represents the value (e.g. function arguments) to assign to the $< subscope >$. An example message using every element is

$$Foo : play\ mclip = myTestClip.mpeg$$

which tells the agent named Foo to play the movie clip "myTestClip.mpeg". For examples of the messages in action see table 3.3.

There are two notable design features of WOZECA. First, WOZECA should perform well in many different network environments such as wireless, for the the network is burdened with only web cam images and a simple protocol. Placing a wireless enabled computer on a mobile platform should allow experiments to relocate to quiet public areas in order to increase experiment participation (bringing the laboratory to users: portable usability laboratories and usability kiosks [64, page 205-206]). Sec-

Table 3.1: Protocol Message Commands

Command	Description	Example Usage
play	plays a movie clip	$Foo : play\ mclip = myTestClip.mpeg$
get	get a property	$Foo : get\ loop$
set	set a property's value	$Foo : set\ loop = True$
value	state a property's value	$Foo : value\ loop = True$
playing	the progress of movie a clip	$Foo : playing\ myTestClip.mpeg = 23$
close	closes the connection or server	$system : close\ server$

ond, the server manages multiple connections simultaneously such that many agents may be controlled by several unique Wizards. This creates more realistic conversation and conversational environment possibilities. That is, between-agent reactions may be more genuine considering that the Wizards are unlikely to know the exact clip that will be played by fellow Wizards.

3.5 Movie Clip Production

WOZECA does not contain the necessary tools for the production of movie clips. This separation is desired since the best tools, platform or environment for the manufacture of movie clips may change. Currently, a digital video camera captures the video which is transfered in real-time, via ieee1394, to a Linux box. Once the recording of video is complete, it is edited and an XML file (the movie clip inventory) is created or updated with the recently captured movie clips; making the movie clips available to the Wizard.

Table 3.2: Protocol Message Properties

Property	Description
debug	enables debugging
mplayerPath	the path to the mplayer executable
mclipRootPath	the path prefixed to all movie clip paths
winx	x of the upper left corner of the window relative to the screen
winy	y of the upper left corner of the window relative to the screen
winWidth	the windows width
winHeight	the windows height
quiet	changes the manner in which the play progress is reported
loop	indicates the movie should play forever
mclipRelPath	the relative path of the currently playing clip or empty
mclipProgress	the progress of the currently playing clip
preempt	whether or not the playing clip should be preempted by "play"
visible	the chat window's visibility (hidden/visible)
cmessage	a complete message entered by the user
pmessage	a partial message entered by the user

Table 3.3: Example of a Protocol Message Sequence

Example Sequence	Sent by	Received by	Response to Previous Message
Foo : *set xyz* = 2	WOZECA	Server	
Foo : *get xyz*	WOZECA	Server	
Foo : *value xyz* = 2	Server	WOZECA	Yes
Foo : *get winx*	WOZECA	Server	
Foo : *value winx* = 50	Server	WOZECA	Yes

3.6 Summary and Evaluation

WOZECA is a low cost cross-platform versatile Wizard-Of-Oz solution that utilizes few network resources. It provides a malleable work area instead of constraining the Wizard to certain notational semantics and allows one or multiple Wizards to control multiple ECAs. Its focus on multi-agent dialogue interfaces distances itself from other WOZ systems. Furthermore, a malleable Wizard environment adds flexibility to WOZECA, a flexibility absent in many WOZ systems.

ECA characteristics (defined in Section 2.2: physical awareness, communicational awareness, learning, modeling the user's emotional state, gesture coordination & understanding, and persona) that affect user response provide a means for assessing WOZECA's capabilities. Following this assessment, ECA appearance and behaviour (Section 2.3) are also considered.

- *Physical awareness* is an agent's ability to track or manipulate physical objects in the user's space and hold communications with respect to these objects.

WOZECA's physical awareness is restricted to the area and details captured by the stationary web-cam (manipulating objects is impossible). ECAs may refer to static objects using prerecorded gestures and speech. However, non-stationary objects are beyond WOZECA's current capabilities. The Wizard can not see objects outside of the web-cam view or refer to these objects using prerecorded video clips. ECAs may circumvent the web-cam restrictions when referring to unique physical objects. For example, the ECAs can refer to a (unique) red square whether it is visible or not.

- *Communicational awareness* is an agent's ability to understand varying types of conversation (speech or text) and their subtleties.

  The ECA's communicational awareness is restricted by the information supplied to the Wizard through input devices. Currently, it is restricted to gross gestures captured via web-cam and the user's ability to converse using text messages.

- *Learning* is characterized by the use of multi modal interaction with human users as ECA learning input. That is, the user trains the ECA via speech and gestures.

- An ECA accounting for (e.g. changing behaviour) and updating information representing the user's emotional state is considered to be modeling *emotional state.*

  Since the Wizard supplies the ECA with its learning abilities, ECA learning is restricted only by WOZECA's multi-modal devices - web cam and text chat.

Similarly, modeling the user's emotional state depends on whether or not Wizard can "read" the emotions on camera or in the user's written text. The Wizard's dilemma is expressing learned elements and the ECA's response to the user's emotional state with a finite set of video clips. However, the Wizard's flexibility, due in part to the use of video clips, should outweigh the limits of this finite set of video clips.

- *Gesture coordination and understanding* is an ECA's ability to communicate to their users without speech or text.

  WOZECA relies on actors for appropriate body language corresponding to ECA statements. With respect to gestural understanding, the web-cam captures gross gestures and obscures details. For example, the Wizard can see the user's head movement as they look to the left but would miss the user's smirk. The current web-cam situated on the computer monitor often deprives the Wizard of facial expressions while the user is typing. Improving gesture understanding is matter of supplying better information to the Wizard via improved input devices such as additional cameras.

- An ECA's *persona* consists of their appearance, role (i.e. purpose), back-story, behaviour patterns and emotions. The ECA persona is one of the least developed aspects of the ECA and, in many cases, their persona is restricted to their function.

  WOZECA controls ECA appearance and emotions through the experimenter's

choice of digital video recording or animation, the applied special affects and the choice of actor. In other words, WOZECA's appearance and behaviour are not limited to (complex) programming. Instead, the recorded videos supply an ECA with these traits. ECA behaviour depends on the ECA video clip inventory and the Wizard's ability to combine video clips. If an ECA requires additional or different behaviour then supplementary video recordings are created. Adjusting the ECA's appearance and supplying video clips with appropriate content complete the ECA's role and back-story.

WOZECA fulfills the role of multi-ECA simulator by placing ECA characteristics and implementation difficulties on the behind-the-scenes Wizard or configurable video clips. WOZECA is restricted by its input and output devices as well as the Wizard's ability to control these devices through her interface. A variety of experiments concentrating on various ECA characteristics are possible using the WOZECA framework (such as the experiment in Chapter 4). Experiments, interested in ECA characteristics not currently available in WOZECA, may extend WOZECA with additional input and output devices, and extend the Wizard's interface to control these devices.

Chapter 4

A WOZECA Experiment: User Response to Single and Multiple ECAs

The previous chapter described a framework (WOZECA) for experimenting with ECAs. This chapter describes an experiment realized using WOZECA. The experiment investigates user response to the addition of a second agent to a one agent group. It also manipulates the ECA group's behaviour, chosen because behaviour appears to be the dominant influence on user's response (Section 2.3). The resulting four conditions (1 ECA and Behaviour A, 2 ECAs and Behaviour B, 1 ECA and Behaviour B, 2 ECAs and Behaviour B) require measures of user response.

Anthropomorphism (Section 2.3.1.1) and the Media Equation (Section 2.3.1.2) suggest user response is fundamentally social and natural. This social aspect of user response motivates the application of theoretical constructs from psychology. This

experiment draws primarily on the existing constructs of group cohesiveness (Section 4.1.1) and valence & arousal (Section 4.1.2) to quantify user response. It also draws on self-efficacy (Section 4.1.3).

Group cohesion (cohesiveness, Section 4.1.1) represents the user's attraction to the ECA group. Valence and arousal capture the user's emotional state (Section 4.1.2) and self-efficacy influences future task performance (Section 4.1.3). This thesis does not propose these effects represent the only effects of an ECA interface, nor does it prove that they are the strongest effects. Past research has investigated other effects such as: group conformity (the pressure to comply to group norms) [44]; trust, liking, and respect towards an ECA[10]; similarity and trust between ECAs and their users [43, 22]. However, these measures do not account for the group aspect of the interface nor do they explicitly rate the ECA group.

Group cohesiveness, valence & arousal, and self-efficacy are seen as interesting measures of user response based on the existing literature. For example, members of cohesive groups rate their groups more positively (e.g. more helpful, more satisfaction). Valence (e.g. happiness) and arousal (e.g. excitement) represent the user's emotional state. Self-efficacy affects performance with strong self-efficacy improving an individual's resilience to failure and the quality of their successes.

This chapter first presents group cohesiveness, valence, arousal, and self-efficacy (sections 4.1.1, 4.1.2 and 4.1.3 respectively). Section 4.4.2 present the independent variables (number of agents, ECA behaviour) and their operationalization of phenomena affecting the constructs. Section 4.4.3 discusses how the independent variables are expected to effect group cohesiveness, valence, arousal, and self-efficacy. The results (Section 4.5) and discussion (Section 4.6) indicate that user response is dependent on

an additional ECA and its behaviour, as well as user traits.

4.1 Effects

4.1.1 Group Cohesiveness

Shaw defines a group as "*two or more persons who are interacting with one another in such a manner that each person influences and is influenced by each other person*" [75, page 8]. This thesis modifies Shaw's definition to include an embodied conversational agent as a group member. Restated, a group is defined as two or more persons, or ECAs, who are interacting with one another in such a manner that each group member influences and is influenced by each other group member.

Prior to 1950, group cohesiveness was a descriptive term [31] referring to a group that hangs together, is tight or close knit - a united group. Festinger, Schachter and Back [31] provided the first formal definition of group cohesiveness which was quickly distilled into a simplified notion: group cohesiveness is *attraction-to-group.* Attraction-to-group is the attraction of a member in the group to other members of the group, or inter-personal attraction. This distilled definition structured a great deal of the group cohesiveness research, focusing group cohesiveness measures on the individual rather than the group. Reluctance to accept this view of group cohesiveness as an inter-member attraction is concisely formulated by Carron: "*operational measures of cohesion based upon attraction underrepresent the concept*" [31]. For example, individuals of a cohesive rowing team worked well together, yet disliked each other (Lenk 1969) [31].

Although traditional human-only groups are likely underrepresented by attraction-to-group, attraction-to-group is well suited for a group containing a single human member (the computer user) and embodied conversational agents. The experiment scenario in this chapter is a single user interacting with multiple embodied conversational agents. This lends itself well to measuring group level phenomena in terms of an individual. The single user multi-ECA group exists only because the user participates, and without user attraction to the ECAs the group is likely disbanded.

The remainder of this subsection describes several causes and consequences of group cohesiveness relevant to this experiment. The causes of interest are cooperation and behaviour. The consequences are increased communication and a more positive evaluation of the group [31].

Behaviour affects group cohesiveness. During a study of mutual support groups for parents of children with special needs and disabilities, parents reported many characteristics of a cohesive group such as friendships or communication of emotions that were not present in other mother and toddler groups [79]. Parents rated their groups as very helpful as well as highly cohesive. More concretely, behaviour such as friendliness, warmth, sensitivity and helpfulness are important factors in the promotion of group cohesion [31, 75].

Two effects of increased cohesion are increased and better communication between members as well as an increase in positive evaluation of the (group) situation and group members [31, 75]. For example, researchers studying walking groups in Queensland Australia concluded that group cohesion was (of those measured) the sole predictor of individual's adherence to walking groups, and a positive attitude towards the group's activity [39].

4.1.2 Valence and Arousal

Wundt in 1896 was the first to propose three dimensions for describing affective meaning of stimuli leading to a categorization of the stimulus response [12]. The stimuli's affect on an individual, thus measures of an individuals emotional state, is categorized on three dimensions named *valence*, *arousal* and *dominance*. Dominance is of little interest because the chosen valence-arousal-dominance instrument is a poor measure of dominance (low correlation). This approach is similar to one taken by Reeves and Nass [70]. From this point onward dominance is ignored.

Valence, originally described by Wundt as *pleasure*, is a measure of a person's contentedness. To further describe valence it is best to borrow from Mehrabian and Russell's Sematic Differential Scale that consists of adjectives describing each (valence and arousal) dimension [12]. The scale measuring valence describes low valence with words such as unhappy, annoyed and unsatisfied; in contrast words such happy, pleased and satisfied describe high valence. Arousal is "*the degree to which an individual feels him or herself to be worked up or emotionally intense about what one is doing*" [35] (Apter, 1989, p9). Arousal is similarly described by word pairs such as relaxed-stimulated, calm-excited and dull-jittery with the left word of the pair describing low arousal and the right word high arousal.

Valence and arousal scales provide insight into an individual's emotional state. For example, erotica pictures are perceived as pleasant and arousing whereas a picture of a flower is seen as pleasant but calm [12]. More recently, Reeves and Nass had subjects score 258 video segments finding that valence and arousal were separate emotional dimensions, and that the video segments caused a wide scope of emotional reactions

[70].

4.1.3 Self-Efficacy

Self-efficacy, introduced by Bandura in 1977 [66], "*refers to the beliefs in one's capabilities to organize and execute the courses of action required to produce attainments*" [7]. The four major sources of self-efficacy being enactive mastery experiences, vicarious experiences, verbal persuasion and psychological & effective state [7].

Enactive mastery experiences are the greatest source of increased self-efficacy with successful experiences producing strong generalized self-efficacy beliefs. The concept of an enactive mastery experience is captured by the quote attributed to Sophocles "*One must learn by doing the thing; though you think you know it, you have no certainty until you try.*".

Success is also judged by comparison of one's self with others. For example, tall people are tall relative to the surrounding geographic population. Vicarious experiences allow individuals to adjust their self-efficacy beliefs by comparing their success and capabilities to other's.

Persuasion such as convincing arguments or encouragement can strengthen or weaken self-efficacy. If persuasion occurs within a plausible context then it may produce an effect on an individual's self-efficacy. For instance, persuading an individual that they can fly is unlikely to have an effect on their belief of such, whereas persuading a programmer of their ability to complete a master's is likely.

An individuals predisposition affects their self-efficacy. For example, a hockey team losing several consecutive games will believe themselves less capable of winning

their subsequent games. Physiological and affective states are information available to an individual for self-judgment of capabilities.

Those with higher self-efficacy set higher goals and are more committed to achieving their goals. Furthermore, these individuals are more likely to succeed because of strong perseverance and persistence [66]. Individuals with a higher self-efficacy also demonstrate better control over their emotional experiences; either by perceiving the experience as less emotional or by more effectively coping with the experience. For example, individuals who believe in their public speaking skills are likely to find public speaking benign or control their anxiety.

4.1.4 Effects Summary

Group cohesiveness, valence, arousal and self-efficacy are the effects of interest in this experiment. Cohesiveness is the attraction of a group member to the group as a whole. Cooperation and behaviour (such as warmth or friendliness) increase the likelihood of a group becoming cohesive. Once cohesive, there is a tendency for the group to communicate more and for group members to evaluate the group positively. The two dimensions, valence and arousal, describe an individual's emotional state. Self-efficacy is an individual's belief in their ability to achieve their objectives. Self-efficacy impacts future task performance with higher self-efficacy generally resulting in better task performance.

4.2 Goals

Although the user believes the ECA group's goal is to help them in their task (Section 4.4.1), the experiment's primary objective is to study the effect of an additional ECA on user response and, in so doing, probe the difference between a single agent and two agents in the interface.

Other potential benefits are as follows: Reproducing cohesion in embodied conversational agent groups should be advantageous if the benefits of cohesive groups can be produced. Cohesive group members evaluate their group positively. Interacting with groups that leave their user happy and energetic, an emotional state described by valence and arousal, has presumptive value. Strong self-efficacy is shown to contribute to an individual's resilience to failure and quality of their success. An ECA group interface that improves user self-efficacy is expected to make further interactions with the interface more successful. A cohesive ECA group interface that increases user valence, arousal and self-efficacy is expected to benefit the user.

4.3 Participants

Experiment participation was solicited via posters placed in high-traffic locations throughout the university campus. The ensuing sample of convenience comprised 32 participants aged between 17 and 45, with a mean of 23.3 and median of 21.5. There were 15 male and 17 female participants. Most were students, either graduate or undergraduate from varying backgrounds, and each was paid 11 dollars (CAD) for one and a half hours of work.

4.4 Experiment Factors

The experiment is a two factor, two level design with one factor, agent behaviour (levels: neutral, positive), measured between subjects and the other factor, the *number of embodied conversational agents* (levels: 1 and 2 ECAs), measured within subjects. Number of agents was measured within subjects and was balanced. Half the participants interacted with one agent followed by two and half the participants interacted with two agents followed by one.

4.4.1 The User Task: The Desert Survival Problem

The desert survival problem (DSP) [40][1] requires participants to order eight items (flashlight, knife, map, magnetic compass, one 2L bottle of water, a piece of a mirror, 6 meter by 6 meter piece of blue canvas, jacket) from most important to least important for their survival in the desert. It provides conversational motivation and context for the participant during the experiment and limits the scope of the human-computer interaction under investigation (see appendix B for details).

Moon [51] constrains a conversation with the DSP in an experiment that examines attitudes and behaviour during distant communications across networked computers; Bradner et al. [13] also employ the DSP to constrain a conversation; Nass et al. [60], Shechtman et al. [76] and Morkes et al. [54] exercise the desert survival problem for the purpose of an experiment scenario while probing human-computer interaction related to relationships, personality and humour; Burgoon and Bunito et al. [11, 18] constrain conversation while studying user reactions to agent appearance and method

[1]The premise of ordering desert survival items is the primary idea borrowed from the original DSP by Lafferty et al. .

of communication.

Likewise, this experiment exploits the desert survival problem for conversational boundaries, motivation and context.

4.4.2 Independent Variables

This section describes the independent variables and explains how they operationalize causes of cohesion in groups. It also briefly reiterates the independent variable's predicted effect on group cohesion.

4.4.2.1 Number of Agents

There are two groups: the single agent group where the participant interacts with one ECA and the two agent group where the participant interacts with two ECAs. The two agent group is more immediately viewed as a group than the single agent group simply by numbers. For example, the definition of group casts the group as multiple individuals [32]. In consequence, it is important to know if the two agent group rates as more attractive than the single agent group.

The participant experiences the one and two agent groups. The embodied conversational agent that appears to the participant in the single agent group reappears to the left of a second ECA during the two agent group. In other words, one agent is reused from the one agent group. The ECAs are virtually identical in this experiment (see Figure 3.3). The same actor performed all ECA roles, was video taped using the same background and camera position, used one dialogue script (for agents having the same behaviour), and performed (for agents having the same behaviour) as similarly as possible providing the ECAs with identical persona. The spoken greeting in the

single ECA group is shared by two ECAs in the two agent group (see appendix C for phrases). To differentiate between the two embodied conversational agents the actor was asked to change shirt colour, not shirt style, and to wear a head-scarf. With many similarities and very few differences, the effects of two agents are attributed to the addition of a second agent.

Due to space constraints, the ECAs in the two agent group are smaller than the ECA in single agent group. The movie clip windows in the two agent group were scaled to accommodate two agents appearing side-by-side (Figure 3.3). Similarly scaling the ECA in the single agent group uncovered a large portion of the desktop which was negatively critiqued in very early pilot studies.

4.4.2.2 Behaviour

Participants experience either a neutral or a positive behaviour[2] while interacting with the embodied conversational agent groups. The behaviour is intended to operationalize causes of group cohesion. The difference between the neutral and positive behaviour is explained using three ECA aspects - voice, script and dialogue - as summarized in table 4.1.

The neutral agent's voice is emotionless, robotic and somewhat monotone whereas the positive agent's voice is energetic. The energetic voice was well received during the actor's nine years of customer service in retail. That is, clients enjoyed the energetic voice based on their feedback. The energetic voice should be perceived as friendly

[2]Terminology such as behaviour A and behaviour B could have been used to differentiate the two behaviours. However, such terminology is found to be more confusing and to make communication more difficult due to the visual similarity and lack of semantic difference between terms. The names neutral and positive, though pejorative, were chosen to increase the readability of this thesis.

Table 4.1: Summary of positive-neutral behaviour differences.

Context	Neutral Behaviour	Positive Behaviour
voice	emotionless	energetic
script	no agent interaction	inter-agent phrases
dialogue	query-response	query-response and volunteer information

and warm (behaviour of cohesive groups and group members), consequently it should increase the participant's cohesiveness rating of the positive group.

One actor read the ECA's dialogue (script) from a computer monitor eliminating mistakes and unintentional dialogue differences between ECAs. The difference in script occurs in the positive two agent group which employs four inter-agent dialogue phrases. An example inter-agent phrase is "*Interesting information, maybe this will help too...*" (see appendix C for all four phrases). These inter-agent phrases allow the ECAs in the positive group to interact with each other as well as the participant. These phrases are meant to increase the cooperation, communication and helpfulness of the positive behaviour groups relative to the neutral behaviour group without introducing additional content. In turn, the participant is expected to evaluate the positive group behaviour as more cohesive.

The Wizard controlled the neutral ECA groups by following three rules established during pilot testing.

- If possible respond to the participant's statement directly. In other words, when several movie clips are possible responses to the participant's statement choose the movie clip that responds to the participant most directly.

- When sent a partial sentence (e.g. keywords), to stimulate a dialogue, request that the statement be rephrased or restated.

- Do not interrupt the participant while they formulate an idea, statement or question. If the participant is inactive (not typing) for three to five seconds then play a movie clip.

These rules establish a query-response behaviour where the participant queries the ECA and the ECA responds with the most appropriate statement available. This behaviour reduces the amount of communication, helpfulness, cooperation and verbal interaction to the minimum level demanded by the circumstances.

Where the neutral agent is restricted to responding to participant statements, the positive agent volunteers information. The previous three rules and the following additional rule establish the positive behaviour.

- The Wizard may provide (contextually relevant) information without being queried.

The Wizard applies this additional rule in two situations: while waiting for the participant to be inactive for three seconds, and after a response to a participant's query (e.g. provide additional information using a second movie clip). This behaviour increases the amount of communication, helpfulness, cooperation and verbal interaction within the ECA-participant group.

The positive behaviour increases the communication, helpfulness, cooperation and verbal interaction between the ECA group and the participant. Thus, the positive behaviour should increase group cohesion (as rated by the participant) relative to the neutral behaviour.

Furthermore, these rules enforce a particular Wizard behaviour across all participant-ECA interactions. This leads to consistent ECA behaviour during the experiment.

4.4.3 Dependent Variables - Predictions and Instrumentation

This section discusses how the independent variables from the previous section are expected to affect group cohesion, valence & arousal, and self-efficacy.

4.4.3.1 Group Cohesion

Reproducing cohesion in embodied conversational agent groups is advantageous assuming the benefits of cohesive groups survive in ECA groups. The neutral and positive behaviour operationalize causes of cohesive groups. An energetic voice, ECAs volunteering information, and four inter-agent phrases are aspects of the positive behaviour expected to increase cohesion in groups.

Group cohesion was measured using the Relative Group Cohesiveness Rating (RGCR) questionnaire (see appendix D.5) adapted from group cohesion questionnaires by Evans and Jarvis', and Stokes [31]. Certain statements duplicated between the two questionnaires and other statements referring to the group temporally were dropped, such as "I dread coming to this group". From each original statement, three were created: one favouring the single agent group; a statement favouring the two agent group; and a neutral statement. For example, the original statement "I want to remain a member of this group." was modified producing "I want to remain a member of the one agent group.", "I want to remain a member of the two agent

group.", and "I do not prefer to remain a member of one group above the other.". Three additional questions directly targeting the human-agent interaction were added to the end of the questionnaire.

Besides treating the questions individually, a cumulative score was calculated by assigning a value of 0 for neutral responses, a value of 1 for responses favoring the two agent group and a value of -1 for responses favoring the one agent group. The sum of these values resulted in scale ranging from -13 (preference for one agent) to 13 (preference for two agents). The RGCR score refers to this cumulative value ranging between -13 and 13.

Two ECAs by their numbers and behaviour are expected to present aspects (e.g. friendly behaviour or inter-member communication) causing cohesive groups (for details see Sections 4.1.1, 4.4.2.1 and 4.4.2.2).

- **Hypothesis 1** Group cohesion will be greater when interacting with two ECAs over a single ECA.
- **Hypothesis 2** Group cohesion will be greater when interacting with positive behaviour ECAs over neutral behaviour ECAs.

4.4.3.2 Valence and Arousal

The emotional dimensions valence and arousal, an aspects of the participant's emotional state, affect user response. Aspects that improve group cohesion and consequently increase overall group members' satisfaction (towards the group) should also increase valence and arousal responses for the group's members. For example, cohesive groups generally increase member satisfaction - an aspect of the valence di-

mension [31]. If the agent's positive behaviour causes group cohesion, valence and arousal ratings should rise accordingly. More simply, the positive behaviour should increase participant arousal and valence relative to the neutral behaviour. In the same way a field of flowers is more arousing than a field with a single flower, interacting with multiple embodied conversational agents is expected to increase a participant's arousal relative to interacting with fewer.

The Self-Assessment Manikin (SAM) [12] measures participant's valence and arousal. The SAM is a standardized measure rating both emotional dimensions on a pictorial nine point scale (5 pictures and four empty areas) and is extremely quick to complete.

- **Hypothesis 3** Participant valence will be higher for the positive ECA group over the neutral group.
- **Hypothesis 4** Participant valence will be higher when interacting with two ECAs over a single ECA.
- **Hypothesis 5** Participant arousal will be higher for the positive ECA group over the neutral group.
- **Hypothesis 6** Participant arousal will be higher when interacting with two ECAs over a single ECA.

4.4.3.3 Self-Efficacy

This experiment does not subject the participant to an enactive mastery or vicarious experience increasing self-efficacy (see Section 4.1.3). Nor do the ECAs claim sufficient or appropriate movie clips to be verbally persuasive (using logical arguments). With the major sources of self-efficacy virtually absent in this experiment,

embodied conversational agents are not expected to influence a participant's desert-survival-item-ranking self-efficacy. However, two agents may be more persuasive than a single agent due to their matching content (i.e. "If two people/ECAs say the same thing then it must be right."). Testing self-efficacy was not a significant addition to procedure complexity and was measured.

The self-efficacy instrument can be found in appendix D. The application of general measures to specific processes results in faulty assessment of self-efficacy [7, page 63]. For example, there is little relationship between an individual's general self-efficacy and their ability to solve quantum physics problems. Similarly, a general self-efficacy test inadequately measures a participant's desert-survival-item-ranking self-efficacy. A specific measure of desert-survival-item-ranking self-efficacy could not be found. Consequently, a questionnaire measuring desert-survival-item-ranking self-efficacy was created and is based on a general self-efficacy questionnaire [74]. The general self-efficacy questions were used as a guideline when formulating the specific desert-survival-item-ranking self-efficacy questionnaire.

- **Hypothesis 7** Self-efficacy will remain unchanged.

4.4.4 Additional Data

This section discusses additional data collected during the experiment.

4.4.4.1 Personal Style Assessment

Personality and personal style commonly affect people's relationships [34] and, intuitively, they should affect participant's relationships with embodied conversational

agents. Although the experiment did not focus on the effects of participant personal style on ECA relationships, collecting data on the participant's personal style did not unduly complicate or prolong treatments. The Jung-Parry Personal Style Assessment Form (PSAF) was chosen because of its simplicity (16 self-administered questions), duration (approximately 20 minutes), and acceptance (still in print since initial research by Jung in 1920 and the basis for the well know Myers-Briggs Type Indicator) [34]. The form assesses four (intuitor, thinker, feeler, sensor) of the eight styles put forth by Jung. The personal styles are briefly described in table 4.2 [34].

Table 4.2: Jung-Parry Personal Style Summary

Intuitor	Thinker	Feeler	Sensor
likely to go off on tangents, not mindful of time, brings imagination and freshness to decisions and problems	is described as ordered, measured, suggests ground-rules for conversations	spontaneous, empathetic, raises questions about one's well being	abrupt, gets to the point, interrupts, needs to control the conversation

4.4.4.2 WOZECA Log Data and Coding

Prior to starting pilot sessions, the ECA's statements were categorized (greeting, positive, negative, neutral, repeat, redirect, water, canvas, knife, mirror, flashlight, map, compass, jacket, general information, and inter-agent phrases). These categories

organized the movie clip inventory and consequently improved the Wizard's response time to participant statements. Participant response, during pilot sessions, appeared to depend on these categories. With the knowledge that categorization of statements (coding) is an accepted step in analysis [5], several natural categories were constructed from the original categorization without the use of a formal framework. This coding of the ECA's script produced the following six categories: general information, item information, prompt, repeat, inter-agent phrases and other (see appendix C). General and Item information are of interest since the user's task - desert survival - relies heavily on this information dispensed by the agents. Prompt and repeat suggest ECA-participant communication cues. The inter-agent phrases facilitate group behaviour, and behaviour is known to be important in user response.

WOZECA records every conversational statement, with accompanying date and time, from both the Wizard and participant in a the conversation log file. The conversation log data and coding permitted analysis of participant-ECA conversation and its duration. Chapter 4 discusses results relating to the categories General information, Item information and Prompt, referring to this coding as conversation style.

4.4.4.3 Interviews

The interview is an opportunity for the participant to explain their response to the ECA group, particularly the reactions (e.g. laughter) noticed by the experimenter and not measured by questionnaires. The participant's experience discussed during the interview often provokes ideas for future experiments.

For this experiment, interviewed participants discussed ECA behaviour and appearance as well as their reaction and thoughts to the entire interaction experience.

The semi-structured interviews began with the general question "What did you think of your experience?" and proceeded to more specific questions such as "During your experience, what did you prefer the most and why?" (see end of appendix B).

4.4.5 Procedure

The agent order (1-2 and 2-1 due to a balanced within subjects design for the factor number of agents) in which the participant experiences the interaction with the ECAs is decided prior to meeting the participant. After receiving the interaction order, the behind-the-scenes Wizard controlling WOZECA waits in the observation and control room [64, page 201]. Subjects meet the experimenter in the experimenter's cubicle and proceed to the laboratory (lab). The lab is a small bland room with a computer situated along one wall. A table, where the participant completes questionnaires and is interviewed, is on the opposite wall. After a brief introduction, the participant signs a consent form and is given a voucher for eleven dollars.

In order to give the participant an idea of their future experience, a description outlining the treatment process is read by the experimenter. Within this description the participant's motivation - the desert survival problem (DSP) - for interacting with the ECAs is introduced; after which the participant completes a small demographic questionnaire (appendix D). The desert survival problem scenario asks the participant to imagine crash landing of a helicopter in the desert from which eight items are salvaged. To help the subject rank the eight items, a black and white picture of a desert with shrubs, sand and rocky hills appears on each DSP item rating questionnaire.

The participant interacts with both the two agent and single agent group because of the within subjects design for the factor *number of agents*. The DSP ranking as well as the participant's self-efficacy, valence and arousal is measured before the participant's first interaction with the embodied conversational agents (time T1), between their first and second interaction with the ECAs (time T2) and after the final interaction (time T3).

The treatment concludes with a measure of the participant's attraction to the one agent group relative to the two agent group, a ten to thirty minute interview and completion of a (Jung-Parry) personal style assessment form.

4.5 Results

This section suggests that users respond differently to a single agent and two agent group. Most importantly, the ECA group's behaviour and the user's age and personal style affect the user's attraction-to-group. To arrive at this conclusion, the results were analyzed using the R (www.r-project.org) statistical package [63]. The data were visualized (e.g. tables, plots, boxplots) to discover where statistical test should be applied; with many of these figures appearing below. The results are presented in American Psychological Association (APA) style (apastyle.apa.org).

4.5.1 A Concise Presentation of the Variables

The main factors, the number of agents and their behaviour, produced no statistically significant effect on group cohesion, valence, arousal and self-efficacy. Collecting data from 32 particpants was expected to be sufficient for producing statistically

significant results having strong effects. However, the powers vary between 0.35 and 0.99 (calculated for a significance level of 0.05) which may suggest a need for a greater N to capture the smaller than expected effects. (Only four power of test calculations produced results below 0.8, of which two were 0.76. Power of test values below 0.8 are presented.)

This absence of strong effects led to the use of several opportunistic tests (related variables Table 4.3) to better comprehend the experiment's data. Figures, such as boxplots, were inspected for interesting, possibly statistically significant artifacts; when found the appropriate statistical test was applied to verify and confirm the visual artifact. To better interpret the opportunistic test results, several alpha levels were calculated based on the Bonferroni criterion[3] (Bon.) [89, page 606]. The Bonferroni criterion was applied (for Table 4.3) as follows:

- Family A included two opportunistic tests for order (1-2, 2-1), likewise for time (a test may be performed for time T1-T2 and T2-T3).
- Family B included 13 opportunistic tests corresponding to the 13 RGCR questions.
- Family C included six conversation styles and three conversation styles derived from the six, for a total of 9.
- Family D is participant age.
- Family E includes the four dimensions of personal style.

[3]The Bonferoni criterion acts as a mechanism to control the inflation of family-wise alpha levels. It states that one can maintain a confidence level, $1-\alpha$, for several tests, m, by setting the confidence level for each test to $1-\alpha/m$ such that the cumulative confidence level of m tests remains equal to or below $1-\alpha$.

- Family F includes six collapsed dimensions of personal style (collapsing two dimensions of personal style can be done $\binom{4}{2}$ ways for a total of 6).

To test for possible confounding between the main factors and variables of interest (due to opportunistic tests), an ANCOVA was run with each variable in Table 4.3; none reached a significance level of 0.3 or less. The model[4] consisted of the two main factors (the number of agents and their behaviour), their interaction and the tested variable, from Table 4.3, as a covariate.

Table 4.3: Consice Presentation of Variables

Variable	Opportunistic	Family	Bonferroni	Sections
Number of agents	-	-	-	4.5.2 - 4.5.4
Behaviour	-	-	-	4.5.2 - 4.5.4
Agent order (2-1, 1-2)	-	-	-	4.5.2, 4.5.3
Time (T1-T2, T2-T3)	√	A	$0.05/2 = 0.025$	4.5.3, 4.5.4
RGCR scores by question	√	B	$0.05/13 \simeq 0.004$	4.5.2
General information	√	C	$0.05/9 \simeq 0.006$	4.5.2, 4.5.5
Item information	√	C	$0.05/9 \simeq 0.006$	4.5.2, 4.5.5
Prompting information	√	C	$0.05/9 \simeq 0.006$	4.5.2
Information sum	√	C	$0.05/9 \simeq 0.006$	4.5.2
Participant Age	√	D	$0.05/1 = 0.05$	4.5.2
Personal style Sensor	√	E	$0.05/4 = 0.0125$	4.5.5
Personal style feeler-thinker	√	F	$0.05/6 = 0.009$	4.5.2

[4]The outcome is predicted by the number of agents, their behaviour and the covariate, according to the following linear model in R lm(outcome ~ numberOfAgents * behaviour + covariate).

Although the main factors produced no statistically significant effect on group cohesion, evidence (details see Section 4.5.2) suggested that users responsed differently to the neutral and positive behaviour. Analyzing these behaviours separately (table 4.4) showed agent order as a marginally significant effect in the neutral behaviour and not statistically significant in the positive behaviour. Table 4.5 contains aspects of the user-positive-behaviour-ECA-group interaction that were likely to effect the participant's ECA group preference. Thus, the ECA's behaviour (e.g. conversation style), the participant's personal style and their age likely affect the participant's preference; possibly explaining the perceived differences between neutral and positive behaviour.

Table 4.4: Preference for Participant's Second Interaction, by Behaviour

Behaviour	R^2	$F(1, 14)$	Alpha	Power
Neutral	0.24	4.46	$p < 0.054$	> 0.42
Positive	0.03	1.5	$p > 0.24$	-

4.5.2 Group Cohesiveness

An additional agent and its behaviour were poor predictors of participant attraction-to-group (Figure E.8 and table E.1). However, Figure 4.1 shows that the order in which the participants experienced the agent groups (agent order[5]) affected their preference[6] for a group, with participants preferring their second interaction, $R^2 =$

[5] *Agent order* results from a balanced within subjects design for the factor number of agents.

[6] When reporting group cohesiveness results, the term *preference* may be used to enhance text readability instead of *attraction-to-group*. For example, "the participant's preference" is equivalent

Table 4.5: Additional Aspects Likely Affecting Participant Attraction-to-Group (Positive Behaviour)

Aspect	Preference	R^2	$F(1,14)$	Alpha	Bon.
Conversation style: more Information discussed with single agent	1 Agent	0.35	7.51	$p < 0.016$	0.006
Personal style: type feeler-thinker	2 Agents	0.24	4.37	$p < 0.056$	0.009
Age: older participants	2 Agents	0.18	3.03	$p > 0.10$	0.05

0.16, $F(1,30) = 5.86$, $p < 0.022$, ($pow > 0.63$). Even though participants typically rated their second interaction as more cohesive, many participants scored items on the cohesion instrument as neutral - preferring neither the single or two agent group. More precisely, there was a (strong) statistically significant neutral scoring for RGCR questions 3, 7, 9 and 11 (table 4.6). These questions used the language *feel* and *care*. For example: "I don't care what happens in one group more than the other".

to "the participant's attraction-to-group".

Table 4.6: RGCR Scores by Question (* indicates $p < 0.004$)

Ques.	No. of Participants Selecting			$c^2(2, N = 32)$	Power
	1 Agent	2 Agents	Neutral		
1	12	16	4	$p < 0.03$	-
2	12	14	6	$p > 0.19$	-
3	4	4	*24	$c^2 = 25.0, p < 0.0001$	> 0.97
4	11	13	8	$p > 0.55$	-
5	9	7	16	$p > 0.12$	-
6	10	14	8	$p > 0.41$	-
7	9	4	19	$c^2 = 10.93, p < 0.0043$	> 0.86
8	9	7	16	$p > 0.13$	-
9	6	2	*24	$c^2 = 25.75, p < 0.0001$	> 0.97
10	11	14	7	$p > 0.31$	-
11	5	4	*23	$c^2 = 21.44, p < 0.0001$	> 0.96
12	10	14	8	$p > 0.41$	-
13	13	15	4	$p < 0.04$	-

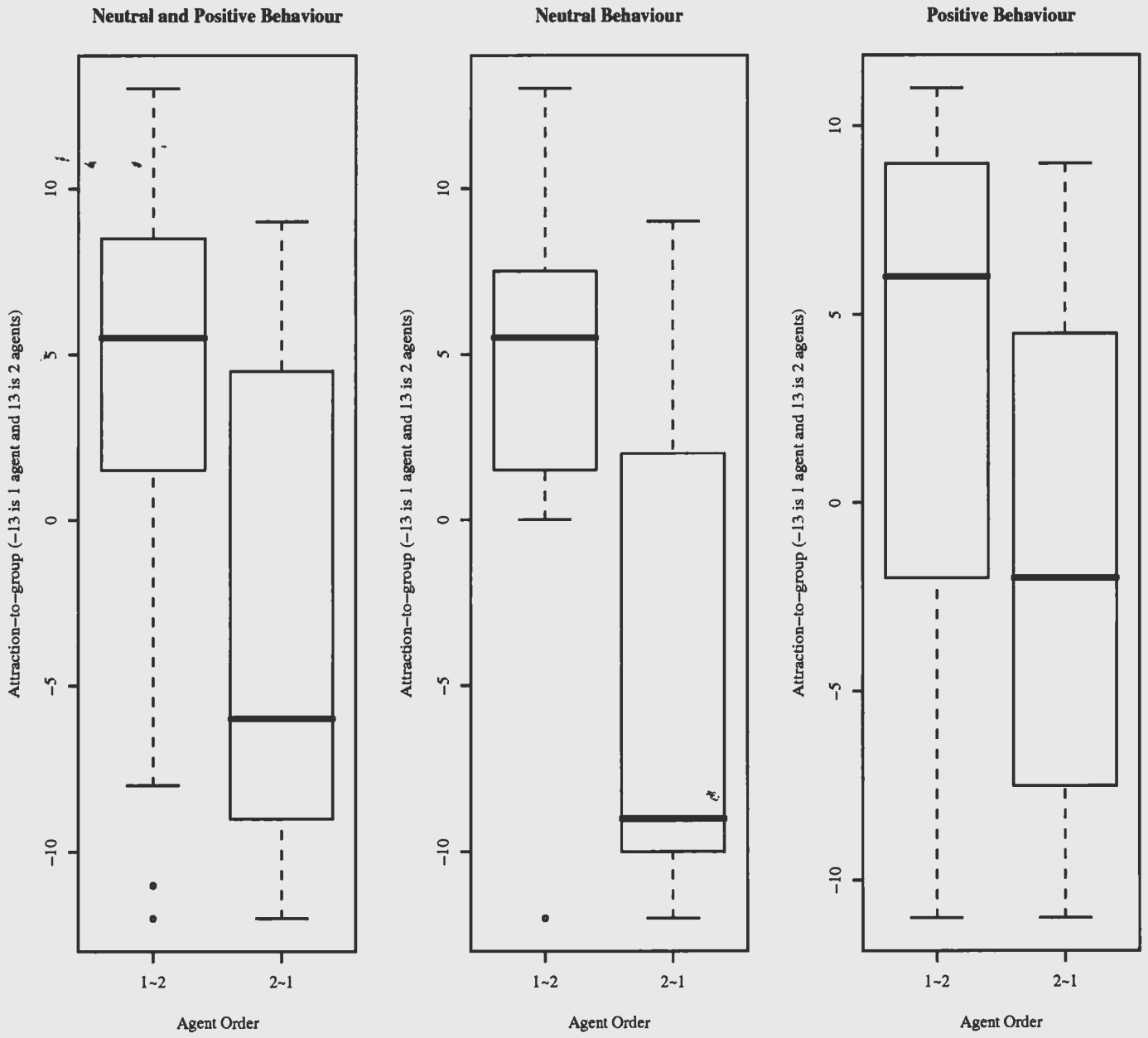

Figure 4.1: Effects of Agent Order on Attraction-to-group.

Interviews (Section 4.5.6) indicated a different user response to the neutral and positive ECA behaviours. Figure E.8, although not statistically significant, supports this difference in response. In the neutral behaviour, summary statistics for RGCR indicated a preference for neither group (mean of -0.44 and median of 1.5, where mean and median of 0 represents preference for neither group and negative is a preference for the one agent group). In contrast, separate summary statistics of the positive behaviour ECAs indicated a preference for the two agent group (mean of 0.88 and median of 3.5). In an attempt to understand this difference in user response the positive and neutral behaviour were then analyzed separately. The next two subsections discuss participant attraction-to-group within each behaviour factor separately (neutral, positive).

4.5.2.1 Neutral Behaviour

The agent order (Figure 4.1) was a marginally significant predictor of participant attraction-to-group for neutral behaviour ECAs, $R^2 = 0.24$, $F(1, 14) = 4.46$, $p < 0.054$, ($pow > 0.42$); with participant's preferring their second interaction.

4.5.2.2 Positive Behaviour

The agent order (Figure 4.1), as a predictor of participant attraction-to-group, was *not* statistically significant for positive behaviour ECAs ($p > 0.24$).

For the positive behaviour ECAs, the older the participant the more General information was discussed with the agents, $R^2 = 0.31$, $F(1, 14) = 6.40$, $p < 0.024$; and there was a strong tendency for younger participants to discuss more Item information, $R^2 = 0.24$, $F(1, 14) = 4.31$, $p < 0.057$. There was also slightly more prompting

in a single agent group; on average 1.6 Prompts versus 1.00 Prompt (almost significant, $p < 0.083$).

For the following results, refer to Figure 4.2 to view the (respective) linear model's variables. The participant's age and the conversation style (for conversation style see Section 4.4.4.2) affected the user's attraction-to-group. An additional agent was preferred by the older participants when its group did not prompt the user and the user discussed little with the single agent. The details of the latter are as follows: The more information (sum of General, Item and Other from Section 4.4.4.2) discussed with the single agent, the more the participant preferred the single agent, $R^2 = 0.35$, $F(1, 14) = 7.509$, $p < 0.016$. Accouting for the participant's age[7] further increased the statistical significance, $R^2 = 0.5$, $F(2, 13) = 6.59$, $p < 0.011$. Also, the less prompting by the two agents, the less information discussed with the single agent, the more the participant preferred an additional agent, $R^2 = 0.47$, $F(2, 13) = 5.77$, $p < 0.017$. Accouting for the participant's age[8] increased statistical significance as well, $R^2 = 0.69$, $F(3, 12) = 9.03$, $p < 0.0022$.

Depending on the participant's personal style, user response to the ECA groups differed. An additional agent is likely preferred by participants that are more feeler-thinker[9], $R^2 = 0.24$, $F(1, 14) = 4.37$, $p < 0.056$. Accounting for the participant's age and the agent order increased statistical significance. The older the participant who is more feeler & thinker and is experiencing the agent order 1-2 (for agent order see

[7]From a linear model having two terms (age and the sum of the counts of General, Item and Other) and response RGCR.

[8]From a linear model having three terms (Prompts, age and the sum of the counts of General, Item and Other) and response RGCR.

[9]The feeler and thinker personal style dimensions were collapsed by summing the feeler and thinker values forming a single dimension.

Figure 4.1) prefers[10] interacting with the two agent group, $R^2 = 0.51$, $F(1, 14) = 4.10$, $p < 0.033$.

[10]From a linear model having three terms (the sum of feeler and thinker, order and age) a second additional agent in the interface and response RGCR.

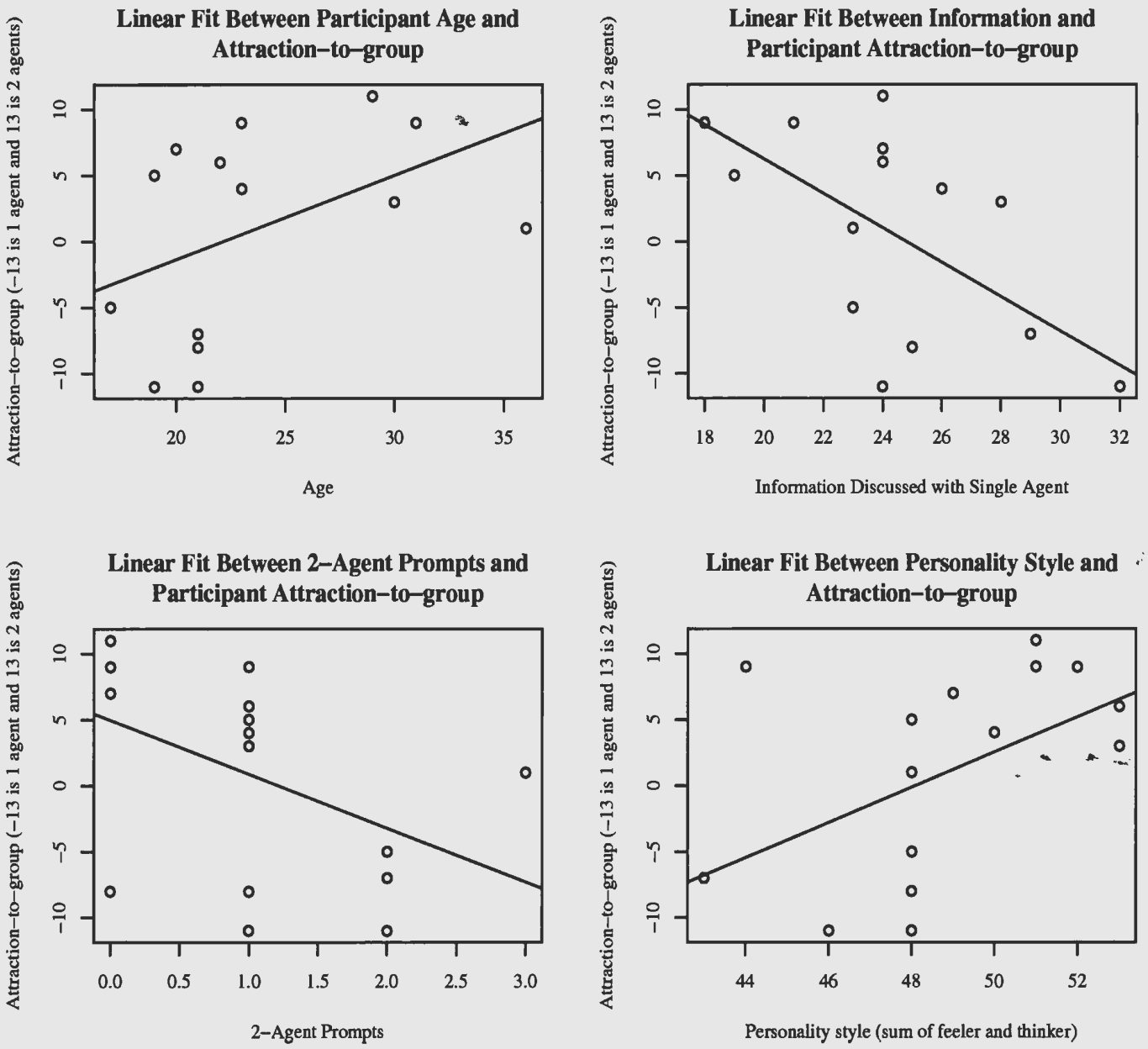

Figure 4.2: Linear Fit Between Selected Variables and Attraction-to-group for Positive Behaviour (several points overlap).

4.5.3 Valence and Arousal

Neither ECA behaviour nor a second additional agent, predicts the participant's valence (Figure E.5, and for more detail see Figure E.4). Participant valence decreases after the first interaction, $t(31) = 2.35$, $p < 0.026$, ($pow > 0.76$); and, as seen in Figure 4.3, continues to decrease when the additional agent is removed from the two agent group (*order 2-1*), $t(31) = 2.15$, $p < 0.5$.

As illustrated by Figure 4.3, the order in which the participant encountered the ECA groups affected their valence. Having participants interact with the one agent group and, subsequently, adding a second ECA to the interface scored higher in valence than beginning with the two agent group and removing the additional agent from the interface, $R^2 = 0.15$, $F(1, 30) = 5.19$, $p < 0.03$, ($p > 0.54$).

The independent variables did not have a statistically significant effect on participant arousal (Figure E.7, and for more detail see Figure E.6). However, Figure 4.4 indicates that participants found the first interaction arousing, $t(31) = 3.5$, $p < 0.0015$.

4.5.4 Self-Efficacy

The additional agent and its behaviour had no statistically significant effect on participant's self-efficacy. On the other hand, Figure 4.5 illustrates that self-efficacy did increase over time; between T1 and T2, and a weaker effect between T2 and T3 (T1-T2: $t(31) = 3.16$, $p < 0.0035$; T2-T3: $t(31) = 2.03$, $p < 0.051$) (for more detail see Figure E.3).

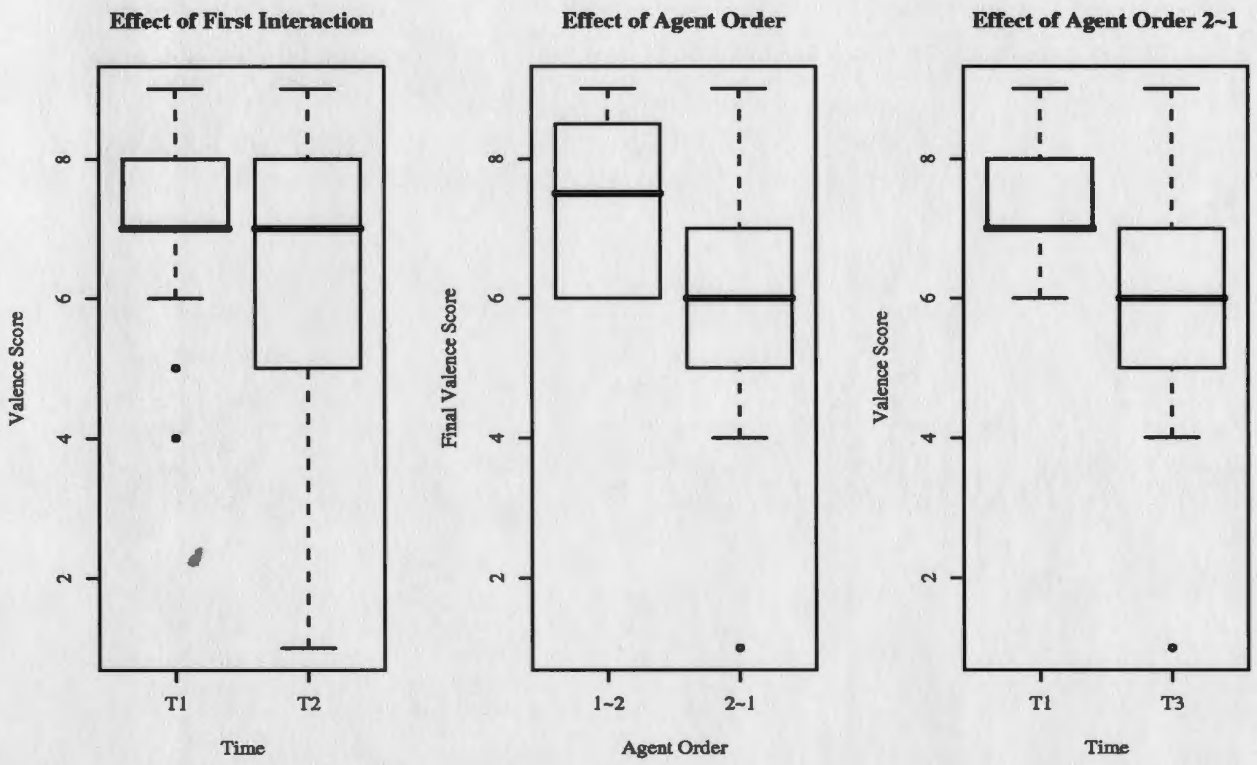

Figure 4.3: Effects of Agent Order on Valence

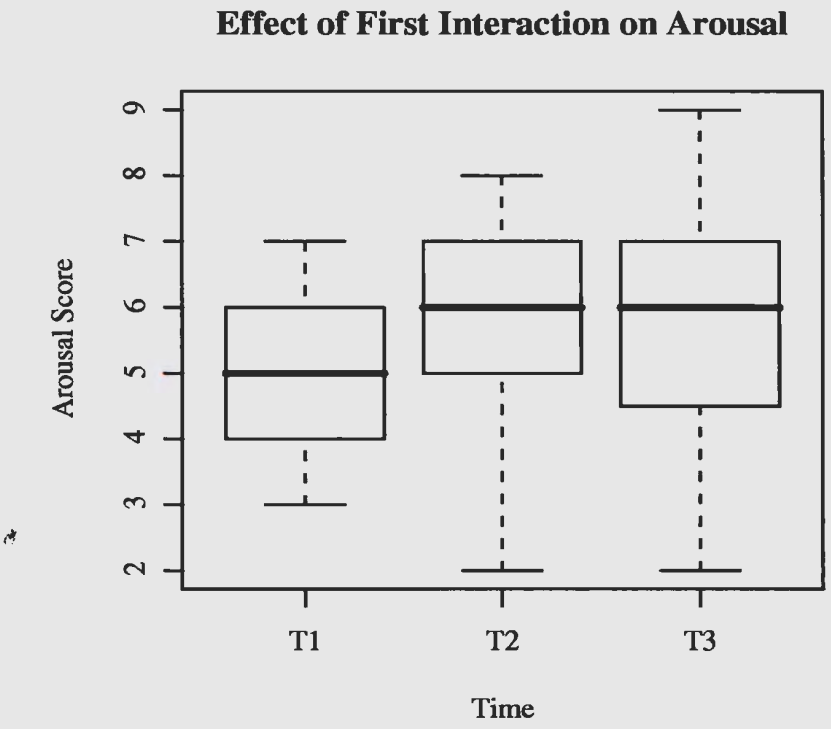

Figure 4.4: Effect of First Interaction on Arousal

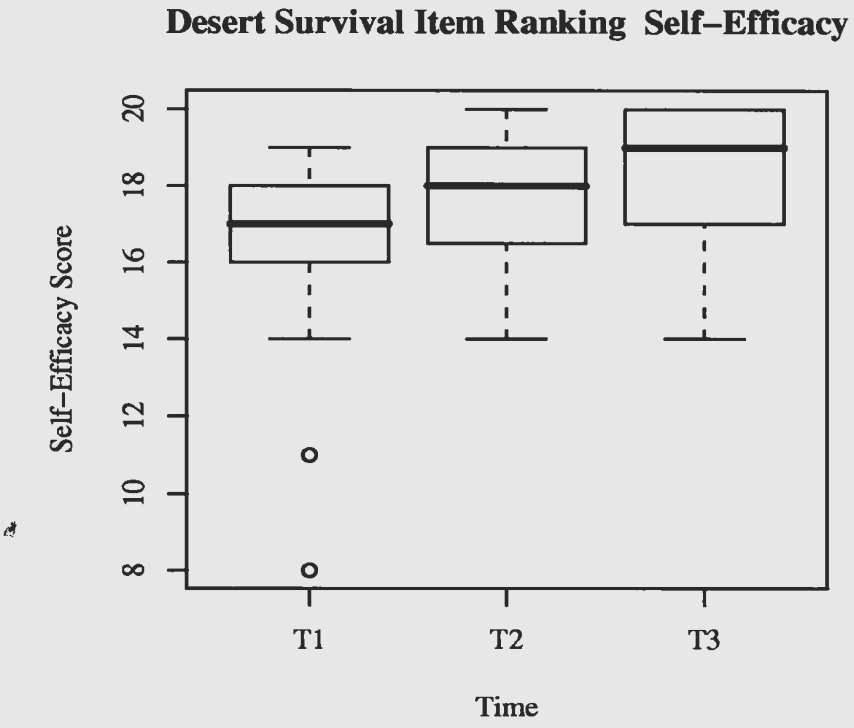

Figure 4.5: Participant Self-Efficacy Over Time

4.5.5 Additional Results

Two (unrelated) additional results that contribute to the discussion in Section 4.6 are:

- As expected (due to the established behaviour Section 4.4.2.2), the Figure 4.6 shows the positive behaviour ECAs dispensed more General and Item information compared to neutral behaviour ECAs (respectively: $R^2 = 0.75$, $F(1,30) = 92.21$, $p < 0.0001$; and $R^2 = 0.21$, $F(1,30) = 8.11$, $p < 0.008$, $pow > 0.76$).

- As seen in Figure 4.7, the stronger the participant's personal style sensor the more Item information they extracted from the single agent,$R^2 = 0.31$, $F(1,30) = 13.37$, $p < 0.00098$.

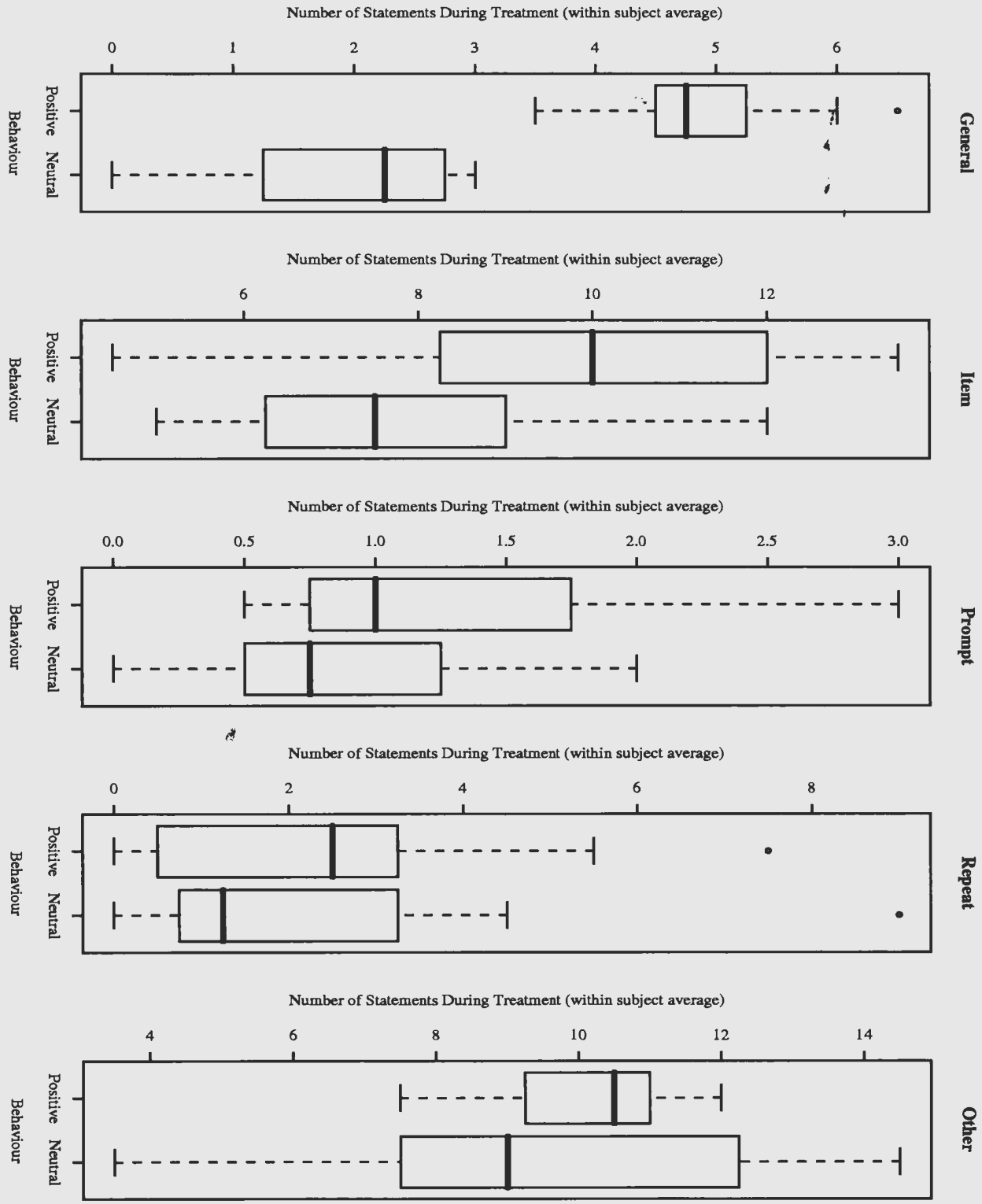

Figure 4.6: Effects of Behaviour on Conversation Style

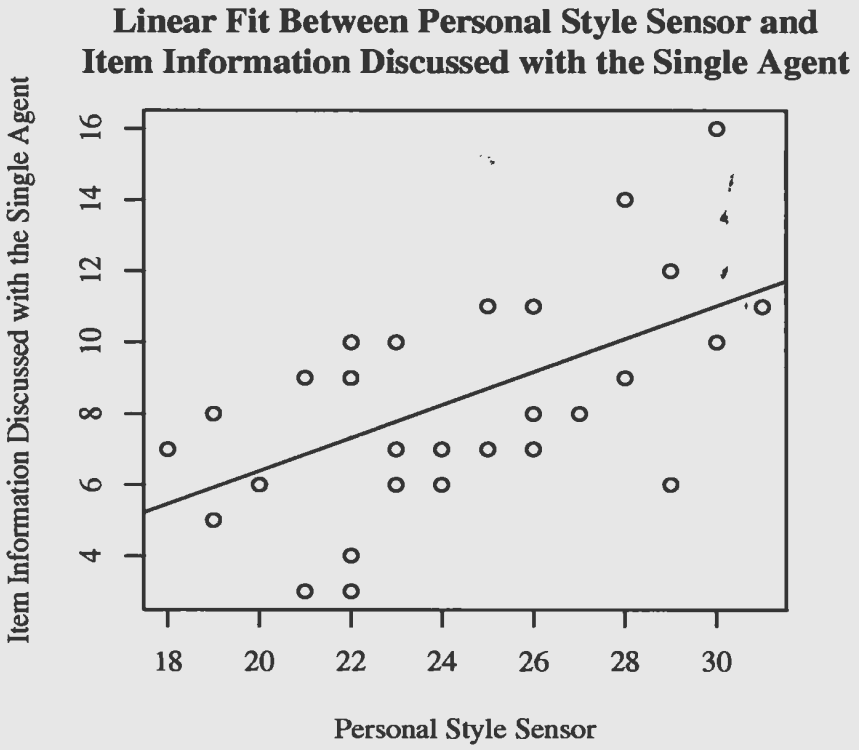

Figure 4.7: Linear Fit Between Sensor and Item Information Discussed with the Single Agent

4.5.6 Interviews

Participants reported the one user one agent interaction as more personal (intimate). This single agent group received few other comments. With an additional agent in the interface, participants focused on the inter-agent dialogue. The positive aspects of the two agent group were different points of view and the inter-agent dialogue which allowed, according to participants, greater time to think and digest the desert survival information. On the other hand, several participants found the inter-agent dialogue distracting and too spontaneous.

Several participants reported a second additional agent, in the neutral behaviour two agent group, was purposeless. Overall, participants communicated that the neutral behaviour lacked body language (e.g. gestures, smiles, eye movement) and required more personality, and that the emotionless voice required improvement. In comparison, the positive behaviour received few comments.

It became very clear during interviews that the greater the participant's (desert) survival knowledge the more poorly the ECAs were perceived to perform. Participants with survival knowledge asked pointed questions which the agents could not answer, frustrating the participants.

Two suggested improvements were: agents should communicate item importance through body language (e.g. more excited for more important items) and the embodied conversational agents should learn from past experiences (e.g. remember answers and information provided by the participant).

4.6 Discussion

This section discusses how attraction-to-group is dependent on the ECA group's behaviour and user traits (age and personal style). It also discusses valence & arousal, and self-efficacy in context of the experiment's independent variables. That is, the first three sections discuss the results in context of the variables and hypotheses presented in Chapter 4. Section 4.6.4 describes additional user responses to ECAs and ECA groups.

4.6.1 Group Cohesiveness

Group cohesion was hypothesized to be greater when interacting with two ECAs compared to one. The difference visible in Figure E.1 is not supported by statistical tests. Collecting more data might increase support for H1. On the other hand, cohesion may involve more factors than accounted for by this experiment.

During interviews, several participants thought little difference existed between the single agent and the two agent group. They questioned the purpose of the second agent. Two participants suggested a difference in personality between the two agents would result in this situation being superior than the single agent group. They said, "*...with variant personality two agents would have been better.*" and "*I definitely think two people would be more effective. They just need different personalities.*". Clearer differences between ECAs will likely improve the multi-ECA interface and its group cohesion rating.

Although clearer differences between ECAs should increase cohesive ratings, the contributions to participant response include aspects external to the interface. There is some indication that older subjects focused more on General information and preferred interacting with a second additional ECA rather than a single agent. The younger subjects focused more on the personal aspect of the conversation (Section 4.5.2.2). These quantitative results were supported by the interviews. For example, an older subject said "*two heads are better...you're going to want more than one opinion*" and a younger subject said "*one on one was more personal*". Older participants desire general information dispensed by two agents whereas younger participants desire targeted information dispensed by a personal agent (Section 4.5.2.2 and 4.5.6).

A surprisingly strong correlation exists between the participant's personal style (sensor) and the amount of Item information extracted from the single agent. Furthermore, participant's personal style (thinker-feeler) affects preference for one or two agent groups. Enhancing the linear model to include learning effect and participant age increases the statistical significance (Section 4.5.2.2). Personal style's influence on attraction-to-group remains the strongest measured effect.

The participant's age and personal style are two aspects external to the interface that may have a greater impact on group cohesion score than differences between ECAs. Even though the user-ECA interaction is complex and factors beyond the ECA's control affect group cohesion rating, intuitively two ECAs should rate as more cohesive than one; simply because a single ECA is not immediately perceived as a group. Even with these additional effects, collecting more data might produce significant results for the two ECA group over the single agent group (see table E.1).

Group cohesion was expected to be greater when interacting with positive behaviour ECAs over neutral behaviour ECAs but little support was found for this conjecture. Without collecting more data or changing the experiment's design from between-subjects to within-subjects for investigating behaviour effects, the results suggest (Section 4.5.2 and 4.5.6) tuning multi-ECA behaviour would increase their group cohesion rating.

When participants responded positively to the single agent group it appears to be due to the personal (one on one) nature of the interaction; whereas user response to the two agent group focused on the inter-agent dialogue. For example, participants discussing the single agent group said "*one on one was more personal...I had a more clear vision when I was talking with the one; she was more forward.*" and "*Talking to*

the one agent is like having a discussion. [With] two agents, I didn't get that discussion on the go (happening)." whereas representative comments on the two agent group were as follows: "*[With the two agents, they] interrupted the conversation.*"; "*The peripheral comments make the computer seem more independent*". Improving inter-agent dialogue should increase group cohesion for those participants who responded negatively to the inter-agent dialogue, with an overall effect of greater group cohesion for the positive behaviour over the neutral behaviour.

From time to time, the ECAs would prompt the user (e.g. "What do you think?"). Prompting was perceived poorly in the two agent group but not in the single agent group even though more prompting (on average) happened in the single agent group. Some participants discussed prompting during the interview with such statements as "*I did not think the 'What do you think?' was relevant.*" or "*I really did not know what to answer to the 'What do you think?'.*". Eliminating prompts should further increase the positive behaviour's group cohesion rating relative to the neutral behaviour's.

When RGCR questions involved language that suggests emotion (e.g. Q3 "I don't care what happens in one group more than the other."), participants preferred neither group more than the other (table 4.6). No other RGCR responses had such strong statistical significance (only questions 1 and 13 were also statistically significant). Comments on ECA emotion such as "*it's not emotion...the emotion is programmed in*" reflected the general belief that ECAs are emotionless. Neither group evoked an emotional connection with the participant, apparently because the participants viewed the embodied conversational agents as emotionless. It should be easier to increase group cohesion with functional changes (e.g. type of information dispensed

and manner in which the information is communicated) rather than enhancing the embodied conversational agent by modeling emotion. In other words, since users appear to reject ECAs as emotional then users may respond poorly to ECAs with emotion no matter who well the "emotion" is implemented.

4.6.2 Valence and Arousal

Participants valence seems more dependent on agent behaviour than on the number of agents (Figure E.5) and their experience using the interface (Figure 4.3). This agrees with previous research stressing the importance of behaviour (e.g. [10, 24, 37, 93]). Adding a second additional agent to the single agent group was expected to increase participant valence relative to the single agent group. This expectation relied on group cohesion being more pronounced in the two agent group than was actually observed. More decisive cohesion results would provide insight into participant valence in single and two agent groups.

The results surrounding arousal are likely explained by a "novelty effect". The first ECA-participant interaction is arousing and subsequent interactions generally maintain this level of increased arousal (Figure E.6). The second interaction lacks the originality that caused initial arousal. Conversing with one agent followed by two, although not statistically significant, maintained participant's arousal relative to the agent order 2-1. This further supports the "novelty" explanation since the agent order 1-2 adds to the novelty of the conversational interface whereas the order 2-1 removes elements from the interface (particularly in the positive behaviour where the ECAs interact).

4.6.3 Self-Efficacy

Self-efficacy increased after each participant-ECA interaction, with the second increase being less significant than the first (Figure 4.5). Some participants were unable to indicate a second increase in self-efficacy, using the instrument, since they had already recored a maximum self-efficacy score. With a modified instrument capable of capturing the second increase in self-efficacy, one may be able to measure a second increase.

Vicarious experience may explain why participant's self-efficacy changed over the course of their interactions. A vicarious experience is discussed as an individual comparing their abilities to others. In the case of the experiment, participants comparing their DSP abilities to prior DSP abilities could contribute to a vicarious desert survival experience. Information obtained from each conversation provided more context for the participant's desert-survival-problem-item ranking. The better the participant understood the DSP the stronger their belief in their abilities to properly rank the items. This would explain the results discussed in Section 4.5.4.

4.6.4 Additional User Responses

Additional responses to ECAs were collected from experimenter observation, statistical data (Section 4.5) and interviews with participants. Although these additional responses do not directly address the hypotheses, they are expected to help in ECA design and the understanding of user response to ECA interfaces.

4.6.4.1 Usability

The participant's ability to use the 2-agent interface improved when they experienced the single agent first. Participants were more content (greater valence) to experience the agent order 1-2 rather than 2-1. Examining the effect of agent order on attraction-to-group for ECAs with positive behaviour, the order 1-2 had a larger effect (father from zero) than 2-1. During the interviews, participants who experienced the agent order 2-1 made comments such as: "*...two agents, that was the problem. Did she finish or was there something next*"; "*Two agents was like a tag team.*"; "*What I liked the least is the random popping up.*". On the other hand, participants experiencing the agent order 1-2 made comments such as: "*I like the one with two people cause you get two different views.*"; "*I like it when they played off each other.*". In multi-ECA interfaces, users should most likely interact with a single embodied conversational agent before interacting with many.

4.6.4.2 Google: Analogizing to the Familiar

Rather than being conversational most users began by entering keywords. When questioned about their reluctance to write in full sentences, the participants responded with statements such as "*I assumed it would understand shorter sentences...It would be like Google - type a word and things come up.*" or "*With the computer you're use to typing short commands...the Google thing.*". Participants expected less functionality of the ECA interface and relied on previous experiences to guide their conversational behaviour. ECAs should most likely demonstrate meta-conversation competence in order to describe their conversational skills and the correct interaction techniques to

their user.

4.6.4.3 Future ECAs

During the interviews, participants were asked to consider a society with embodied conversational agents and to comment on such a society. Participants concentrated on loss of employment, ECA purpose and user-ECA relationships. This questioning was developed serendipitously and, as such, this section relates more closely to ECAs as a destination for the evolving computer interface rather than the thesis' main theme.

Participants felt ECAs were well suited as information dispensers such as information kiosks staff or for uninterrupted service positions such as tech-support. These traits were seen as positive attributes for a home security ECA - an ECA replacing the conventional home alarm system. Although participants responded positively to an ECA society, a major concern was loss of employment. Embodied conversational agents were expected to replace students in "menial" (terminology used by participants) jobs such as a cashier. However, participants were convinced that ECAs would fail in positions requiring emotions or an emotional understanding.

Participants had privacy concerns over the information collected by ECAs during user-ECA relationships since embodied conversational agents could facilitate access to people's private information. Viewing the ECA as an information collecting system paralleled the participant's expectation that an emotional link would be absent in user-ECA relationships. However, many participants feared people would become addicted to interacting with their ECAs. They expected people to lose their social capabilities and sever the human connection. In several discussions, participants voiced concerns that people could fall in love with their ECAs and abuse the technology

pornographically.

Participants seemed uneasy with the prospect of software replicating human function. In some sense asking themselves, "What purpose would humans serve if machines were human?".

4.6.5 Discussion Summary

The group's positive behaviour appears to increase user preference for the two agent group. The results suggest that user preference for an ECA group depends on factors other than the number of agents and their behaviour. Age, personal style, dialogue Prompts and information dispensed by the single agent group affect user attraction-to-group. With respect to the user's emotional state, valence increases for the agent order 1-2 and arousal increase after the first group interaction. The increase in valence is attibuted to interface usability and arousal to the "novelty effect". The increase in the user's self-efficacy over time is attributed to a vicarious experience (i.e. users comparing their abilities over the duration of the experiment).

Chapter 5

Conclusion and Future Work

This thesis set out to study ECA groups through WOZECA, a framework for simulating user-ECA group interaction, and an experiment that investigated user response to an additional second agent in an ECA-group. Section 5.1 discusses WOZECA and Section 5.1.1 suggests several improvements to the implemented framework. Section 5.2 distills the experiment and Section 5.2.1 proposes future work.

5.1 WOZECA

WOZECA simulates user-ECA group interaction using configurable video clips and a behind-the-scenes Wizard. The ECA's characteristics (see Section 2.2) are driven by a Wizard's ability to use the provided input and output mechanisms (web cam, chat program, agent movie clips) via his interface. Of these characteristics, ECA persona depends heavily on the available movie clips. Non-linear conversation style (a point at which the conversation can take multiple directions, such as two questions asked simultaneously) and users choosing the topic of conversation are two possible

weaknesses discussedin the following paragraphs.

When the user asks two or more questions at once (e.g. Is the ball green and the square blue?), it is often difficult to answer both questions in a natural manner. Since users seem to tolerate breakdown in the conversation, one mechanism for handling non-linear conversation is to ignore all but one aspect of the user's statement reestablishing conversation linearity. For example, if the user asks "*Is the ball green and the square blue?*", the embodied conversational agent answers "*The square is not blue. Maybe we can talk about the green diamonds.*".

Users will choose the topic of conversation if none is supplied. This makes a dialogue between the user and ECA more difficult to maintain. Restricting the conversation topic and a time constraint (e.g. 6 minutes) can improve the Wizard's ability to maintain a dialogue with the user. In this context, two important conversational aspects are "greetings" and "redirection" for these maintain the conversational topic within bounds that are addressable by the recorded clips. A proper greeting constrains the conversation and provides a direction for the first user-agent dialogue, and redirections are statements that allow the Wizard to change the topic when it deviates from the expected and feasible.

These weaknesses in the set of video clips and agent persona are addressed by recording additional video clips. Weaknesses related to simulation of the remaining ECA characteristics did not surface during the pilot sessions or experiment. The framework successfully simulated the experiment's one agent group and two agent group. Furthermore, WOZECA is capable of a variety of ECA experiments but may be extended (e.g. new input/output devices) to increase the scope of possible experiments.

5.1.1 Future WOZECA

WOZECA is primarily restricted by its input and output mechanisms (web cam, chat program, agent movie clips). Improvements to WOZECA's current input and output mechanisms, or additional mechanisms, should increase the variety and scope of possible experiments.

Audio would augment the Wizard's ability to understand body language and user dialogue messages. For example, audio could help differentiate between a frustrated (grunts or sighs) and happy (giggles) user. Audio would also permit the Wizard to listen to the dialogue between the experimenter and subject during an experiment (e.g. when the subject request help from the experimenter); reacting accordingly.

Using several (two or three) high-quality web cams could provide the Wizard with a more complete and detailed view of the participant. This would also extend ECA physical awareness. Future experiments could more comfortably include physical aspects of the lab environment.

Adding artificial intelligence to the Wizard interface that highlights, and possibly filters, movie clip buttons would increase the speed, accuracy and consistency of the Wizard's response. This could avoid an inappropriate response mistakenly selected by the Wizard. A more complicated mechanism, that would also improve the ECA's communicational awareness, is software that converts text to speech. The Wizard would be free to respond to users with any written statement which would, then, be spoken by the ECAs using the text to speech software. With this software, ECAs would be capable of natural dialogue (from the Wizard). This software would also address user suggestions to have the ECA's learn from past experiences (e.i. the

Wizard would be free to refer to past experiences).

From a user's perspective, the interface's fluidity and consistency needs improvement. Movie clip windows should no longer appear; rather an alive ECA should be present during the entire interaction. Moving to animated ECAs would remove the minor inconsistencies (these id not appear to affect experimental results) in lighting and actor location, relative to the window frame, captured during recording. Furthermore, animated ECAs may facilate the implementation of communicating item importance through body language.

WOZECA's weakest implementation component is Revolution due to its awkward scripting language and runtime environment. Replacing Revolution with a equivalent component programmed using Python would provide programming language consistency across every WOZECA component. This consistency is likely to increase WOZECA's maintainability by allowing programmers to focus on mastering one language rather than several and permitting the entire system to be more easily version controlled.

An immediate improvement to WOZECA would see integration of audio, use of animated video clips, and components programmed using Revolution replaced with equivalent components written in Python.

5.2 The WOZECA Experiment

Designing an experiment consisting of between and within subject measures complicated data analysis in R. For example, comparing repeated measures across both within and between subject groups involved R programming (to convert raw data to

a format expected by the built-in R functions). A design incorporating within subject measures (only) balanced by a Latin square would have simplified data analysis in R.

The ECA's chosen shirt colours and the difference in user interface window size, between the one-agent group and two-agent group, did not appear to affect user response (e.g. users did not comment on these aspects during their interview).

The *attraction-to-group* construct successfully represented user preference for an ECA group. It captured participants' preference for the one agent group even though, during the interview, several participants expressed the expectation that the two agent group more aptly suited the DSP context.

A within-subjects experiment directly comparing the positive and neutral behaviour might confirm or refute the expected link between behaviour and group cohesion. Another approach that might confirm or refute this link is collecting more data (increasing N) under the current experiment design. The selected approach would likely depend on convenience and ease of implementation. For example, a within-subjects experiment should require fewer participants and would be a better suited for cases where there is low experiment participation. Since the outcome is desired to produce a large effect (to manipulate group cohesion), it may be more appropriate to modify ECA behaviour to produce a large effect rather than increasing N to identify small effects.

To confirm or refute the link between the number of agents and group cohesion, changes in the current experiment are suggested. Based on interview responses, the agents in the two agent group should differ in order to produce a strong response from experiment participants. For example, ECAs could differ by having conflicting DSP item rakings and associated reasoning.

These experiments should continue to measure the user's emotional state (valence and arousal) due to the insight into user response to ECA groups that this measure has provided. However, it is suggested that these experiments no longer measure self-efficacy, as they are not expected to affect user self-efficacy given self-efficacy's dependence on time. Instead, experiments could be specifically designed to better comprehend user self-efficacy and ECA groups.

Within the current experiment, ECA behaviour influences user attraction-to-group and valence more than the number of agents, whereas arousal and self-efficacy depend on time. Within the positive ECA behaviour, the user's attraction-to-group is determined by their age and personal style. Generally, older users and users that scored higher on the feeler-thinker personal style scale preferred two agents. With two agents in the interface, participants focused on the agent-agent interaction. The results suggest that these participants were judging the ECA's social abilities and that appropriate ECA behaviour within the two agent group likely differs from the single agent group (e.g. prompting).

User's personal style had a much stronger impact on group cohesion and user response than expected. Not only could ECA interfaces accommodate personal style, computer systems interfaces may benefit from considering personal style and personality during design. For example, operating system users could complete a personal style questionnaire and as a result their initial default interface (e.g. colours, icons, style, button positions) would account for their personal style.

Users approach the ECA interface with the preconception that ECAs are emotionless search tools. Although this notion is easily extinguished, it leads to the thought that ECAs should be capable of meta-conversation to establish ECA capabilities.

Meta-conversation is expected to become particularly relevant when users routinely encounter ECAs with varying degrees of capability.

Before users interact with a multi-ECA interface, users could interact with a single ECA to establish an understanding of the ECA's abilities (e.g. gestural, intelligence, conversational). Once the user understands how to interact with a single ECA, a second is much more likely to improve the interface rather than upset the user.

Users hold preconceived notions of future ECAs' role in society. By and large, participants worried that future ECAs would weaken human social connections (to the extreme of pornographic addiction to ECAs) as well as reduce employment opportunities. Embodied conversational agents may need to overcome users' preconceived notions to increase their effectiveness.

Given an understanding of ECAs and ECA groups acquired throughout this thesis, the following are suggestions for the ECA and ECA group interface:

- Use an ECA-group in an information dispensing context
- Provide ECAs with meta-conversational abilities to help users understand the ECA interface's capabilities
- Introduce the user to a single ECA prior to an ECA group
- ECA group members should differ in appearance and behaviour
- ECA groups and their members' appearance should be consistent with their behaviour
- Concentrate on perfecting the ECA and ECA group's behaviour

- Allow the user to return to their original ECA group

5.2.1 Future Experiments

Although embodied conversational agents are becoming more sophisticated and natural, their value depends on discovering situations in which the ECAs and their groups are effective for users. Future work should refine and discover the characteristics that affect user response within ECA groups. This subsection suggests several avenues for refining these characteristics. They are: delimiting ECA meta-conversation requirements; the effects of personality and personal style on user response to ECA groups; and user response to embodied conversational agents with conflicting views.

An embodied conversational agent's meta-conversational abilities are anticipated to be important, for conveying the purpose and abilities of the ECA interface. The meta-conversational questions posed by users may fall outside of conventional questioning. For example, users may be interested in the amount of personal information stored by the ECA, as well as who has access to this information and the ECAs physical components. Consequently, meta-conversational ECAs will likely require a model of their self that may include details beyond those encountered during normal human conversation. Perhaps a more difficult problem will be how an ECA conveys this *self* information using terminology appropriate to each user. Future research could develop an understanding of the meta-conversational boundaries and build the components required to implement ECA meta-conversational abilities.

A more elaborate experiment concentrating on the effect of personal style on user response could clarify the impact and breadth of a user's personal style on their ECA

interface. One possibility is adopting an approach employed by Reeves and Nass [70] - locate a successful psychology experiment related to personal style and groups, and replicate it using ECAs as group members. Assuming users respond to ECAs in accordance to the existing knowledge on personal style, ECAs should benefit from the ability to measure their user's personal style. In other words, ECAs could use small-talk to perform a real-time personal style assessment.

It would be interesting to discover user response to ECAs with conflicting views and ECAs arguing amongst themselves. An immediate application of argumentative ECAs could be electronic games. Under this context, group cohesion may not be an appropriate instrument for measuring user response to argumentative ECAs given the objective of these ECAs may be to entertain (e.g. "evil" game characters arguing with game "heroes").

A possible experiment would begin to establish when users strongly prefer a two agent group over a single agent group. This experiment would be similar to the experiment described in Chapter 4. It would be a balanced within subjects design with one factor (the number of agents) having two levels (one agent, two agents), where the two agents differ in persona.

5.3 Final Thought

Adding an additional second agent to a single agent group appears to modify user response. As shown by Morishima et al. (Section 2.1), this second ECA should be beneficial, under certain conditions, to ECA interface users. The thesis results suggest these conditions include ECA groups with group members differing in appearance

and behaviour. By simulating ECAs and their characteristics, WOZECA will help establish which ECA differences are beneficial to ECA interfaces.

Bibliography

[1] Autotutor. Wold Wide Web, December 2005. www.autotutor.org.

[2] Fascinating haptek history. Wold Wide Web, December 2005. www.haptek.com.

[3] Meet tank. World Wide Web, January 2006. http://www.roboceptionist.com/meet_tank.htm.

[4] J. M. Allbeck and N. I. Badler. Creating embodied agents with cultural context. In S. Payr and R. Trappl, editors, *Agent Culture: Human-Agent Interaction in a Multicultural World.* Lawrence Erlbaum Associates, 2004.

[5] C. F. Auerbach and L. B. Silverstein. *Qualitative data : an introduction to coding and analysis.* New York : New York University Press, 2003.

[6] B. P. Bailey and J. A. Konstan. Are informal tools better?: comparing demais, pencil and paper, and authorware for early multimedia design. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 313–320, 2003.

[7] A. Bandura. *Self-Efficacy : The Excersice of Control.* W. H. Freeman and Company, 1997.

[8] T. Bickmore. Unspoken rules of spoken interaction. *Communications of the ACM*, 47(4):38–44, 2004.

[9] T. Bickmore and J. Cassell. Relational agents: A model and implementation of building user trust. In *The SIGCHI conference on Human factors in computing systems*, pages 396–403, 2001.

[10] T. Bickmore and R. Picard. Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction*, 12(2):293–327, June 2005.

[11] J. A. Bonito, J. K. Burgoon, and B. Bengtsson. The role of expectations in human-computer interaction. In *Proceedings of the international ACM SIGGROUP conference on Supporting group work*, pages 229–238, 1999.

[12] M. M. Bradley and P. J. Lang. Measuring emotion: The self-assessment manikin and the semantic differential. *Behaviour Therapy and Experimental Psychiatry*, 25(1):49–59, March 1994.

[13] E. Bradner and G. Mark. Why distance matters: effects on cooperation, persuasion and deception. In *Proceedings of the ACM conference on Computer supported cooperative work*, pages 226–235, 2002.

[14] S. Brave, C. Nass, and K. Hutchinson. Computers that care: Investigating the effects of orientation of emotion exhibited by an embodied conversational agent. *International Journal of Human-Computer Studies*, 62(2):161–178, 2005.

[15] C. Breazeal. Socially intelligent robots. *ACM Interactions*, 12(2):19–22, 2005.

[16] C. Breazeal, G. Hoffman, and A. Lockerd. Teaching and working with robots as a collaboration. In *The Third International Joint Conference on Autonomous Agents and Multiagent Systems*, volume 3, pages 1030–1037, 2004.

[17] A. v. Breemen, X. Yan, and B. Meerbeek. icat: An animated user-interface robot with personality. In *The fourth international joint conference on Autonomous agents and multiagent systems*, pages 143–144, July 2005.

[18] J. Burgoon, J. Bonito, B. Bengtsson, C. Cederberg, M. Lundeberg, and L. Allspach. Interactivity in human-computer interaction: a study of credibility, understanding, and influence. *Computers in Human Behavior*, 16:553–574, 2000.

[19] J. Cassell. Embodied conversation interface agents. *Communications of the ACM*, 43(4):70–78, April 2000.

[20] J. Cassell. Embodied conversational agents: Representation and intelligence in user interfaces. *AI Magazine*, 22(4):70–71, 2001.

[21] J. Cassell, C. Pelachaud, N. Badler, M. Steedman, B. Achorn, T. Becket, B. Douville, S. Prevost, and M. Stone. Animated conversation: rule-based generation of facial expression, gesture & spoken intonation for multiple conversational agents. In *Proceedings of the 21st annual conference on Computer graphics and interactive techniques*, pages 413–420, 1994.

[22] J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, editors. *Embodied Conversational Agents*, pages 374–402. MIT Press, 2000.

[23] A. Cavalluzzi, G. Clarizio, B. N. De Carolis, and F. de Rosis. A persona is not a person: Designing dialogs with ecas after wizard of oz studies. In *HUMAINE WP6 Workshop on Interaction and Communication*, March 2005.

[24] K. Dautenhahn. Socially intelligent agents in human primate culture. In S. Payr and R. Trappl, editors, *Agent Culture: Human-Agent Interaction in a Multicultural World.* Lawrence Erlbaum Associates, 2004.

[25] A. De Angeli, W. Gerbino, E. Nodari, and D. Petrelli. From tools to friends: Where is the borderline? In *Proceedings of the UM99 Workshop on Attitude, Personality and Emotions in User-Adapted Interaction*, pages 1–10, June 1999.

[26] F. de Rosis, C. Pelachaud, and I. Poggi. Transcultural believability in embodied agents: A matter of consistent adaptation. In S. Payr and R. Trappl, editors, *Agent Culture (Human-Agent Interaction in a Multicultural World).* Lawrence Erlbaum Associates, 2004.

[27] R. Dillmann, R. Becher, and P. Steinhaus. Armar ii a learning and cooperative multimodal humanoid robot system. *International Journal of Humanoid Robotics*, 1(1):143–155, 2004.

[28] R. Gockley, A. Bruce, J. Forlizzi, M. Michalowski, A. Mundell, S. Rosenthal, B. Sellner, R. Simmons, K. Snipes, A. C. Schultz, and J. Wang. Designing robots for long-term social interaction. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2166–2204, 2005.

[29] M. Goldstein, G. Alsio, and J. Werdenhoff. The media equation does not always apply: People are not polite towards small computers. *Personal and Ubiquitous Computing*, 6:87–96, 2002.

[30] A. C. Graesser, A. Olney, B. C. Haynes, and P. Chipman. Autotutor: A cognitive system that simulates a tutor that facilitates learning through mixed-initiative dialogue. In C. Forsythe, M. L. Bernard, and T. E. Goldsmith, editors, *Cognitive systems: Human cognitive models in systems design.* Erlbaum, 2005.

[31] M. A. Hogg. *The Social Psychology of Group Cohesiveness.* New York University Press, 1992.

[32] Houghton Mifflin Company, Boston, Ma. *The American Heritage Dictionary of the English Language*, 4 edition, 2004.

[33] W. Ju, S. Nickell, K. Eng, and C. Nass. Influence of colearner agent behavior on learner performance and attitudes. In *CHI '05: CHI '05 extended abstracts on Human factors in computing systems*, pages 1509–1512. ACM Press, 2005.

[34] C. Jung and S. Parry. *Personal Style Assessment Jung-Parry Form.* HRD Press, Inc., 2000.

[35] J. H. Kerr. *Motivation and Emotion in Sport*, page 90. Psychology Press, 1997.

[36] S. R. Klemmer, A. K. Sinha, J. Chen, J. A. Landay, N. Aboobaker, and A. Wang. Suede: A wizard of oz prototyping tool for speech user interfaces. In *Proceedings of the 13th annual ACM symposium on User interface software and technology*, pages 1–10, 2000.

[37] T. Koda. User reactions to anthropomorphized interfaces. *IEICE Transactions on Information and Systems*, E86-D(8):1369–1377, 2003.

[38] S. Kopp, P. Tepper, and J. Cassell. Towards integrated microplanning of language and iconic gesture for multimodal output. In *Proceedings of the 6th international conference on Multimodal interfaces*, pages 97–104, October 2004.

[39] L. Kwak, S. Kremers, A. Walsh, and H. Brug. How is your walking group running? *Health Education*, 106(1):21–31, 2006.

[40] J. C. Lafferty, P. M. Eady, and J. Ehners. *The desert survival problem.* Plymouth, Michigan: Experimental Learning Methods, 1974.

[41] B. Laurel, editor. *The art of human-computer interface design.* Addison-Wesley, 1990.

[42] B. Laurel. *Computers As Theatre*, page xi. Addison Wesley Longman, Inc., 1991.

[43] E.-J. Lee and C. Nass. Does the ethnicity of a computer agent matter? an experimental comparison of human-computer interaction and computer-mediated communication. In *Proceedings of the Workshop on Embedded Conversational Characters Conference*, pages 123–128, 1998.

[44] E.-J. Lee and C. Nass. An experimental test of normative group influence and representation effects in computer-mediated communication. *Human Communication Research*, 28(3):349–381, 2002.

[45] K. M. Lee and C. Nass. Designing social presence of social actors in human computer interaction. In *Proceedings of the conference on Human factors in computing systems*, volume 5, pages 289–296, April 2003.

[46] J. C. R. Licklider. Man-computer symbiosis. In *IRE Transactions on Human Factors in Electronics*, volume HFE-1, pages 4–11, March 1960.

[47] H. Maldonado and B. Hayes-Roth. Toward cross-cultural believability in character design. In S. Payr and R. Trappl, editors, *Agent Culture (Human-Agent Interaction in a Multicultural World).* Lawrence Erlbaum Associates, 2004.

[48] J. Martin, S. Buisine, and S. Abrilian. 2d gestural and multimodal behavior of users interacting with embodied agents. In *Proceedings of the workshop on Balanced Perception and Action in ECAs*, New York, USA, March 2004. AAMAS.

[49] P. Metzner. *Crescendo of the virtuoso : spectacle, skill, and self-promotion in Paris during the Age of Revolution.* Berkeley: University of California Press, 1998.

[50] M. Minsky. *The Society of Mind.* Simon and Schuster, 1985.

[51] Y. Moon. The effects of distance in local versus remote human-computer interaction. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 103–108, April 1998.

[52] L.-P. Morency, C. Sidner, C. Lee, and T. Darrell. Contextual recognition of head gestures. In *The 7th international conference on Multimodal interfaces*, pages 18–24, October 2005.

[53] Y. Morishima, H. Nakajima, S. Brave, R. Yamada, H. Maldonado, C. Nass, and S. Kawaji. The role of affect and sociality in the agent-based collaborative learning system. In *Affective Dialogue Systems: Tutorial and Research Workshop, ADS 2004*, volume 2068, pages 265–275. Springer-Verlag GmbH, 2004.

[54] J. Morkes, H. K. Kernal, and C. Nass. Humor in task-oriented computer-mediated communication and human-computer interaction. In *Proceedings of the conference on Human Factors in Computing Systems*, pages 215–216, 1998. (conference summary).

[55] B. A. Myers. A brief history of human computer interaction technology. *ACM Interactions*, 5(2):44–54, March 1998.

[56] H. Nakajima, S. Brave, C. Nass, R. Yamada, Y. Morishima, and S. Kawaji. The functionality of human-machine collaboration systems mind model and social behavior. In *Proceedings of the IEEE Conference on Systems, Man, and Cybernetics*, volume 3, pages 2381–2387, October 2003.

[57] Y. I. Nakano, G. Reinstein, T. Stocky, and J. Cassell. Towards a model of face-to-face grounding. In *The 41st Annual Meeting on Association for Computational Linguistics*, volume 1, pages 553–561, 2003.

[58] C. Nass, E.-Y. Kim, and E.-J. Lee. When my face is the interface: an experimental comparison of interacting with one's own face or someone else's face. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 148–154, June 1998.

[59] C. Nass and K. M. Lee. Does computer-generated speech manifest personality? an experimental test of similarity-attraction. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 329–336, 2000.

[60] C. Nass, Y. Moon, B. J. Fogg, B. Reeves, and C. Dryer. Can computer personalities be human personalities? In *Conference companion on Human factors in computing systems*, pages 228–229, 1995.

[61] C. Nass, J. Steuer, and E. R. Tauber. Computers are social actors. In *Proceedings of the conference on Human factors in computing systems*, pages 72–78, 1994.

[62] C. Nass, J. Steuer, E. R. Tauber, and H. Reeder. Anthropomorphism, agency, & ethopoeia: Computers as social actors. In *Proceedings of the conference on Human factors in computing systems*, pages 111–112, 1993. (conference companion).

[63] J. Neter, W. Wasserman, and G. A. Whitmore. *Applied Statistics.* Allyn and Bacon, 4 edition, 1993.

[64] J. Nielson. *Usability Engineering.* Morgan Kaufmann, 1993.

[65] N. Ole Bernsen and L. Dybkjær. Evaluation of spoken multimodal conversation. In *Proceedings of the 6th international conference on Multimodal interfaces*, pages 38–45, 2004.

[66] F. Pajares. Self-efficacy beliefs in academic settings. *Review of Educational Research*, 66(4):543–578, 1996.

[67] S. Pasquariello and C. Pelachaud. Greta: A simple facial animation engine. In *6th Online World Conference on Soft Computing in Industrial Applications*, September 2001. Session on Soft Computing for Intelligent 3D Agents.

[68] C. Pelachaud. Multimodal expressive embodied conversational agents. In *The 13th annual ACM international conference on Multimedia*, pages 683–689, November 2005.

[69] H. Prendinger and M. Ishizuka. Social role awareness in animated agents. In *Proceedings of the fifth international conference on Autonomous agents*, pages 270–277, 2001.

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

[71] M. Rehm, E. André, and M. Wissner. Gamble v2.0 - social interactions with multiple users. In *The fourth international joint conference on Autonomous agents and multiagent systems*, pages 145–146, July 2005.

[72] R. Rickenberg and B. Reeves. The effects of animated characters on anxiety, task performance, and evaluations of user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 49–56, 2000.

[73] D. Riecken, P. Maes, B. Shneiderman, and D. C. Smith. Dealing with complexity: uniting agents and direct manipulation. In *Conference companion on Human factors in computing systems*, pages 191–192, 1995.

[74] R. Schwarzer and U. Scholz. Cross-cultural assessment of coping resources: The general perceived self-efficacy scale. In *First Asian Congress of Health Psychology: Health Psychology and Culture*, August 2000. http://web.fu-berlin.de/gesund/publicat/world_data.htm.

[75] M. E. Shaw. *Group Dynamics: The Psychology of Small Group Behaviour.* McGraw-Hill, 1981.

[76] N. Shechtman and L. M. Horowitz. Media inequality in conversation: How people behave differently when interacting with computers and people. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, volume 5, pages 281–288, April 2003.

[77] C. Sidner and M. Dzikovska. Hosting activities: Experience with and future directions for a robot agent host. In *The International Conference on Intelligent User Interfaces*, pages 143–150, January 2002.

[78] A. K. Sinha and J. A. Landay. Capturing user tests in a multimodal, multidevice informal prototyping tool. In *Proceedings of the 5th international conference on Multimodal interfaces*, pages 117–124, 2003.

[79] M. Solomon, N. Pistrang, and C. Barker. The benefits of mutual support groups for parents of children with disabilities. *American Journal of Community Psychology*, 29(1):113–132, 2001.

[80] T. Stocky and J. Cassell. Shared reality: Spatial intelligence in intuitive user interfaces. In *The International Conference on Intelligent User Interfaces*, pages 224–225, January 2002.

[81] B. Stronks, A. Nijholt, P. van der Vet, and D. Heylen. Designing for friendship: Becoming friends with your eca. In *Embodied conversational agents - let's specify and evaluate them!*, July 2002. In conjunction with The First International Joint Conference on Autonomous Agents and Multi-Agent Systems.

[82] W. Swartout, J. Gratch, R. Hill, E. Hovy, S. Marsella, J. Rickel, and D. Traum. Toward virtual humans. In *working notes of the AAAI Fall symposium on Achieving Human-Level Intelligence through Integrated Systems and Research*, October 2004. Download from section publications of www.ict.usc.edu.

[83] Y. Takeuchi and Y. Katagiri. Identity perception of computers as social actors. In *Second International Conference on Cognitive Science*, July 1999.

[84] K. Thórisson. *Communicative Humanoids: A Computational Model of Psychosocial Dialogue Skills.* PhD thesis, Massachusetts Institute of Technology, September 1996.

[85] K. Thórisson. Gandalf: An embodied humanoid capable of real-time multimodal dialogue with people. In *The first international conference on Autonomous agents*, pages 536–537, 1997.

[86] K. Thórisson, C. Pennock, T. List, and J. DiPirro. Artificial intelligence in computer graphics: a constructionist approach. *ACM SIGGRAPH Computer Graphics archive*, 38(1):26–30, February 2004.

[87] D. Traum and J. Rickel. Embodied agents for multi-party dialogue in immersive virtual worlds. In *The first international joint conference on Autonomous agents and multiagent systems: part 2*, pages 766–773, July 2002.

[88] US Army, Guilford, CT. *U,S, Army Survival Guide*, 2002.

[89] Wackerly, Mendenhall, and Scheaffer. *Mathematical Statistics with Applications.* Wadsworth Publishing Company, 5 edition, 1996.

[90] T. Winograd and F. Flores. *Understanding Computers and Cognition.* Addison-Wesley, 1986.

[91] D. Wonisch and G. Cooper. Interface agents: preferred appearance characteristics based upon context. In *Proceedings of the workshop Virtual Conversational Characters: Applications, Methods, and Research Challenges*, November 2002.

[92] M. Wooldridge. *An Introduction to MultiAgent Systems.* Cambridge University Press, 2002.

[93] J. Xiao, J. Stasko, and R. Catrambone. An empirical study of the effect of agent competence on user performance and perception. In *The third international joint conference on Autonomous agents and multiagent systems*, pages 178–185, July 2004.

[94] X. Yuan and Y. S. Chee. Embodied tour guide in an interactive virtual art gallery. In *The International Conference on Cyberworlds*, pages 432–440, 2003.

[95] J. Zheng, X. Yuan, and Y. S. Chee. Designing multiparty interaction support in elva, an embodied tour guide. In *The fourth international joint conference on Autonomous agents and multiagent systems*, pages 929–936, 2005.

Appendix A

Brief History of ECAs

Throughout the ages, humankind has struggled to understand and artificially replicate characteristics and abilities such as the articulation of members, personality or the ability to solve puzzles.

In the eighteenth century, organicist automaton makers duplicated human appearance and processes using machines to discover the similarities and differences between themselves and the machines. The nineteenth century saw the creation of automatons for the purpose of entertainment rather than explanation [20]. A well know automaton is the Flûteur (Flute Player) created by mechanician Jacques Vaucanson (1709-1782) in 1738. He unveiled the Flute Player to citizens of Paris, and amazed them, when the wood structure filled with cords, axles, levers, pulleys, chains, pipes, valves and bellows played perfectly harmonized music [49, page 162].

Artists share this self-fascination and have depicted humanoid life-like robots in a multitude of media over many years. In 1920, the theatrical piece R.U.R (Rossum's Universal Robots) by Karel Capêk introduced the world to the word "robot", the

Czech word for "worker" [84]. HAL-9000, of "Kubrick's 2001: A Space Odyssey", was a notably malicious life-like humanoid; life-like in soul rather than appearance. Robots are also depicted as helpful and intelligent, such as Star War's C3PO. People of varying backgrounds across many disciplines have molded a similar idea - life-like robots - from the same creative clay.

In 1950, Alan Turing proposed the Turing test which measures a computer's competence at human-like conversation and, if passed, is a mark of an intelligent machine. From the time of the Turing test, various humanoid components have come to exist either in the form of an idea, an algorithm or a physical component. Licklider realized, possibly not the first, that more effective computers would converse and think at a human level [46]. He discussed such things as speech-recognition and talking to machines. Shortly after, the discipline of Human-Computer Interaction imagined and realized trainable gesture recognition (Teitelman in 1964) and augmented reality (Ivan Sutherland from 1965 to 1968) [55]. Eventually, Minsky applied and popularized[1] the term agent with his book "The Society of Minds" [50, pages 17-37]. Minsky's agents are atomic elements of intellect; existing alone they are unintelligent but connected, Minsky believes, they are truly intelligent. Such simple beginnings have given rise to a substantially more complex agent. The human puzzle is slowly being assembled from its pieces.

People such as writers, artists, film makers and scientists have dreamed of the embodied conversational agent housed in a physically-real robotic body. However, only a primitive realization of the latter exists; as Bickmore notes, work with ECAs has only begun [8].

[1]From my research and according to Nakajima et al. [56].

Appendix B

Detailed Procedure

Below is the procedure followed during the experiment with material read to the participants in quotes (e.g. "*Hello*"). A minor variation on this procedure exists where the participant interacts with one ECA followed by two ECAs (i.e. interchange one and two, and correct the English accordingly).

1. Pre-treatement
    - The participant's computer speakers are turned on to the correct volume (sound check).
    - The screen saver is off.
    - Assign a number (#C2-1 or #NC2-1) to the participant and record the number on all questionnaires.
    - The server is started.
    - The Chat window is properly centered on the participant's computer and is the correct size.
    - The background is set to 2 agents.
    - Sign voucher; Record name and MUN #;

    "*I will be reading the instructions to you so that I do not forget something. The goal of this experiment is to understand what is required to build successful virtual people, or agents. A person that exists on a computer screen that is controlled by a computer is a virtual person, or an agent. I am with you during*

the entire experiment which will take about one and a half hours. I'll give you the questionnaires at the appropriate times and teach you how to talk to the agents. I may take some notes during your conversation with the agents. Before we start, I'll need your consent."

2. Participant reads consent form; sign.

3. Ask participant: Do you have any questions?

4. Scenario: "*You will complete the Desert Survival Problem. The Desert Survival Problem requires that you rank eight items from most important to least important for survival in the desert. Once you have answered two questionnaires and ranked the Desert Survival Problem items, you will interact with agents on the computer. The agents will share their desert survival knowledge during a six minute period. After the six minutes have elapsed, you will again answer some questions and rank the Desert Survival Problem items; and so on...*"

   Ask participant: Do you have any questions?

5. "*The long term goal is to build future agents that work well with you, and work well for you. In order to do this, an understanding of who you are is required. The next questionnaire will help make agents more sensitive to individual differences.*"

   Participant completes: data collection 1

6. "*Imagine that you crashed-landed a helicopter in the middle of the desert and that you were able to salvage items from the body of the helicopter. You know that search and rescue teams are looking for you but you do not know whether they will find you in hours, days or weeks. The Desert Survival Problem requires that you rank eight items from most important to least important for survival in the desert.*"

   Participant completes: DSP ranking labeled "before"

7. "*Again, the long term goal is to build future agents that work well with you, and work well for you. The next questionnaire will help make agents more sensitive to individual differences.*"

   Participant completes: data collection 3 labeled "before"

8. "*This next questionnaire will help capture how the agents affect you.*"

   Participant completes: Valence, Arousal Images labeled "before"

9. Experimenter teaches participant how to use Chat tool

   "*It's time to sit in front of the computer. I'm going to teach you how to use the chat tool before you have a conversation with the agents. To Chat with the*

agents about desert survival you will have to use the Chat tool. You can see it on the screen (experimenter points). It works similar to MSN or IRC; if you know what they are. You must type the messages you wish to send and when you are ready to send your message you must hit the enter key. Let's send a message together. Type "hi". You can see your message here (experimenter points). Now hit the enter key. You have sent "hi" to the agents. This is the message (experimenter points) history; the messages you send appear in the message history."

10. Give Memory Aid: "Desert Survival Items" to participant.

    "This should help you during your conversation with the agents."

11. Instructions just prior to interaction

    "Remember, the goal of this experiment is to understand what is required to build successful virtual people, or agents. The agents will share their desert survival knowledge during a six minute period. Don't forget that the Desert Survival Problem requires that you rank eight items (point to the memory aid) from most important to least important for survival in the desert. After your six minute conversation you will complete four more questionnaires. The agents understand more than the 8 desert survival item words. The agents will tell you when they do not understand something. When you are ready to talk to the agents type "hello" and press enter."

    6 MINUTE INTERACTION

    Agent terminates conversation

12. *"There are three questionnaires that I would like you to complete. The first questionnaire captures your reaction; please complete it now."*

    Participant completes: Valence, Arousal Images labeled "middle"

13. *"The Desert Survival Problem requires that you re-rank the eight items from most important to least important for survival in the desert."*

    Participant completes: DSP ranking labeled "middle"

14. Change background to 1 agent

15. *"Again, the long term goal is to build future agents that work well with you, and work well for you. The next questionnaire will help make future agents more sensitive to individual differences."*

    Participant completes: data collection 3 labeled "middle"

16. *"Instructions just prior to interaction: It's time to sit in front of the computer, again. The agents will share their desert survival knowledge during a six minute*

period. After your six minute conversation you will complete four more questionnaires. When you are ready to talk to the agent type "hello" and press enter."

6 MINUTE INTERACTION

Agent terminates conversation

17. *"I would like capture your reaction."*

    Participant completes: Valence, Arousal Images labeled "after"

18. *"Could you re-rank the eight items from most important to least important."*

    Participant completes: DSP ranking labeled "after"

19. *"This will help make future agents more sensitive to individual differences."*

    Participant completes: data collection 3 labeled "after"

20. *"This questionnaire should help determine which situation you liked best."*

    Participant completes: data collection 5

21. Post-Experiment Interview

    "At this point, I would like to conduct an interview to better understand your perspective on this experience. After the interview there is one more questionnaire and then we're done."

22. Start digital audio recording.

    Some interview questions:

    - What did you think of the experience?
    - How much do you normally use a computer?
    - Which situation/interaction (with the ECAs) did you prefer the most, and why?
    - Which situation/interaction (with the ECAs) did you prefer the least, and why?
    - Did you feel like a team?
    - Did the agents feel like a real person or a computer generated person?
    - Did you feel that you had a relationship with the agents?
    - Do you think a virtual person could affect your life? make you happy or sad? make you stop smoking, exercise more or something similar?
    - Imagine a future society with ECAs, ECAs far more advanced than those you experienced. Could you comment on such a society.

23. "*Again, the next questionnaire will help make future agents more sensitive to individual differences. Please read the instructions carefully before completing the questionnaire.*"

    Participant completes: data collection 2 [use booklet for questions [34]; starts on page 3]

Appendix C

Script and Coding

The video clip statements available to the embodied conversational agents are presented below. The desert survival information is drawn from the US Army Survival Guide [88]. The coding of these statements can be found in square brackets. For example, the category water was coded as item information.

Greeting [other]

- Hello, my name is Alpha. I am a computer generated personality and I use small movies to talk with you. I have been taught some basic desert survival. I will be able to give you information and insight into the items that you must rank. What item would you like to discuss first?
- My name is Alpha.
- bye
- I'm sorry but we've run out of time, goodbye.

(Two Agent Greeting) [other]

- Hello, my name is Alpha. We are computer generated personalities and we use small movies to talk with you.
- And my name is Beta. We have been taught some basic desert survival. We will be able to give you information and insight into the items that you must rank. What item would you like to discuss first?

Positive [other]

- yes
- I agree.
- Definitely true.

Negative [other]

- no
- I disagree.
- Definitely not.

Neutral [other]

- maybe
- I'm not sure.
- Perhaps yes, perhaps no.
- I don't know.

Repeat [repeat]

- I'm not sure I understood, could you rephrase that.
- That was confusing could you say that again.

Redirect [other]

- This is an unfamiliar topic area. We should focus on desert survival.
- Consider this...
- Which item should we discuss next?
- What do you think? [coded as *prompt*, not other]
- Did you know that...
- Continue...

Water [item information]

- the one 2L bottle of water
- You drink the water.

- You can use the empty 2L bottle to trap condensation at dawn, providing more water.
- I would rank the bottle of water as the most important item.

Canvas [item information]

- the 6 meter by 6 meter piece of blue canvas
- The canvas can provide shelter from the sun. It is difficult to build a shelter in desert environments so you must have one with you.
- The canvas can keep you warm at night.
- Using the canvas, you might be able to trap condensation at dawn, providing more water.
- I would rank the canvas as the second most important item.

Knife [item information]

- the knife
- The knife is necessary if you wish to build anything to increase your chance of survival.
- The knife can given you a feeling of security.
- I would rank the knife third.

Mirror [item information]

- the piece of a mirror
- The reflection of sunlight off of the mirror can alert the search and rescue teams of your whereabouts.
- You can use the mirror to inspect parts of your body, like your back, that are not easy to see. This can be useful if you fall and scrape your back, or worse.
- I would rank the mirror fourth.

Flashlight [item information]

- the flashlight
- At dusk and at night light travels great distances, the flashlight can alert the search and rescue teams of your whereabouts.

- The flashlight can help you see at night making nighttime experiences less frightening.
- I would rank the flashlight fifth.

Map [item information]

- the map
- Using the map you could move to a nearby location that would increase your chances of survival, like a small cave.
- You need the compass if you plan to use the map.
- I would rank the map sixth.

Compass [item information]

- the magnetic compass
- If you have map reading skills the compass allows you to orient the map
- properly.
- I would rank the compass before last.

Jacket [item information]

- the jacket
- The jacket might be useful during a couple of hours at night when the desert cools.
- The jacket could provide shelter from the sun during the day by holding it above your head.
- I would rank the jacket as the least important item.

General Information [general information]

- Deserts can be cold during the night and hot during the day. Most deserts are around 10 degrees C at night and up to 60 degrees C during the day.
- In the desert, the shade can be as much as 20 degrees C cooler than in the sun. Being in the shade helps you conserve water by not sweating as much as in the sun.
- People sitting around in 45 degree C heat can require up to 15 liters of water.

- People need at least two liters of water per day and deserts only receive 10 centimeters of rainfall annually water is important.
- Food digestion requires water so eating food increases your need for water.
- Deserts have very few plants, that generally do not provide shade and can not be eaten.
- Moving around - doing physical activity - will only increase your need for water so it's best to stay where you are; stay in the shade and move around at dawn and dusk.
- Moving around in the dark is dangerous because you can't see the dangers around you, like a cliff or snakes.
- Most search and rescue teams work during the day; not at night.
- Often, people that want to survive yet do not have survival skills do better than people who have survival skills, but don't have the will to survive.
- The most important aspect to surviving in the desert is water and shelter.
- If you can survive, the search and rescue teams will most likely find you.

Inter-agent Phrases [inter-agent phrases]

- Interesting information, maybe this will help too...
- I would like to add this...
- That made me think of this...
- I agree with your ranking.

Appendix D

Questionnaires

This appendix contains the non-standard questionnaires used in the experiment.

D.1 Consent Form

Assigned # _____

I, Neil Barrett, am a student in the department of Computer Science at Memorial University. I am investigating human-computer interaction - the study of how people perceive and interact with digital media such as computers. I am requesting your permission for you to take part in this study. You will be asked to interact with a virtual person that appears on the computer monitor. Together, you and the virtual person will work on the Desert Survival Problem. Your impressions and experience will be noted and discussed. To do so, you will be required to complete some short questionnaires and discuss your experience with the experimenter. The experiment will take about one and a half hours. All information gathered in this study is strictly confidential and at no time will individuals be identified. I am interested in how peo-

ple, as a whole, interact with computers and not in any individual's performance. Participation is voluntary and you do not have to answer questions that make you feel uncomfortable and may withdraw at any time (for students: your student status is not affected). The proposal for this research has been approved by the Interdisciplinary Committee on Ethics in Human Research at Memorial University. If you have ethical concerns about the research (such as the way you have been treated or your rights as a participant), you may contact the Chairperson of the ICEHR at icehr@mun.ca or by telephone at *telephone number*. The results of my research will be made available to you upon request. If you have any questions or concerns please do not hesitate to contact me, Neil Barrett, at Memorial, *telephone number*.

I, _____, hereby agree to take part in a study on how people interact with computers being undertaken by Neil Barrett. I understand that participation is entirely voluntary and that I can withdraw permission at any time. All information is strictly confidential and no individual will be identified.

Signature _____

D.2 Demographic data

Assigned # _____

Sex: M F

Age: _____

Occupation / Vocation (if student indicate discipline) : _____

Answer the following questions to the best of your ability.

I have read at least two books on wilderness survival.

True / False

I have taken at least one course on wilderness survival.

True / False

How many consecutive days have you spent camping with a tent. _____

I have seen a desert. (in person)

True / False

I have been camping in a desert.

True / False

D.3 Desert Survival Problem Ranking

Assigned # _____

Rank the items from 1 to 8 where 1 is the most important to you and 8 is the least important.

_____flashlight
_____knife
_____map
_____magnetic compass
_____one 2L bottle of water
_____a piece of a mirror
_____6 meter by 6 meter piece of blue canvas
_____jacket

Figure D.1: Image appearing on the DSP ranking questionnaire (www.zeledi.com)

D.4 Self-Efficacy

Assigned # _____

Circle the appropriate answer.

1) I am confident in my ranking of the desert survival items.

Not at all true — Hardly true — Moderately true — Exactly true

2) If I spent time in the desert, I would maintain the ranking of the four items I ranked highest or most important.

Not at all true — Hardly true — Moderately true — Exactly true

3) I would survive in the desert for 24 hours if I had my six highest ranked or most important items.

Not at all true — Hardly true — Moderately true — Exactly true

4) I would survive in the desert for 48 hours if I had my six highest ranked or most important items.

Not at all true — Hardly true — Moderately true — Exactly true

5) In a desert survival situation, I would make decisions that increase my chance of survival.

Not at all true — Hardly true — Moderately true — Exactly true

D.5 Relative Group Cohesiveness Rating

Assigned # _____

For each group of three statements, select one statement that best describes you (circle the letter associated with the statement). In other words, read statements a, b

and c then circle the letter identifying the statement that is closest to how you feel.

1. (a) I want to remain a member of the two agent group.
   (b) I want to remain a member of the one agent group.
   (c) I do not prefer to remain a member of one group above the other.
2. (a) I like my one agent group.
   (b) I do not like one group more than the other.
   (c) I like my two agent group.
3. (a) I don't care what happens in one group more than the other.
   (b) I don't care what happens in the two agent group.
   (c) I don't care what happens in the one agent group.
4. (a) I feel involved in what is happening in my two agent group.
   (b) I feel involved in what is happening in my one agent group.
   (c) I do not feel involved in what is happening one group more than the other.
5. (a) I am no more dissatisfied with one group over the other.
   (b) I am dissatisfied with my two agent group.
   (c) I am dissatisfied with my one agent group.
6. (a) In spite of individual differences, a feeling of unity exists in the one agent group.
   (b) In spite of individual differences, a feeling of unity does not exist in one group more than the other.
   (c) In spite of individual differences, a feeling of unity exists in the two agent group.
7. (a) I do not feel part of the one agent group's activities.
   (b) I do not feel part of the two agent group's activities.
   (c) I do not feel part of one group's activities more than the other.
8. (a) I feel distant from the two agent group.
   (b) I feel distant from the one agent group.
   (c) I do not feel more distant from one group over the other.
9. (a) I do not feel that my absence would matter to one group more than the other.

(b) I feel my absence would not matter to the two agent group.

(c) I feel my absence would not matter to the one agent group.

10. (a) If I was asked to participate in another project like this one, I would like to be with the one agent group.

    (b) If I was asked to participate in another project like this one, I would like to be with the two agent group.

    (c) If I was asked to participate in another project like this one, I would not like to be with the one group more than the other.

11. (a) I do not feel comfortable working with the two agent group.

    (b) I feel no more comfortable working with one group more than the other.

    (c) I do not feel comfortable working with the one agent group.

12. (a) I do not feel that I worked well with one group more than the other.

    (b) I feel that I worked well with the one agent group.

    (c) I feel that I worked well with the two agent group.

13. (a) I prefer interacting with one agent rather than two agents.

    (b) I am indifferent as to whether or not I interact with one or two agents.

    (c) I prefer interacting with two agents rather than one agent.

Appendix E

Additional Figures and Tables

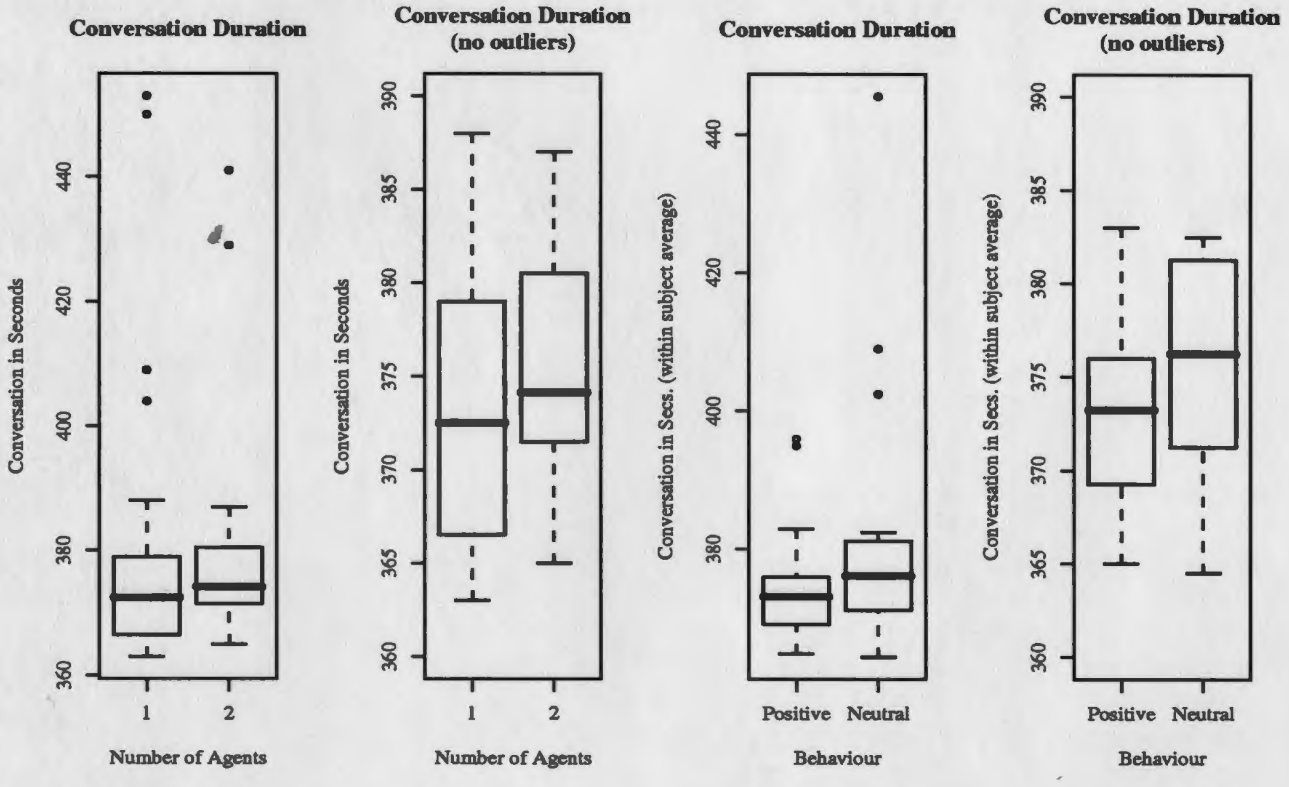

Figure E.1: Effects of Independent Variables on Conversation Duration

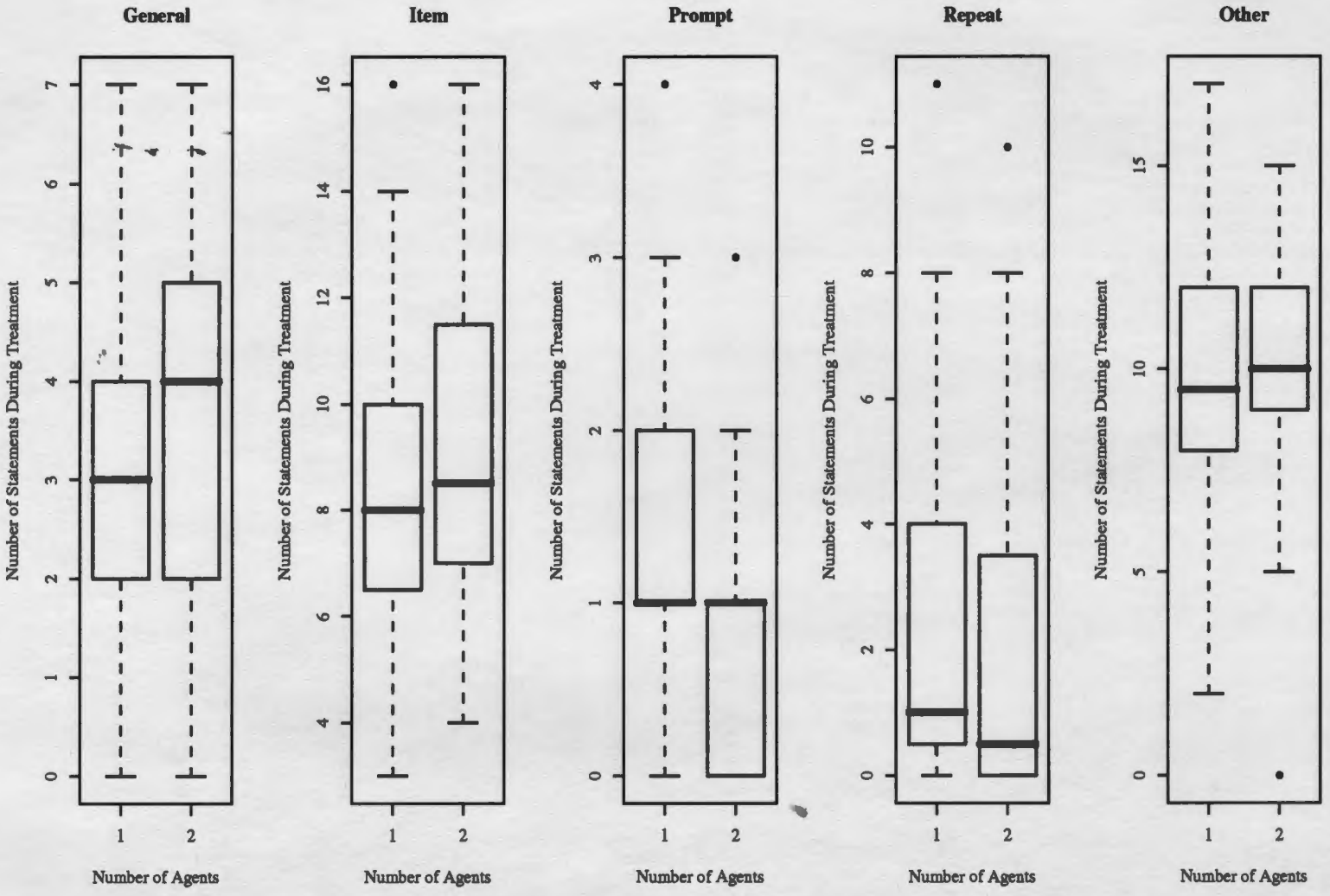

Figure E.2: Effects of Number of Agents on Conversation Style

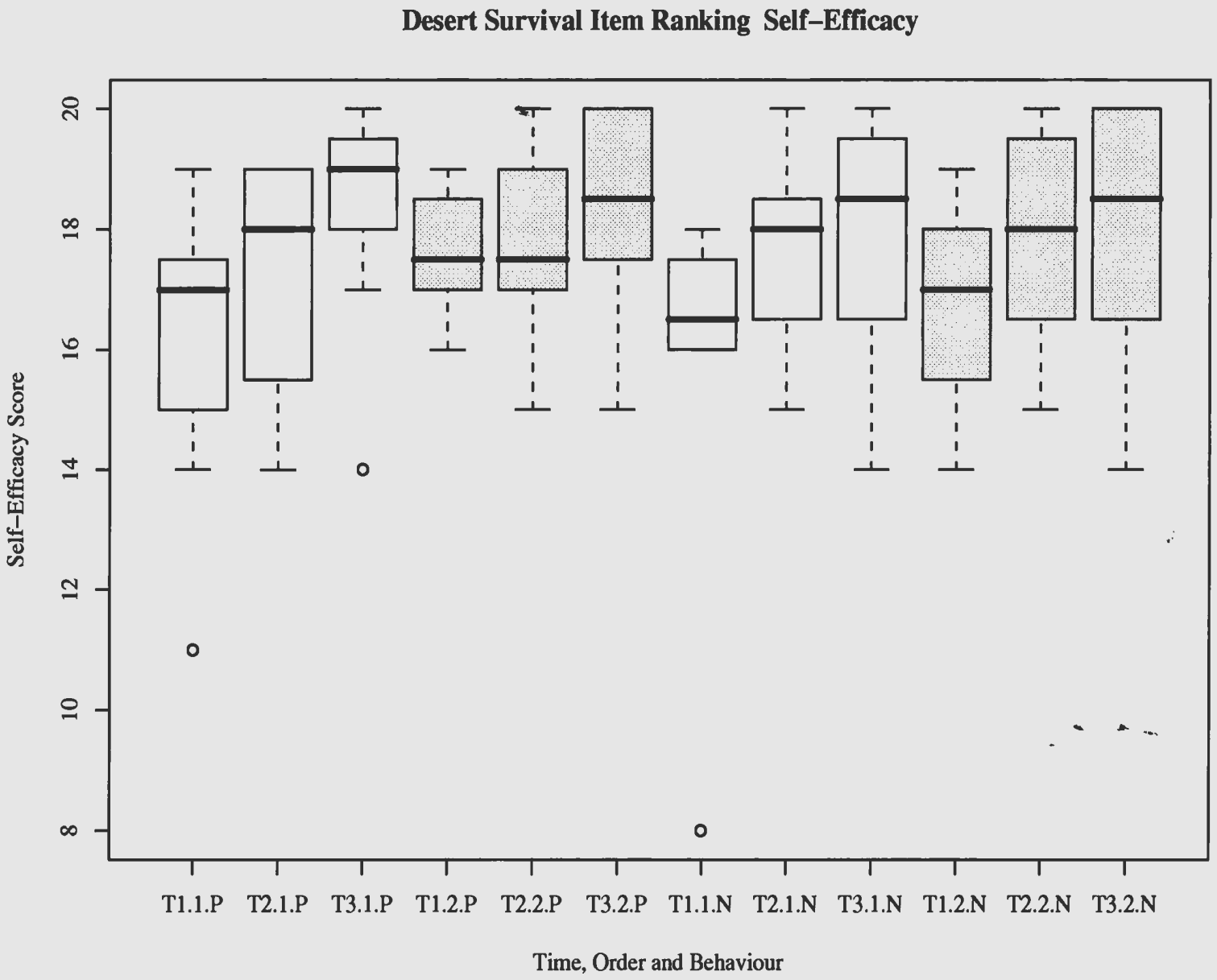

Figure E.3: Self-Efficacy Separated by Factors Time, Order and Behaviour

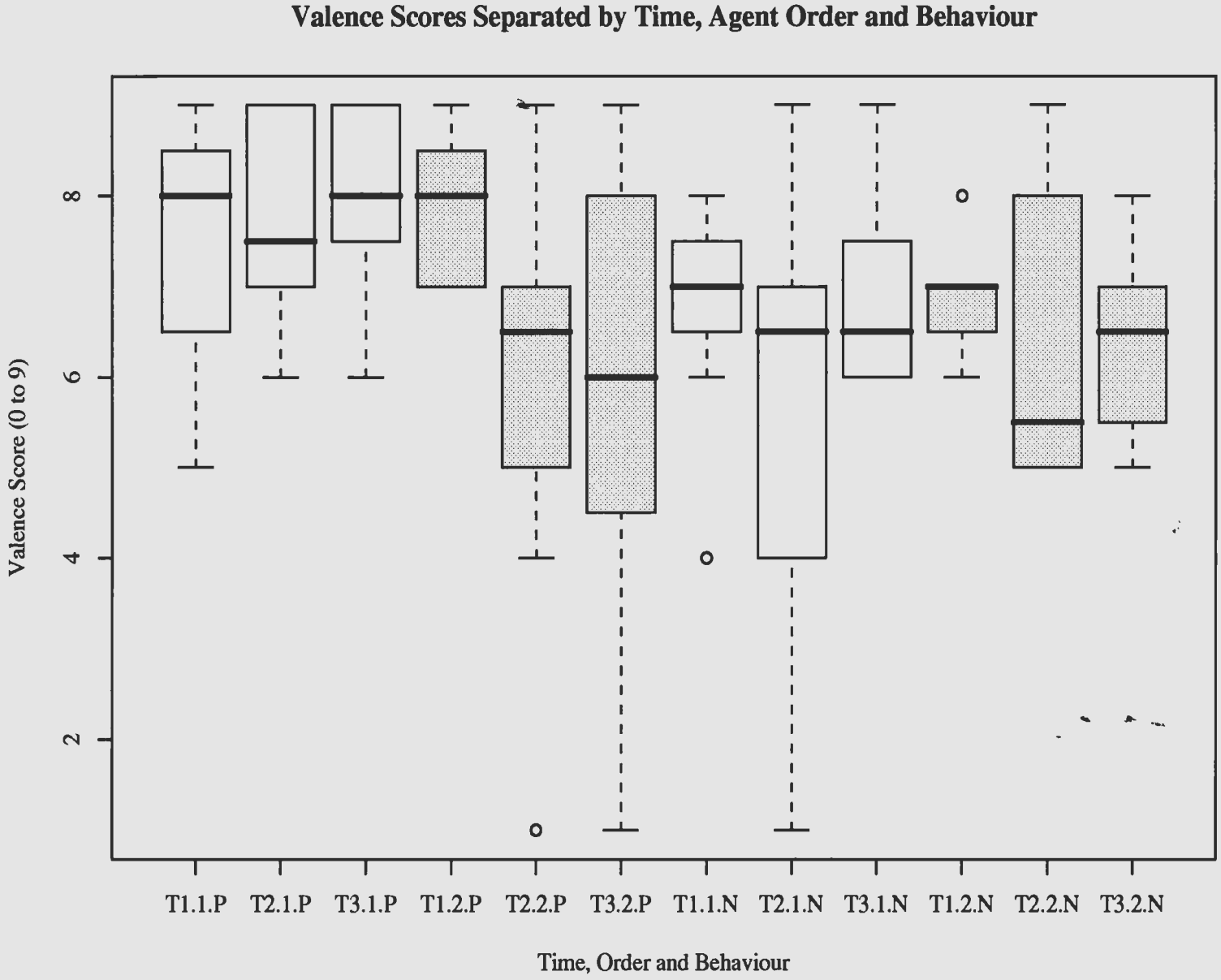

Figure E.4: Valence Separated by Factors Time, Order and Behaviour

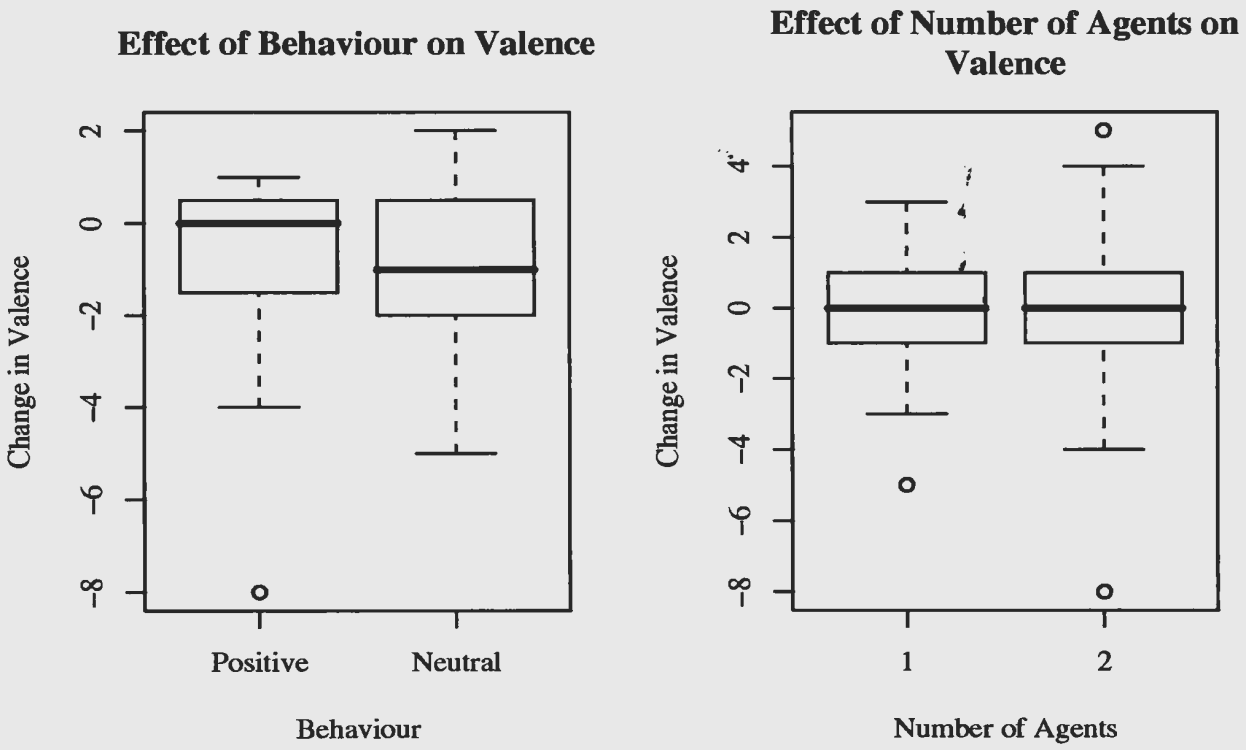

Figure E.5: Effects of Independent Variables on Valence

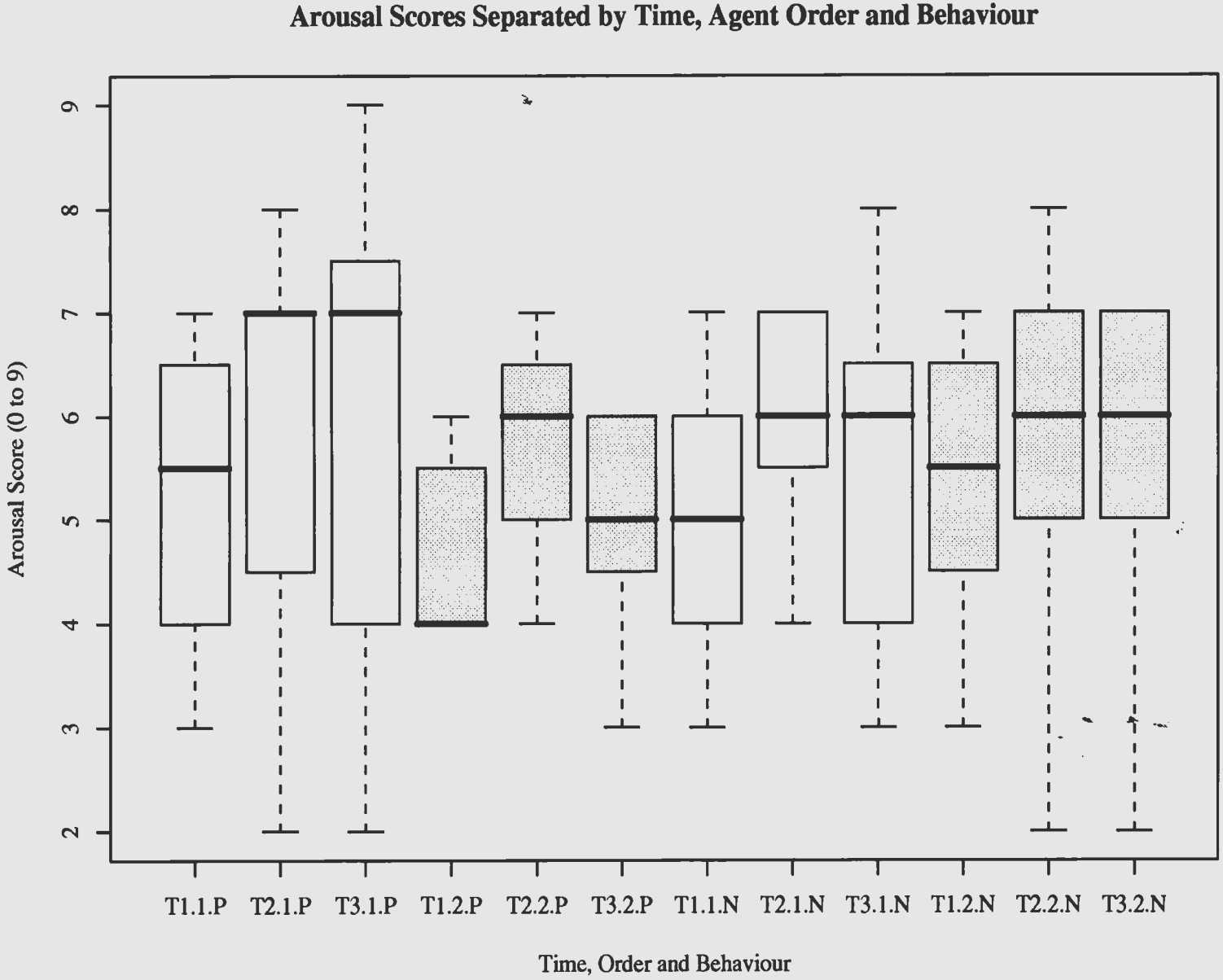

Figure E.6: Arousal Separated by Factors Time, Order and Behaviour

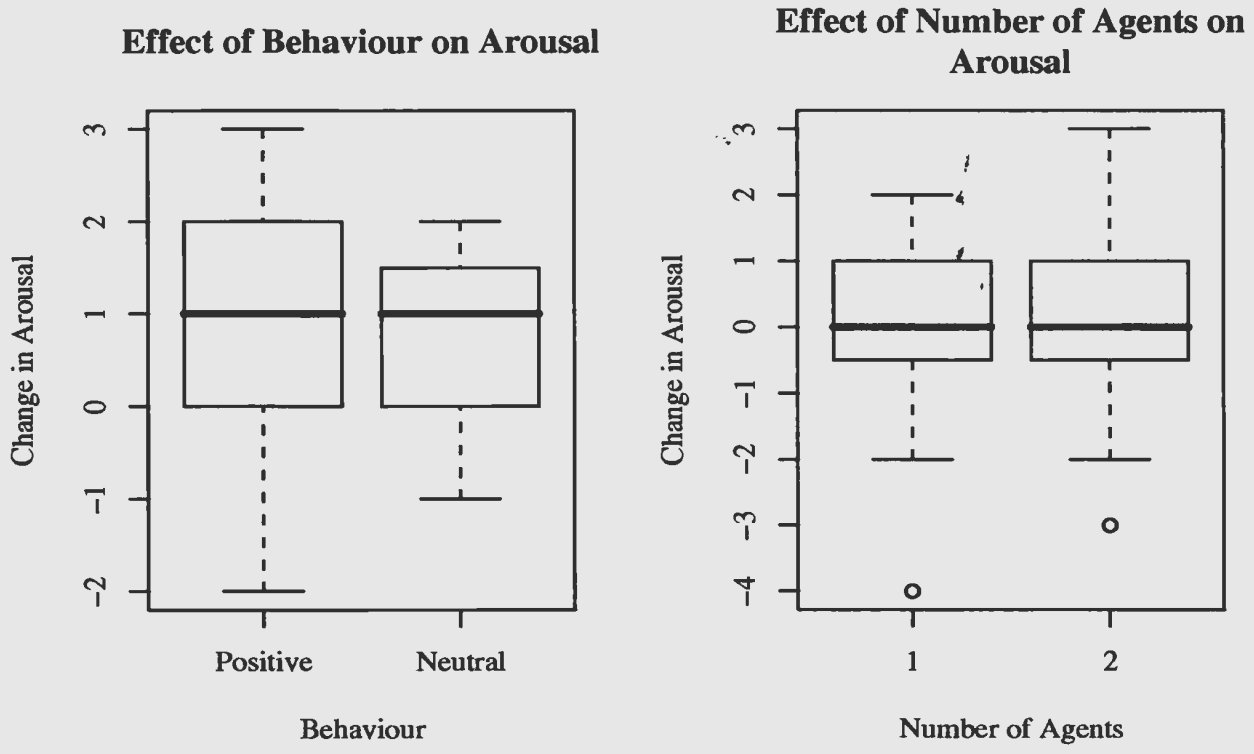

Figure E.7: Effects of Independent Variables on Arousal

Table E.1: Effect of Number of Agents on Participant Attraction-to-group.

	Number of Agents		
	One	"Neither"	Two
Number of Participants Attracted-to-group	13	1	18

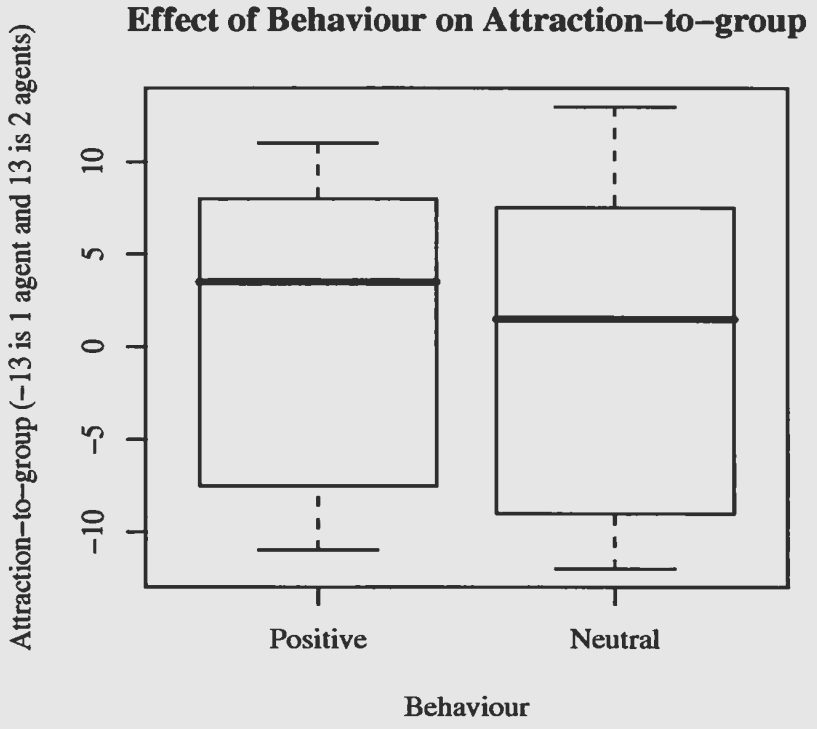

Figure E.8: Effect of Behaviour on Participant Attraction-to-group.