ÖFAI-TR-2001-04

Virtual Encounters:
Avatars, Actors, Agents

Sabine Payr
Österreichisches Forschungsinstitut für Artificial Intelligence
Schottengasse 3, A-1010 Wien
Email: sabine@ai.univie.ac.at

Work on this report has been undertaken in the framework of the research project "An Inquiry into the Cultural Context of the Ddesign and Use of Synthetic Actors"
BMBWK, GZ 21.074/1-III/A/3/99

Preface

Work on the research project "An Inquiry into the Cultural Context of the Ddesign and Use of Synthetic Actors" has started in spring 2000. In the first phase of the project, we collected examples of applications and tried to get an overview of the state of reseach and development in this field.

An overview of current research issues is given in part I of this report in the form of examples. Fictitious diary entries of a computer user in the near future serve as a first approach and introduction.

Part II is concerned with the issues and problems of a cultural studies approach to this domain. As a key issue, we discuss the double nature of virtual characters - both as technical products and as social actors. This question cannot and should not be decided prematurely, as we presume that this double character is essential for understanding the speicific issue at hand.

Part III is an attempt to open the field from the viewpoint of social interaction between human and virtual characters. We refer to sociological action theory as a theoretical framework, in particular to Habermas' "Theory of Communicative Action". The discussion of "communicative rationality" is to be seen as a preparatory work for the inquiry into the cultural aspects of social interaction with virtual characters which are the main subject of the project.

Cover page illustration: http://ligwww.epfl.ch/~thalmann/research.html

Contents

Part III – Agents as Actors

Part I

Presenting Projects and Issues

Preliminary remark

Project reports usually are written for the sponsors and project managers. As a result, they are a) not-too-honest (stressing what has been achieved more than what has not), b) boring (the readers are not supposed to get involved in the subject matter) and/or c) unintelligible (to confirm that both authors and readers are state-of-the-art).

This part of the report tries to be something different: it was written with a public, and a non-technical public at that, in mind. This is not to mean that the report as it appears here is to be published, but it is intended to develop one or more publications in the framework of this project. The idea here is to make a field of information technology development accessible to interested non-technical persons and to raise awareness of interesting questions about the implications of this technology for the individual and for society.

In a first approach, I have tried to "open" the field by mixing fictitious scenarios with current research and technical development. Not all aspects of research have been reflected. I felt that reporting every strand and direction of current research efforts would drag me too deep into the discussion inside the research and design community, converting me to the "engineers' viewpoint", while my real concern is with the (future) "users' viewpoint". It is difficult because there are, as yet, not many "real users" of virtual characters in whatever role. It is also difficult because the issue is one that escapes traditional studies of user behavior: virtual characters are "objects to communicate with", and their users are therefore not simply "users of objects" but at the same time participants in social interactions. This discussion will be taken up in part II.

Two stories

Tuesday morning

Tuesday, 9:00 a.m. I've just returned home from the day care center where I've left my little son. I turn on the PC and fetch my third cup of coffee in order to get ready for the day's work.

This afternoon at 3 p.m. we will have a meeting of the EU project consortium. The partners are from Finland, U.K., Spain and Germany. This time the colleagues from Valencia are responsible for the organisation of the meeting. They have rented a conference room, sent the instructions on how to get there, and other guidelines for the meeting. Participants should arrive at the conference location at 2:30 p.m. at the latest.

I have received the agenda and some other papers last week. Item 2 on the agenda concerns our part of the project. So I will have to prepare a short presentation of our interim report.

The meeting is virtual. The conference room is located on some server out there in the Internet. Technically, the meeting is an audio conference via Internet with application sharing. The participants will be represented by their replicants.

I don't want to send my replicant to the meeting without preparation and check. I had it produced some time ago at a computer shop. The procedure reminded me of the photo booths: inside the brightly lit booth, photos from different sides are taken. Some minutes later, the figure of your replicant appears on the monitor, showing some of the standard movements. Some details and parameters can be corrected at this stage before the replicant is finished and uploaded on the server from where I can download it remotely.

My replicant has not been in use for over three months, so it would be useful to have a look at her and to become again familiar with the controls. I start the software and load the image. The replicant wears jeans and a sweater: this is not suitable for the formal occasion, and looks somehow strange now in summer - even if the Internet does not have seasons. The virtual wardrobe I bought for the replicant is from fall, so not really right either. But I don't have time to download the summer collection right now. A dark, "neutral" looking suit must do for today.

I let the figure spin around to view it from all sides. Well, I have put on a little weight since the replicant has been made, so the avatar looks better than I do. I hope I will have lost the extra pounds before the next face-to-face project meeting. Obviously flattering replicants are not well received in business and scientific communities. Interestingly enough, men tend more to embellish themselves than women - making their bellies smaller and their shoulders broader, for example.

Before the photorealistic avatars became available practically everywhere, it was not easy for women to get a reasonable replicant. the geometry of the body was mostly pre-defined and you could only adapt some of the parameters. And the "basic model" from which you had to start was a sort of sexy doll figure - a small boy's or a programmer's dream, maybe, but a nightmare for professional women. Small wonder that many women refused to show their personal avatars. They preferred some neutral or phantasy figure. The new photo-realistic replicants are a much better fit, at least where looks are concerned. The movements are still standardized, and there is only one animation repertoire for all the replicants, which has obviously been modelled with male replicants in mind. Female replicants therefore look a bit curious when they move, with long steps and sweeping gestures. I've read recently that different motion libraries will soon be available, but for the moment, we have to get along with "unisex" movements.

The conference room is already accessible for a trial run. I log in and display the avatar in the room. First I have to adapt size and position: the figure should match the room and should move on the floor. Then I

look for my appointed seat at the conference table and place the avatar there. The command for having the avatar sit down is simply chosen from a control panel. I store this position as the default position for the meeting: I will be able to make the avatar return to it from everywhere with a single mouse click. In the past, most Internet connections were too slow to handle realistic movements like sitting down in real time. The avatars just stood or floated around somewhere in the virtual room.

Next step: I rehearse some simple movements like turning the head, raising a hand, nodding, etc. These are all movements that I can execute using the control panel and the mouse. They are the simplest movements, being independent of the room and of the other avatars. In the beginning, it was a little bit awkward and distracting to control the avatar while being in a meeting where important issues were being discussed (most of the time text-based, as in chats). But it is important to make your replicant look "alive" and aware. It is very disturbing to be in a meeting where the participants do not move and react at all. You know they are only avatars but still you can't help feeling isolated and strange. Once I had to give a talk in such a meeting with completely unmoving avatars, and I felt genuinely rejected. That was right at the beginning of virtual meetings, when most users were not familiar with the technology and did not master the animation controls. Today, it is regarded as bad manners when you leave your avatar sitting or standing somewhere without having it display some sign of "life" from time to time. Moving your avatar is not just toying around with it: the motivation for using replicants in virtual meetings is, after all, to bring in some of the non-verbal elements that are so important in communication. Researchers have found out that replicants make meetings more efficient and help to avoid misunderstandings.

There are still the context-sensitive movements to plan and record. First, I let the avatar turn the head towards the labelled seat of each of the other (not yet present) participants. I record these looks by the names of the persons, so that I can execute these movements later on by simply choosing the name. I do the same with a pointing gesture that I can use

at need to address a participant individually (but I probably won't use it, because pointing to persons is considered no more polite in virtual worlds than in real life).

The most difficult part of the preparations concerns my presentation. In our project meeting, we simulate a lot of traditional meeting behavior: speakers do not only stand up, but also walk around to the head of the table and talk to the audience. Presentation slides are displayed on the "wall" of the virtual room, so that the avatar can even simulate pointing at the slides. I train and record the necessary sequences of movements: standing up, walking, turning round is one of them. Turning to the slides and pointing to either the upper or the lower part of them are two others. That should be enough. There are people who can manage the motion controls so well that they can animate their avatar on the fly. But I am not so good at that and would only risk losing my train of my thought if I tried it during my talk. So having a few sequences ready helps a lot. To finish off with my preparations, I rehearse my presentation with replicant and slides, so that I feel quite confident about this afternoon.

Wednesday evening

Wednesday, 8:30 p.m. My son has just fallen asleep. I've got some time to get on with my online course in Spanish. I put on my headset and log into the learning environment with my student account. After some connecting and loading, Carmen appears in my browser window and greets me: "¡Hola, Sabina! ¿Que tal?"

Carmen is one of the virtual tutors in this Spanish course. She is a lifelike embodied agent and really looks like a Spanish teacher - or what I think a Spanish teacher should look like. She talks and understands my answers - well, most of them: it works quite well as long as my pronunciation is halfway correct and as long as I don't give unexpected answers.

First of all, she asks me how much time I've got today. I tell her that I would like to work for one and a half hour. She will choose lessons and

activities accordingly. Theoretically, I will be able to override her suggestions at any time, but I've rarely felt the need to do so in the past. She keeps track of the things that I've already learned, but also of the things where I have difficulties: in-between new activities, she will present some exercises or reminders of older material. She sets up a lesson plan with varying activities and contents that does not only take into account my current progress, but is also varied and entertaining. Actually, the time I spend on the course always passes incredibly fast. Only at the end I realize that I've been working hard.

Carmen is my partner in dialogues, but she also introduces and explains the other materials of which the course is put together: video clips, cartoons, texts, oral and written exercises, and so on. She is the one to give me feedback on my actions and answers, and hints when I am at a loss. She remembers exactly where I have problems, and I can be sure that the failed task or a similar one will crop up during the next session, and afterwards in gradually increasing intervals.

I am not the only student of the online course. In fact, there are lots of people following it in all parts of the world. As it is an open course, the participants' progress is very different. We all work with different speed and intensity. Communication among students - in Spanish - is an important part of the course, and a lot of fun. It takes place in the form of a shared chatroom where you can talk and listen and act in writing. However, you can never be sure whom you will find in the chatroom and whether it is worthwhile to go there. Sometimes it is "occupied" by a group whose language level is so far above or below yours that it is either boring or frustrating to participate in the chat. This is another one of Carmen's tasks: she keeps track of all the students who are online at a given time and how long they plan to be there. As she knows what level each of these students is on, she can put together an ad hoc chat group whenever there is a more or less homogeneous group of people online at the same time. In this case, she will build or alter the individual lesson plans and suggest to each student to join her in the chatroom. There she will follow the discussion and record the utterances of the students.

Sometimes she will suggest a topic for the discussion; of course, we are free to ignore it when we have other things to talk about. But especially at the beginning of the course the topics were welcome even if not really "hot". Some of it was the typical "language class" stuff, you know: in everyday life, you would not spend an hour discussing someone's family relationships. But by accepting Carmen's topics, you can be sure that she has all the necessary vocabulary ready on demand. In fact, this is the only situation where she translates words into Spanish. So you can ask her what a "great-grandfather" is called in Spanish (*bisabuelo*, by the way - I remember that from last week's lesson).

Of course, there are also "real", that is, human facilitators. They can be contacted via e-mail, and at certain times they can be reached via phone hotline. There are also scheduled chats with the facilitators. But with students dispersed in nearly every time zone of the globe, and with widely varying, self-defined course schedules, online communication with the facilitators is not feasible for everyone and at every time. So the human facilitators come in when there are questions that the virtual tutor cannot answer and when it is not necessary to get an immediate reply. For example, students send their essays and exam papers to the facilitators, because the virtual tutors, after all, cannot analyse free text or assess your overall progress. On the basis of a student's work, the facilitators will also make some manual adjustments in Carmen's goals, if necessary: for example, open a new set of lessons for her to choose from, or add a subject to the list of remedial actions.

Tonight, Carmen takes me on a shopping trip through the city of Madrid. She proposes to enter a boutique and makes me ask about dresses and trousers. I am not represented by an avatar as in yesterday's meeting. The scene is shown "through my eyes", so-to-speak, i.e. from a first-person viewpoint. I try to keep up the dialogue, but whenever I don't know what to ask or say, Carmen intervenes. I know that this shopping tour will come up again in one of the next sessions - but then Carmen will be very much less eager to help me out with my turns of the dialogue. It seems to make her genuinely sad when I don't remember things she

thinks I have already learned, so I will try not to disappoint her and will have a look into the suggestions for the shopping dialogue I can find in the materials.

Writing this, I notice that I tend to describe Carmen as if she were human. But this comes easy, given that she does show some traits of a genuine personality and seems to react emotionally to her students. She appears genuinely distressed when I repeat the same mistake over and over, and finally offers a remedial lesson for my problem. On the other hand, my progress makes her "happy".

From time to time, I get the opportunity to rate the various types of lessons and exercises so that the program gets some idea of what type of learner I am and of my preferences in learning style: whether I prefer talking to writing, images to sound, complex real-life situations to step-by-step-learning, immediate feedbak to delayed assessment, and the like. This does not mean that I get only tasks of my preferred varieties - sometimes I have to do it the other way as well. But the feedback I give makes Carmen follow my predilections as far as possible, and I have noted that she makes an effort to schedule one of my favourite activities at the end of each lesson.

Projects and Implementations

The two stories with which I have started this paper are certainly fictitious, but in some respects hardly so. The "virtual meeting", for example, appears as a tiny and modest detail of the "metaverse" imagined by Neal Stephenson in his novel "Snow Crash". As of today, some of the techniques are still experimental, while others are already market-ready, but used in isolation. What makes the stories futuristic is the seamless integration of all the different techniques in one interactive and user-friendly application. In what follows, I will describe in more or less detail projects and programs on which I based my story.

The two stories illustrate two different types of virtual characters: replicants as a special form of avatars in the first and virtual tutors as a special form of synthetic actors in the second. Avatars are like electronic puppets: they do not act on their own but represent their "owners" in a virtual world. Synthetic actors normally do not have a specified owner. They act on their own, which implies that they make plans and try to execute them by interacting with their environment - which includes the user. Note that this differentiation in terminology is not consistently made in literature, but I will stick to it in this paper for the sake of clarity.

Of course, there is no "natural" borderline between the two types of virtual characters. Once a replicant has been created, it could be used as an autonomous agent in whatever environment or role. Inversely, a user could choose any kind of virtual figure - animal, phantastic, geometric or another human - as his or her representative. The difference between the two, therefore, does not reside in appearances, but in functionality. Autonomy is the key criterion that allows us to distinguish between avatars and actors.

However, it is not an all-or-none-criterion, but rather an axis with "complete autonomy" on one end and "complete control" on the other. The examples of virtual characters - real or imagined - that I present can be ordered along this axis. Avatars, in my definition, will be found more at the "control" end, actors at the other. But we could imagine an avatar that is given certain degrees of freedom by its owner, e.g. to act according to standard procedures in certain routine situations, pursuing its owner's goals without being under permanent control (Cook et al. 1999). It would acquire some characteristics of an actor. On the other hand, no synthetic actor (so far) is totally autonomous: at least its overall goals are pre-defined by its designers, and only within these goals it can plan its actions autonomously. For example, a "virtual salesman" is pre-determined to sell a certain product, e.g. cars. It cannot make plans and take actions outside its overall goal, e.g. to court a potential buyer instead of praising the car it tries to sell.

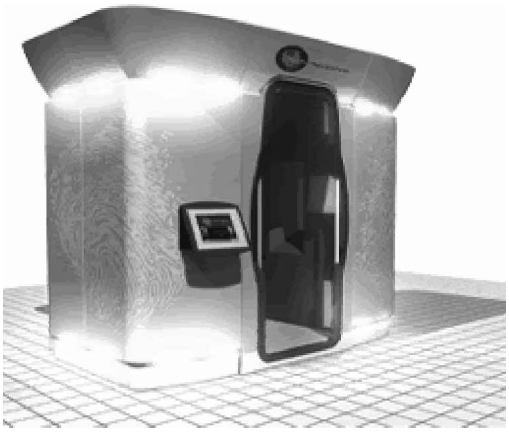
Replicants and virtual meetings

A replicant is an avatar that reproduces the features of a "real" human being, be it living or dead. There are already some replicants of famous persons. In my story, however, a replicant reproduces its user's (or owner's) features and can represent her in virtual worlds.

Creation of replicants

The automatic creation of a replicant mentioned in the story is not a futuristic invention. Such a technique already exists:

The company AvatarMe has developed a booth, comparable to photo booths, in which four photos are taken of the client. An assistant helps you to take the right position. In a few minutes, the avatar is constructed from the photos:



AvatarBooth (© AvatarMe Ltd.), http://www.avatarme.com

Technically, the AvatarBooths represent an integration of two breakthroughs: the instantaneous acquisition of photo-realistic 3D shape information from high-resolution digital colour images and the subsequent automatic reconstruction of articulated human models. They work by initially capturing four images of a person standing in four specified postures, from the front, back and two sides. Controlled illumination is used to reliably extract the silhouette outline for each camera view - tight-

> fitting clothing produces the best results - while analysis of the silhouette contour accurately locates a set of key-feature points and enables the silhouettes to be separated into body parts. This allows the software to automatically generate a fully-jointed skeleton for the model, while a transformation algorithm is applied to each body part that uniquely maps all points inside the silhouette contour of a captured image to a corresponding point inside the silhouette of a model. Transformations from multiple views are then applied to the generic avatar to morph the three-dimensional shape to exactly that of a particular person. Digital colour information captured from the images is finally texture mapped onto the transformed generic model to create an avatar with the shape and appearance of the person.

The result is a fairly realistic replicant at a moderate file size - the company's paper mentions 400 to 500 KB. The client can download the avatar from the Internet, together with player software that allows him to view his "incarnation" from all sides and to run the animation in which the avatar walks, runs, climbs stairs and so on.

Avatar, produced with the AvatarBooth and animated (© AvatarMe), http://www.avatarme.com

The AvatarBooth was first shown at the SIGGRAPH '99 conference. Currently, it is in use at the Millenium Dome (London), where visitors can "incubate" their avatars for free. The company's plans are ambitious: on the production side, booths should be installed in malls, first in the USA. The company's goal is of "incubating avatars of people from 25 % of US households within 6 years". On the application side, the company cooperates with computer game producers to insert the avatars into new or existing games as characters.

Modification

Tools for the modification of these avatars are planned, but have not yet been developed. Software for construction of realistic avatars does however exist, and it is plausible that a tool for the post-production modification of the avatar by the user could look somewhat like the following.

AvatarStudio: Adjusting the avatar's measurements, http://www.2nd-world.fr/avatarstudio/default.htm

The number of parameters that can be adjusted is impressive. It allows for subtle changes, but also for major transformations of the avatar. With

this tool, avatars can be created "from scratch". It offers, however, also the option to import a digitized photo of a face to be grafted on the avatar's head, which is another method of creating one's realistic replicant.

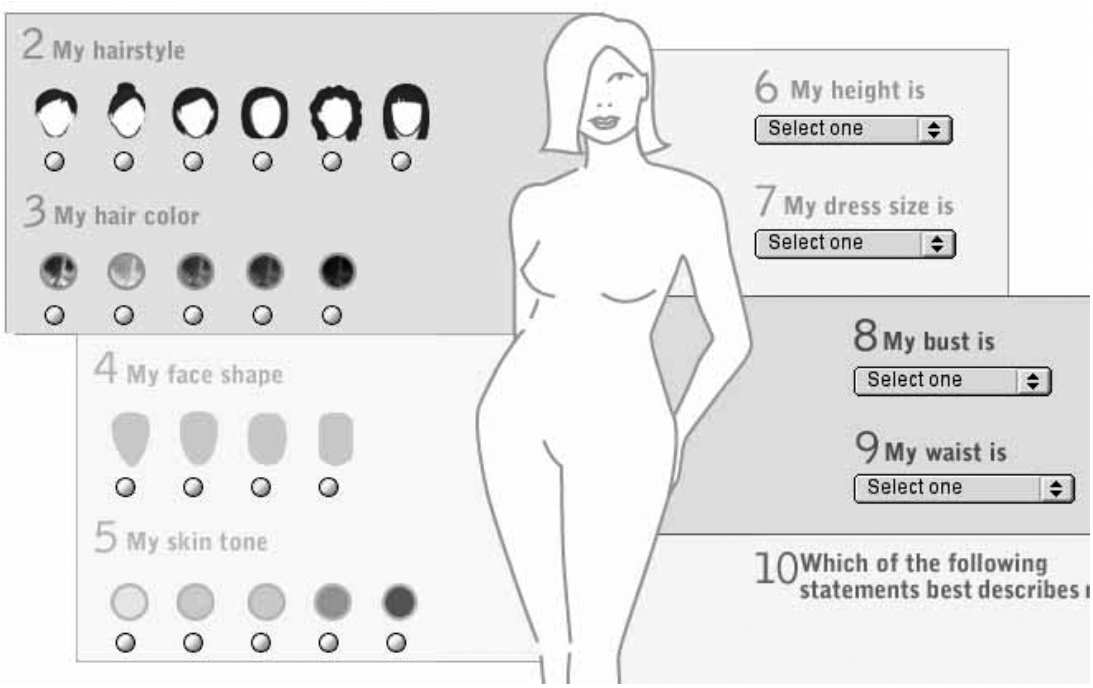
With a tool like this, the "photorealistic" avatars could be adjusted to the taste and needs of the user. He or she might not be happy to appear in the Internet exactly like in real life. Any form of cheating would be possible and presumably welcome, as virtual body shaping is certainly less tedious than doing it in the fitness studio.

Clothing

One shortcoming of the avatar constructed from a photo is also its invariable clothing. For the application of the avatar in games and chatrooms, it may be sufficient, but other applications - such as the formal meeting mentioned in the story - might require appropriate and changing clothes. AvatarStudio offers an extensive wardrobe from which to choose. Interestingly enough, there are three layers of underwear to choose from. It can be presumed they are not essential to the functionality of the avatar. The idea of avatars in the role of virtual sexual playmates is obviously present in the minds of the designers as a promising commercial application.

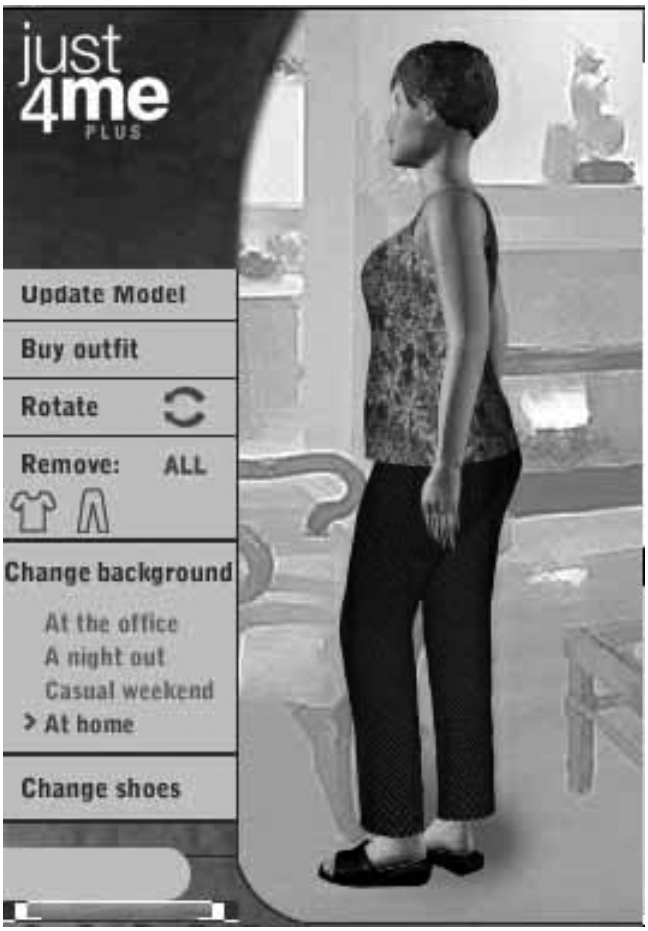
AvatarStudio: Choosing the avatar's clothes, http://www.2nd-world.fr/avatarstudio/default.htm

The mail-order store J.C. Penney's uses a not so very different tool for selling clothes via the Internet. On the basis of a simple dialogue, in which the user can enter a handful of her features, such as colour and length of hair, size, body shape and so on, a simple, static model is constructed and displayed.



Just4me: Definition of model, http://www.just4meplus.com/

The model will be shown wearing the clothes that the user selects from the products offered. It is possible to view from all sides the effect of the clothes on the model before ordering them.



Just4me: Display of clothes, http://www.just4meplus.com/

This is obviously an application where the user is interested in achieving as much realism as possible. Given the rather run-of-the-mill choice of clothes offered for electronic sale and the standardized appearance of the figure, it remains doubtful whether this facility really attracts visits of the online shop - which is clearly its purpose.

Dressing a static figure is, after all, a relatively simple process. Things become more complicated once the avatar is moving. Each (visible) piece of clothing has to follow body movements, adding extra complexity to the animation. It is not astonishing, hence, that the users of the AvatarBooth are recommended to wear close-fitting clothes.

But this is only a short-term solution. A resarch group at the Swiss MIRALab has been working for years on the modelling of virtual characters' clothes (Magnenat-Thalmann and Volino 1997, Furukawa et al. 2000). The challenge is not only to model the clothes as separate objects, but also to animate them in correspondence with the body's movements.

Animation and facial expression

While the photorealistic avatars I have described above certainly are quite good in showing a person's individual look, they are still far from displaying also the individual way of moving. When you watch people in the street, you will notice quite quickly that no two persons have the same way of walking. The same is true of course for running, jumping, turning around, or pointing to something with a finger.

There are differences not only in the way the same gestures are made, but even more so in the repertoire of gestures, and the choice which gesture is made in which situation. Part of the repertoire and its use are culturally constrained, but inside these limits there is still room for individual variation. Much psychological work has been done on face recognition and the perception of facial expressions, but to my knowledge we know little about how perception of individual posture and movement contributes to the recognizability and characterization of persons.

Current research in animation of virtual characters focuses on the generation of movement on the basis of physical laws and models (cf. e.g. Ventrella 2000, Bret 2000) instead of "traditional" animation techniques. The synthesis of any life-like movement or facial eexpression in real time is, at the moment, enough of a challenge, so that the individual variation has to be left out of the picture for the time being. The technology of motion-capturing is used to capture an individual's movements and gestures from which the repertoire of an animated character can be developed. However, this complex technology is far from being as close to a large-scale commercial product as is the booth for the production of static photorealistic avatars. So the company AvatarMe, at present, can

only equip the avatars with a standard set of movements, executed in the same way by all the avatars.

Virtual meetings

Virtual meetings as the one described in the story would be a special case of inhabited virtual worlds (IVW). Large-scale conferences in virtual worlds like AV98 and AV99 (Damer et al. 2000) have already taken place, with several thousand participants.

View of the Avatar98 Conference, http://www.activeworlds.com

The roots of inhabited virtual worlds lie in the text-based multi-user environments that have been around since the 1970s and continue on the WWW. They now develop from text-based environments to visual "worlds" in which users are represented by avatars. "The trend is towards rich multimedia digital conferencing on the Internet." (p. 3):

> Central to the concept of IVW are Avatars, the visual embodiment of people in Cyberspace. By using Avatars, you can go beyond simple virtual meetings in which only (spoken) text and documents are exchanged. Avatars can, increasingly so, also embody such

> important aspects of human social interaction as gstures, proximity to the group, and emotion. ... A conversation between people as Avatars in an IVW comes closer to the feeling of a face-to-face conversation than a conversation which is only the exchange of text by e-mail.
>
> A second important characteristic of IVWs is the possibility to construct various types of virtual meeting spaces. For the design of these types of meeting spaces, we can use different generative metaphors. The meeting space can be small, being functionally equivalent with "rooms" in physical space. ... IVWs can be the meeting space for the private auditorium, where "one-to-many" presentations are given to usually closed audiences with the assistance of bot-driven slide shows or audio. IVW can also cope with EMRs (electronic meeting rooms, SP), where a "two-way interaction" exists with one or more speakers and responding audiences in a distributed environment. This can be done by web links, live video and audio streaming and backchannel chat for questions and answers." (Damer et al. 2000, p. 5)

Existing web-based virtual meeting rooms such as ActiveWorlds are still rather poor in representing and handling avatars. Animation is restricted to few gestures, and facial expression (e.g. of emotions) is out of the question at current resolution and data transfer rates.

The virtual meeting described in the story is, compared to these virtual conferences, "traditional" on the one hand: for example, participants are sitting at tables and walking around on ground level while there is no reason why they shouldn't fly or rest suspended in mid-air. On the other hand, managing not only a two-way, but a many-to-many interaction as in meetings is still a technological challenge.

Other uses of replicants and avatars

Chatrooms and computer games seem to be the first and most important implementation of avatars at the moment. Other fields of use are

naturally those where lifelike *appearance* in virtual environments plays a role: models and "dummies".

The virtual character "Jack" has been developed at the University of Pennsylavania's Center for Human Modeling and Simulation. In this figure, most attention has been paid to realistic functioning of the body. The reason is that Jack is mainly used for "virtual user testing": vehicles, machines or rooms that have been designed on the computer can be tested without the need to build them in "hardware".

This is the range of tests for which Jack can be used in engineering design *(http://www.transom.com/Public/app-design.html)*:

> Fit and comfort—Is your design optimized for user fit, comfort, visibility and access to controls?
>
> Visibility—What can different sized people see when they operate your equipment or vehicle?
>
> Ingress and egress—Can your target population easily climb in and out of equipment or vehicles?
>
> Reaching and grasping—Are controls placed so everyone can reach and operate them?
>
> Foot pedal operation—Can variously sized people comfortably operate foot pedals?
>
> Multi-person interaction—How do multiple people interact with your product?

> User maintenance—Can everyone perform maintenance tasks for your products, like changing a tire.
>
> Strength assessment—Does operating your product require inordinate strength or create the potential for injury?

Current research on Jack however goes beyond the physical simulation of the human body to build in aspects of planning, decision-making and reaction to the environment. The motivation is that "different conditions can lead to different motions because people adapt their behavior to circumstances through their unconscious reactions and conscious decision-making." (Badler, Reich, and Webber 1997, p. 44). It is interesting to note that this research program, which started out to simulate human physics, comes to the conclusion that even human movement involves much more than physics:

> Human movement realism reflects decision-making in context. For realistic animation, synthetic humans must engage in such decision-making if we want them to share human qualities. Sensed human motions are insufficitient because they reflect only decisions that have already been made. Physics is insufficient because there are no decisions outside the outcome of the mathematical laws. Conscious humans are neither puppets nor mannequins. They coninually assess their surroundings (to validate expectations, avoid obstacles, minimize surprises etc.) make choices, and plan for the future. Different cognitive "styles" of sensing, decision-making and planning ... can make animated human agents behave differently in response to the same environment and symbolically-specified activity. (Badler, Reich, and Webber 1997, p. 46)

These findings point directly to the subject of this study which is based on the hypothesis that not only conscious and situational choices influence the behavior of humans down to their body movements, but also the inconscious cultural background of acquired attitudes, habits, postures, distances, etc.

There is another field where virtual characters promise to be successful: as models.

Webbie Tookay (by Illsuion2K), http://www.illusion2k.com/

Tyra (by Vierte Art), http://www.nodna.com/

The motivation behind this application seems not to be to model humans as realistically as possible, but rather idealistically. And this does not only cover the visual appearance of the virtual models, but also the "ideal" behavior of models. Statements like the following obviously come to journalists' and marketing people's minds when writing about virtual models:

> ... the model with the perfect specs: she doesn't gain weight or party all night, she doesn't have a boyfriend and she never complains. (Newsweek)
>
> The perfect model: she won't age, gain weight, or throw tantrums (Wallstreet Journal)
>
> Never before realtime models have been so beautiful, so flexible, so fast and so valuable. ... With ease and elegance, always fresh and lively ... (noDNA, agency for virtual models and actors)

The advantage of virtual models, then, is precisely that they don't plan or make decisions.

And finally, an important field for animated virtual humans are movies and computer games. Maybe the most famous virtual character is, at the moment, Lara Croft from the "Tomb Raider" game series. Obviously, she too is the outcome of her designers' ideas about an "ideal woman".

Lara Croft,
http://www.gamespot.co.uk/pc.gamespot/features/tombraider_hist/

The game developers obviously thought it necessary to invent a personality and a biography for her.

LARA CROFT'S BIOGRAPHY

Nationality: British

DOB: 02/14/68

Birthplace: Wimbledon, Surrey

Marital Status: Single

Blood Group: AB-

Height: 5 feet 9 inches

LARA'S EDUCATION

Private Tutor: 3-11

Wimbledon High for Girls: 11-16

Gordonstoun Boarding School: 16-18

Swiss Finishing School: 18-21

Lara Croft is the daughter of Lord Henshingly Croft and was raised from birth as an aristocrat. A lover of rock climbing and acrobatics, Lara is construed by some to be a mercenary, big-game hunter, and master thief. She has survived a plane crash and has learned how to thrive under extreme circumstances.

The Earl of Farrington, the man Lara's parents hoped she would wed, is patiently waiting for her to get all this adventure out of her system. But Lara has found contentment in using her keen wit and intellect to write travel books. She has several other published works as well.

Among her many achievements, Lara has set a record time driving to the Alaskan Highway from Tierra del Fuego and has gained some notoriety for having slain an actual Bigfoot in North America.

She can stitch tapestry, ride a motorcycle, repel, and perform a host of stringent physical tasks. She is very adept with a firearm as well. But through all of this she manages to have a warm heart. She is a fan of Brian Blessed, the voice of Boss Noss in Star Wars: Episode I, and wants to see him one day attain his goal of climbing Mount Everest.

If we follow this description, the ideal female - at least for a computer game - is Jane Eyre, Emma Peel and Karen Blixen all rolled into one. To list "stitching tapestry" as one of her achievements is particularly

endearing, together with her aristocratic background. Some 19th century virtue appears still to be desirable in females - even if they are protagonists of action games.

While computer game designers start to pay more attention to the development of characters and plots (Blache n.d.), they come closer to the domain of interactive literature (Hirsh 1998). In both fields we can expect to see a crossing of the boundaries between avatars and synthetic actors, as virtual characters in games and stories will be equipped with personality and autonomy to act "freely", albeit inside the limits of the general plot or game world (Hayes-Roth 1998).

Autonomous synthetic actors and online learning

Synthetic actors can be used in various roles and "jobs": guides, salesmen, tutors, information brokers, and so forth. As more and more goods and services are offered online, there is also a growing need to make online-shopping easier and more "personal". Together with the fact that Internet services are offered worldwide and permanently, this trend favours the development of virtual assistants of many different kinds who seem to be a promising way to make the services more usable and to take over some tasks of human workers who cannot serve huge numbers of customers around the clock. I have chosen online learning as an example for a field of application of synthetic actors not only because I am more familiar with this domain than others, but also because it appears prominently on the reasearch agenda, although (or because?) it is a challenging enterprise to model the highly complex and sensitive interactions between teachers and learners.

The felt need for pedagogical agents might be motivated by the steps in the development of online learning that we currently assist:

1. *The "surfing model":* In a first step, learning materials are put online, mainly as static web pages with few software interaction. Free Internet browsing serves as a model of this phase that we could therefore call the "surfing phase" of online learning. Apparently justified by

simplified constructivist learning theories, explorative learning and autonomous student activity are foregrounded. Drop-out rates, however, are very high, and students tend to be less active than presumed.

2. *The "facilitation model":* In a second step, therefore, online materials are combined with human assistance, e.g. under the label of "facilitation". Facilitation can cover anything from technical support to active tutoring. The most frequent case is to have a facilitator who actively reminds students of deadlines, answers technical questions, receives and comments students' work, and constantly but mildly pushes them towards participating actively, e.g. in collective activities such as discussions. Human facilitators, however, are not available at all times, and can serve only so many students at once.

3. *The "intelligent agent model":* In a third step, some of the facilitators' tasks are transferred to software. This includes interactive courseware which can now, with better connectivity, also be delivered over the Internet. Some of the software support is in the form of pedagogic agents. Pedagogic agents need not be visualized as virtual characters, so students need not be aware that they are there. Some are visualized, and in fact can serve as the common front-end for several software agents or software interactions. Some do what intelligent tutoring systems were supposed to do, and some only or also take on functions that are specific to web-based training.

As I have discussed current research and development in the field of visualization of virtual characters in the first part of this chapter, I will, in what follows, concentrate on their functionality as more or less autonomous agents. Visualization of the agent is an additional feature but serves not only as "embellishment". The fact of having a virtual persona as a counterpart in an interaction will be shown to raise specific questions later on. Here I will outline the potential functionality, again using the fictitious story as an illustration. What I present is therefore a combination of what synthetic pedagogic characters are able or supposed to do with what intelligent tutoring systems are able or supposed to do. The fictitious pedagogic agent in the story integrates these functions and

even has additional capabilities, some of them to be found already in (not necessarily pedagogic) agents, and some of them, to my knowledge, not yet implemented.

One of the futuristic capacities of "Carmen" is the ability to organize, co-ordinate and support lerners' group activities. The problems of co-ordinating agents in learning environments and an approach to them are discussed in detail in (Wasson 1998).

Adaptive and intelligent tutoring systems

Brusilovsky (Brusilovsky 1999) gives a good overview of intelligent technologies for Web-based Education which I will resume shortly. He distinguishes four core functions of ITS:

Curriculum sequencing: The goal is obviously to offer to the student individually the optimal sequence of tasks and steps. Active sequencing means that the program pursues a learning goal for the student. Passive sequencing means remediation: the system reacts to the student's problems (gaps in knowledge, misconceptions ...) by offering remedial steps but does not have its own learning goal. One can further distinguish between systems that propose learning sequences on a high level, by determining subgoals and proposing lessons, and low-level sequencing where only the subtasks inside a unit are proposed by the software. Sequencing did not get much attention in ITS but has become important in web-based learning environments where the selection of appropriate learning materials and the optimal navigation through the learning space become a problem. While sequencing was considered a secondary function of ITS, compared to problem solving support, it has gained more importance with the switch from stand-alone to web-based educational media. In the web, navigating through the multitude of available materials and activities is often even regarded as the first and foremost interaction of the user with the learning environment, corresponding to the "surfing model" of online learning.

Intelligent analysis of student solutions is a technology where the final answers and results of the student's work are examined.

> To be considered as intelligent, a solution analyzer has to decide whether the solution is correct or not, find out what exactly is wrong or incomplete, and possibly identify which missing or incorrect knowledge may be responsible for the error. ... Intelligent analyzers can provide the student with extensive error feedback and update the student model. (Brusilovsky 1999)

Interactive problem solving support does not wait for the final solution, but intervenes, if necessary, at each step that the student makes, offering intelligent help. Help can vary from simply signalling an error to hints to executing the step in the student's place. These systems are often called "interactive tutors" in the narrow sense of the word.

Example-based problem solving is a new and special case of this type of intelligent support by suggesting cases from the student's earlier experience as examples in the course of the problem solving process.

While interactive problem solving support technology dominates in stand-alone ITS, the web again changes the picture:

> Both intelligent analysis of student solutions and example based problem solving support appears to be very natural and useful in Web context. Both technolgies are passive ... and can be relatively easy implemented on the Web using a CGI interface. ... an important benefit of these two technolgies in the Web context is their low interactivity: both usually require only one interaction between browser and server for a problem solving cycle. This is very important for the case of slow Internet connection. These technologies can provide intelligent support when a more interactive technology will be hardly useful. (Brusilovsky 1999)

However, these purely technical limits are only a temporary obstacle for more interactive intelligent tutoring in web-based education.

Animated pedagogic agents

The new dimension that animated agents as virtual tutors contribute to the field of educational media has been summarized as follows:

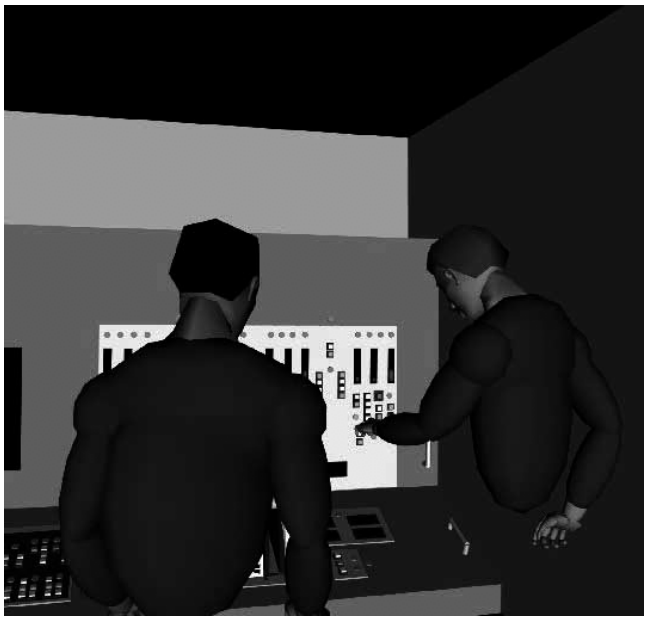
> As in previous work, students can learn and practice skills in a virtual world, and the computer can interact with students through mixed-initiative, tutorial dialogue in the role of a coach or learning companion. However, the vast majority of work on tutorial and task-oriented dialogues has focused on verbal interactions, ... an animated agent that cohabits the learning environment with students allows us to exploit such nonverbal communication. The agent can demonstrate how to perform actions. It can use locomotion, gaze, and gestures to focus the student's attention. It can use gaze to regulate turn-taking in a mixed-initiative dialogue. Head nods and facial expressions can provide unobtrusive feedback on the student's utterances and actions without unnecessarily disrupting the student's train of thought. ... Moreover, the mere presence of a lifelike agent may increase the student's arousal and motivation to perform the task well. (Johnson, Rickel, and Lester 2000)

Both of the examples with which I will illustrate the state of research on lifelike pedagogical agents have been developed at the University of South California's Center for Advanced Research in Technology for Education (CARTE). They are called Steve (Soar Training Expert for Virtual Environments) and Adele (Agent for Distance Learning: Light Edition) (cf. Johnson, Rickel, and Lester 2000, Elliott and Brzezinski 1998).

Steve interacts with students in networked immersive virtual environments. The figure is a snapshot from an application of Steve to the training of operators of engines aboard US navy ships.

Steve demonstrating the operation of the ship's gas turbine propulsion system, http://www.isi.edu/isd/VET/steve-demo.html

The engine room is modelled as a virtual world in which trainees operate machines via the data glove. The agent has information about the current state of the virtual world. Steve can therefore not only demonstrate the operation of the machines, but also react to the trainee's actions with feedback and hints. Steve uses speech to communicate with the user. In a next step, Steve will also be used for team training, where several agents and/or agents and users should co-operate as team.



Two agents in a training team, http://www.isi.edu/isd/VET/steve-demo.html

While Steve acts in a 3D virtual world, Adele has been designed as an agent for web-based training, to be used with conventional interfaces and browsers. At the moment, Adele is used to deliver distance continuing education to medical doctors.

> Adele consists of a pedagogical agent and a 2D animated persona, which is implemented as a web-based Java applet. Adele adapts the presentation of the material as needed, provides hints and rationales to guide student actions, and evaluates student performance. Adele is currently being developed to work with both medical and dental students. She will be used in two medical education systems: case-based diagnosis and trauma care. Her current work in dentistry involves assisting students with simulation exercises for a course on geriatric patient care. (http://www.isi.edu/isd/carte/carte-projects.htm)

Adele guiding the student through the simulation of medical cases, http://www.isi.edu/isd/ADE/ade.html

Goals, Plans and Autonomy

When are agents autonomous, and why do pedagogic agents need autonomy? Autonomous agents "can seamlessly integrate planning and execution, adapting to changes in their environments." (Johnson, Rickel, and Lester 2000). Like autonomous agents in general, pedagogical agents must exhibit robust behavior in rich, unpredictable environments. The key characteristics however, which separates autonomous agents from avatars and animated figures in games and movies, must be the capacity to pursue a goal and to plan actions in order to reach the goal.

The goal can be set by the user (or designer) at any level of complexity. But in order to deserve the predicate "autonomous", the agent must be able to choose among a certain range of possible ways, in function of the situation, to reach the goal . This definition does not prescribe what the choice is based on: on reasoning, on a personality model, on an emotional model - or, ideally, on all three plus other factors that have not yet been considered by the designers.

Whatever the definition of an autonomous agent, it is an engineering criterion: the user cannot know the decision space in which the agent she has to deal with is acting. It makes sense to discuss autonomy in replicants that can act outside direct control by their owner and are even capable of making commitments in his/her name (Cook et al. 1999). But the former definition of autonomy would apply to any adaptive tutoring system and is not specific to agents, whether visualized or not.

The story describes a pedagogic agent with a high degree of autonomy. It has to pursue several goals at the same time: Goals at the highest level are, for example, the student's progress and simultaneously, the student's entertainment. Each choice of lessons and types of exercises would very probably require a complex decision process, where the student's preferences have to be reconciled with the (in-built) plan of learning steps that would lead to an optimum learning outcome.

At this point, the human designers come in with their theory of learning: what types of activities in which order would lead the student to success are basic assumptions that are built into the agent. The agent ist necessarily a follower of one or the other learning theory. The fact that each (inter)action of the pedagogic agent is based on one or the other theory of learning - and therefore of teaching - is made explicit in the experimental learning environment ETOILE (Dillenbourg et al. 1993), where the different aspects of tutors even take the name of the researcher whose model they follow.

Communication with agents

Communication with agents is currently considered one of the big problems and obstacles in the field. Communication in spoken natural language as described in the story is still far from being mastered satisfactorily, especially in semantically open domains. At the same time, speech recognition and understanding is certainly an essential criterion for "interpersonal" communication with agents and therefore for social interaction with them.

Curiously enough, in the literature that I have consulted on the subject of pedagogical agents, this major problem is at once recognized and put aside. Compared to research work on aspects like deixis, facial expression, motion, emotions and personality models, verbal communication gets not much of the attention of the "agent community".

On the one hand, there are "chat-bots" either without visual appearance or represented by cartoon figures with only little animation and expression. Using different techniques and tricks, they can keep up conversations in writing. An advanced research project on an "embodied conversational agent" (Cassell, Sullivan et al. 2000) allows for verbal conversation in a specific domain, but its focus is on non-verbal behavior (see part III). Some projects get around the problem of verbal communication with human users by having two agents interact with one another (André, Rist, and Müller 1999; André et al. 2000).

Language learning as a field of application for a synthetic tutor is therefore highly unrealistic at the moment, but not only because of language processing.

It is not by mere chance that most intelligent tutoring systems can be found in sciences and mathematics. Researchers prefer the paradigm of tutoring and problem solving [Baumgartner and Payr 1994, 2000] to build their systems on, and these content domains seem to offer clearly defined problems of which not only the solution, but also the process that leads to it are known. The system needs a clear-cut "ideal" procedure and solution to compare the student's actions with and to give hints and support. The same is true for medical "cases", as in the ADELE example: while the student may have the choice of different diagnostic methods and their sequencing, the result or "solution" is unambiguous.

In the case of foreign language acquisition, there is no such clear-cut end result, but only the vague notion of "mastery" of the foreign language. Modern foreign language pedagogy has recognized this and, as a consequence, sees the main role of the teacher and the learning setting to offer rich, varied and adequate opportunities to gain mastery. How and to

what degree the student benefits from them cannot be determined from the outside - the teacher - but only by the student herself. The term "language acquisition" as opposed to "teaching" underlines this view of the learning process. The intervention of the human facilitator in the story acknowledges the fact that "mastery" of a language cannot be evaluated fully and holistically by a formal system. But with a certain degree of human intervention, I consider a virtual language tutor neither easier nor more difficult to implement than a mathematical one.

PART II

Problems and issues for a cultural studies approach to synthetic actors

Approaches and motivation

Constructing the user's point of view

The two stories have been taken from everyday life on purpose: research on avatars and synthetic actors has focused on computer games and MUDs/MOOs (Turkle 1998) or on fiction about them (for example Hayles 1999). Social science and literary studies need some "material" basis as a primary source to start from, an "object" to study - be it a work of fiction.

Computer science as a domain of engineering, on the other hand, lives from what does not yet exist, trying or at least promising to bring it into existence and proposing ways of doing just that. It creates its own "material".

This contrast may be one of the reasons why sociological and cultural studies always lag behind the changes that manifest themselves in society. they have to wait for things to have happened in order to study them *post festum*. This time-lag, however, becomes a problem when things tend to change at ever increasing speed, as they do in the information technology. Social science risks to look old and to lose its grip on everyday life (if it has ever had it). For example, what we can find today as sociological studies of the Internet leaves us strangely unsatisfied: statistics of Internet use, profiles of surfers and chatters, studies on media consumption and e-commerce. These approaches are by far not sufficient to get a glimpse into the very near future (or present for some of us) where cyberspace with all the possibilities of "virtuality" that are being programmed right now is a substantial part of our everyday culture.

"Everyday fiction" like the texts at the beginning of this report is a possibility to bridge this gap. The stories are fictitious, but as I have shown in the chapter on their technical background, not very much so. The technology is already there (and even beyond), but still in bits and pieces that do not yet fit together. The fictitious part of the story is the

piecing together of the various projects and products and to sketch the puzzle as it might look once some of the pieces are in place.

The picture might be wrong and remain wrong, technology might take a completely different turn somewhere, and the domain of synthetic actors could become a dead-end street, as many other domains in hardware and software technology have turned out to be. There might be barriers that prevent a breakthrough even if everything were in place. Use of synthetic actors could remain restricted to special groups and situations and never reach the broader public. There are any number of potential barriers I could think of: from complex psychological and cultural ones to the simple question of telecommunication cost and network capacities.

Desktop video-conferencing technology can serve as an example here. There can be no doubt that the technology is already there and has been so for a while already: tiny cameras and microphones, codecs (coder/decoder for video and audio signal) as plug-in cards or simply as software, easy-to-use control software, even (to some extent) standards that guarantee the interoperability of different systems. But desktop video-conferencing has still not made its way to the ordinary workplace or home computer, at least not in Europe. What are the reasons? It might be the price for a connection (at least when you use ISDN lines). It might be simply too unusual to use the PC for this purpose, and special, single-purpose equipment could make it on the market (contrary to the prophesies about convergence). It might be technical problems, or it might be that video-conferencing has too often been showcased with high-end equipment like large screen projections, several cameras, requiring a host of cameramen and technicians. It might also be that employees feel "down-graded" when they are not allowed to travel but have to use video-conferences instead (this problem has actually been observed). The breakthrough of desktop video conferencing could be just round the corner - but it could also never come at all. Or it could find uses designers and developers have never even dreamed of.

It is impossible to foresee whether avatars and virtual characters will really play an important role in practical computer use. However, I

consider it profitable for socio-cultural research in this area to imagine a situation where avatars and agents have become part of everyday life. For the story-teller, there is nothing new and thrilling about sending her representative to a meeting or about being expertly tutored by a virtual character. The technology, in this scenario, has become one computer application among many others - like word-processing, or e-mail, and can be studied on the same terms as it has been done for some of these applications.

This scenario makes it possible to bring in the user's point of view into a field where "real users" still are a rare species, and to create a subject for social science. To construct the user's point of view is essential for the kind of questions I want to study. Designers and developers answer or pretend to answer these questions implicitly by their design choices, often enough without being aware of the questions. They act out of their understanding of computer culture and users' needs.

A window in space and time

But first of all, they act out of their own cultural background. Culture, in this sense, is all that we need not think consciously about but that we experience as given and - paradoxically - as natural.

> One of the functions of culture is to provide a highly selective screen between man and the outside world. In its many forms, culture therefore designates what we pay attention to and what we ignore. (Hall 1981a)

Avatars and agents, being designed by people with a certain cultural background, come laden with features that seem so self-evident to their developers that they are not even able to think about them. When the users share large parts of this background, they will not notice this tacit dimension either. Soon enough, and by way of habituation both groups will take for granted what has been, in the beginning, a set of arbitrary choices.

There are two motivations for my attempt to probe into the cultural background of this technology:

First, a new technology opens a small "window of insight" as long as it is new and unusual. What Michael Heim wrote about word processing in 1987 - "The present moment is a felt transition" (Heim 1987, p. 10) - holds for so many other technologies and so many other moments in our fast-moving time that it becomes hard to even be aware of any specific turning point: transition is becoming the normal state of technology. Taking a look at an emerging cultural technology requires to privilege a certain point in time - the present moment - as a standstill. This is certainly an illusion, but nevertheless necessary. It is impossible to turn and look around while in full motion. Heim argued his concern with writing technologies as follows:

> It would be easier to judge or evaluate the phenomenon of word processing after some historical distance from it has been established, once detached, historical accounts of the computer revolution became available. ... A philosophical study of digital writing made five or ten years from now would be better than one written now in the sense of being more comprehensive, more fully certain in its grasp of the new writing. At the same time, however, the felt contrast with the older writing technology would have become faded by the gradually increasing distance from typewritten and mechanical writing. ... It is incumbent on us then to grow philosophical while we can still be startled ... (Heim 1987, p. 7)

And second, the fact that avatars and agents are, at present, developed by a male, white and predominantly North-American population of technicians offers me the chance to view it from the outside, at least in a few respects (I can't help being white, too): being female and European, where "Europe" here means first of all a continent with a huge variety of languages and cultures. Only in this sense of "European", I admit, can I find the necessary contrast that gives me a chance to take the outsider's viewpoint, because there is another aspect in which European culture is not really different from North-American, especially in the fields of research and technological development where globalization is creating a

homogeneous, albeit somewhat pidginized, language and culture worldwide. The fact that I write this text in English pays tribute to my being also part of this global scientific community.

Through this small window in both space and time, I can try to look at virtual characters and to reflect their development and use in an attempt to make the tacit dimension visible, at least in some aspects. The view from outside is instrumental in that it can reveal "strange" things to wonder about. Everyday culture being, by definition, the part of life that we do not and need not think about, is hard to see. Only its breakdown (Baumgartner 1993) offers the opportunity to make one or the other aspect explicit. The breakdown is experienced as a frustration, an unfulfilled expectation, or simply as a momentary loss of orientation. We all know the breakdown we experience when we are in a foreign country and people do not act as they "should" according to our assumptions. In this situation, we can find out something about other people's culture, but much more about our own, that is, what we expected and why.

The study of virtual characters from an intercultural perspective

A look into a future where virtual characters have become part of everyday culture is one such attempt at crossing cultural boundaries - not in space, but in time.

The fact that virtual characters have become part of the culture explains why the narrator does not and in fact cannot think about them in an existential way. Even if we look at this culture at its very beginning - with imperfect technology, with refusals and reservations still being made, etc. - certain questions are not raised any more by the person using virtual characters.

Only from today's viewpoint can we try to wonder about this other culture:

- Why are replicants used in a virtual meeting? What is their purpose in the eyes of their users (or owners)? What is their effect in our "outside" view?
- Why are synthetic tutors used in online courses?
- What practices are involved in using them? What sort of work, manners, discourse surrounds the use of virtual characters?
- What is the relationship between the avatar and its owner, the avatar and the other people it deals with?
- What is involved in experiencing the relationship with a pedagogical agent as social and "interpersonal"? What do users expect once they enter into a social relationship with a virtual character?
- What do virtual characters change in computer and Internet use? More specifically: in what does the agent-guided online course differ from traditional ones?
- What does it mean to have a virtual replicant?

Looking more at the present and at practical questions of virtual character design, the findings of studies on intercultural communication can equally be useful.

What is called "Internet culture" emerged from and simultaneously encouraged scientific and technological globalization, and in the beginning it looked certainly like an overwhelming push towards homogenization of regional particularities, its language being (sort of) English and the "electronic lifestyle" being North-American. But there is a growing number of Internet users who do not speak English, and whose use of the Internet is predominantly local, regional or national. The majority of users would not know what to look for in the Internet worldwide, being mainly interested in (national) news or (local) leisure time activities. Internet is being regionalized instead of its users being globalized, or more precisely: globalizing effects are limited to a part of

users, probably larger than it would be without the Internet, but still (or again) a minority in the cyberspace population.

The long-standing problems of "localization" in software engineering e.g. (del Galdo and Nielsen 1996) therefore will stay with us instead of dissolving into a uniform "cyberculture". With virtual characters who/that bring back features of face-to-face-communication and (tele)presence into a medium where, so far, text-based and "faceless" communication has avoided a number of potential sources for interpersonal and intercultural conflicts and breakdowns, the issues of cultural particularities on the non-verbal level will play an important role in acceptance and evaluation of interfaces, e.g. (Hall 1981a, 1981b, 1982):

- *Space*: for example in interpersonal distances that have very fine-tuned and specific meanings in different cultures, as have dimensions and structures of virtual spaces, motion in space and relative to other persons.

- *Time*: speed and duration of interactions, attitudes towards schedules and delays, etc.

- *Context*: Hall distinguishes "high-context" from "low-context" cultures, whereby he resumes deep-seated cultural differences in the structuring of social relationships, institutions, work, families, indeed everyday life.

Things may become even more complex when we look at the level of verbal communication. Far from being the clear and logical means of communication that grammaticians tend to foreground,

> Socio-cultural conventions affect all levels of speech production and interpretation from the abstract cultural logic that underlies all interpretation to the division of speech into episodes; from their categorization in terms of semantically relevant activities and interpretive frames, to the mapping of prosodic contours into

> syntactic strings and to selection among lexical and grammatical options. (Gumperz 1982, p. 186)

Theoretical foundations for the study of virtual characters

Should virtual characters be studied as artifacts or as humans? The question seems trivial: they are artifacts. However, humans do not interact with them as if they were artifacts. Even for simple dialogues with software applications, it has been shown that users tend to consider them as social partners and not as objects or tools (Reeves and Nass 1996). Evidence for this anthropomorphization can be found in behavior that was, in the past, reserved for persons.

Reeves and Nass did experiments with a range of settings - from computer hardware to TV programs and cartoons. The smoothness of the experiments and the homogeniety of the results raise a number of critical questions with regard to the method, but also considering the hypothesis where all these phenomena are subsumed under the concept of "media" and contrasted to a notion of "real life" that has long since become obsolete. This view is what Heim calls "naive realism"[1] and contrasts with "virtual realism" (Heim 1998, p. 37).

One can reproach naivity to the authors or any number of other things, but the statement they make - the "media equation" - can raise doubts whether the opening question can or should be put this way. We have to look at software in general and at virtual characters in particular as simultaneously artifacts and social beings. No anthropomorphizing of virtual characters is involved from my side if I insist on studying communication with them as social interaction: users enter in social interaction with them, and taking the user's point of view implies following this turn. Virtual characters are, beyond doubt, artifacts: products of software engineering. But from the moment the user meets

---

[1] Naive realism in his sense is not to be confused with the philosophical term e.g. (Searle 1983) where it stands for equating the outside world with our perception of it.

them, they are also "persons". Whether the user considers them a special kind of persons and the communication with them a different kind of interaction than with humans, remains open to speculation at this point in time. Reeves and Nass suggest that there would be no difference. This would imply that users implicitly bring all their rules of communication and even more tacit expectations to the encounter with virtual characters. Given that designers and researchers, by trying to model more and more aspects of personality in their virtual characters, make an effort to reach precisely this goal, it seems legitimate to view the relationship between users and virtual characters as an interpersonal one and to apply methods from the fields and disciplines that deal with this relationship. (this is the approach taken in part III)

However, the fact that virtual characters are artifacts opens a perspective that should not be ignored. It may seem paradoxical to adopt an approach that insists on the material object in a first step only to leave it again in a second one on the grounds that the object in question is not simply an object, but also a social actor. But only such an approach makes it possible to study them in a way related to the study of "material culture", based on the ethnological tradition. Only in this way we can avoid the trap to view virtual characters only in "present tense", that is, in the moment and the way in which the user has to do with them. They are products of technology, and as such already bring with them a history of design, goals, assumptions, limitations and - again - cultural background which are those of the developing community and the society in which this community works, gets funding, and/or hopes to market the product.

A socio-semiotic approach could help to analyze the processes of meaning production that can occur in the relationship between user, object and producer (Gottdiener 1995, p. 180 ff). Socio-semiotics, in Gottdiener's definition, overcomes the idealistic bias of postmodernism by re-introducing the object in its material existence and its everyday functionality into the semiotic analysis:

> According to socio-semiotics, any material object constitutes the intersection between social context and the codified, connotative ideologies of social practice, on the one hand, and the material, objective, production or design practice which produces the object world, on the other. (Gottdiener 1995, p. 56).

This approach appears promising for a study of virtual characters as objects. It allows for their study in terms of objects of everyday use and, at the same time, as signs with all the processes of signification and connotation that are implied by this term, and with the creativity that users can apply to products in order to adopt them to their (sub)culture. Although this approach stresses social interaction as the basis of semiotic processes, it obviously does not extend social interaction to take place *with* the object - as is the case with virtual characters. It will be interesting to study the additional complexity of the tripartite relationship when this process is taken into consideration.

To conclude this discussion - for the moment only - I will take a pragmatic view. Trying to determine approaches and methods from the outset can be helpful for getting a "handle" on a new field, but it is also dangerous, in that it narrows down perception . Too much is happening in this field to allow any individual researcher to pretend that she/he has already put the phenomenon into its corresponding epistemological box (for an example of "pre-judice" in this sense cf. (Schefe 1997)).

Issues for further study

At the end of this part, I will resume the domains and questions that need further study and, at the same time, outline further project activities that should contribute to that end.

Theoretical issues

Personal avatars in use

It will be highly interesting to follow up on the acceptance and use of personal and/or personalized avatars:

- Are people in fact interested in replicating their physical appearance, or would they prefer phantastic or at least "embellished" versions of themselves?
- How does the option of lifelike replicants fit with Turkle's findings about the use of chat-rooms, MUDs and MOOs as a playground for experimenting with multiple identities? (Turkle 1998)
- Or will the widespread use of lifelike replicants confirm the opposed view held by some sociologists (especially feminists) that society in the Internet is just "more of the same", i.e. replicates class, race and gender discrimination?

As the commercial application of personalized avatars is still in its infancy, it is expected that this question cannot be pursued fruitfully during the lifetime of this project.

Social interaction with virtual characters

A whole range of research issues comes up when the interaction between humans and agents is considered as a social interaction (and not one between subject and object). With "Embodied Conversational Agents" (Cassell, Sullivan et al. 2000) being designed to interact as naturally as possible with human users - through language, gestures, gaze, deixis - virtually the same questions that have been raised about human social interaction arise for the "hybrid society":

- How do subjects achieve coordination of action through communication?

- How do manipulation and the exercise of power translate into verbal/nonverbal behavior? Is interaction without them even possible?
- Why do "breakdowns" in communication occur and what can be done to repair them?
- What is the role of the cultural background in ensuring the possibility to communicate? How much of it is culture/group specific?
- If we assume that social interaction is constitutive and/or reproductive of social structures, what does human-agent-interaction reproduce and/or constitute?

The inquiry into this question has been taken up in part III of this report, but the problems are far from being discussed exhaustively. It will be suggested there that Goffman's level of study, the "order of interaction", might lead out of the current dilemma where agent research concentrates either on 1:1 communication (reducing it to "conversation") or on rough simulations of society without the fine-grained detail of communicative action.

Non-verbal communicative action

There appears to be a gap between the study of nonverbal communication and the study of interaction (in the narrow sense of the word): while the first is the realm of conversational analysis and psychology, sociological action theories concerned with the second start out from the "rational man" model which implies the sole use of language in communication. "Real" conversation, however, involves a large nonverbal part.

The attempts of the agent research community to model and implement one or the other aspect of nonverbal behavior in virtual characters draw our attention to open problems in this domain:

- What does nonverbal communication contribute to the interaction?

- To what degree is nonverbal interaction rational and communicative, to what degree is it strategic (in Habermas' sense)?

While communicative "breakdowns" on the verbal level still open the possibility of verbal repair (question, challenge, meta-linguistic discourse), repair work becomes much more difficult if not impossible for the nonverbal parts of communication.

- What are cues for a breakdown in nonverbal communication?
- Which elements of nonverbal behavior can be expressed verbally and therefore brought into (verbal) discourse?
- Is there anything like "nonverbal repair work"?
- Which culture-specific nonverbal behaviors can lead to breakdowns in cross-cultural communication?
- How can breakdowns in the nonverbal channels be avoided or repaired in virtual characters?

At the end of part III, a first approach to this vast area of study is made by analyzing gestures from the perspective of communicative rationality. The first subgoal is to draw a line between "rational" nonverbal communication which can be - in principle - expressed verbally and thus brought into the range of argument, and "non-" or "pre-rational" nonverbal communication.

Research agenda of the scientific community

Heaton (Heaton 1998) did a comparative study of software design for CSCW (computer supported collaborative work) in Denmark and Japan. She shows that the development teams in both cuntries pursue widely different goals in design: while Danish engineers focus on democracy and quality of working life, the Japanese pay most attention to the quality of communication and on technical sophistication. The cultural differences

are ubiquitous not only in the design process, but also in the resulting products:

> The preferred Japanese approach to CSCW design is to provide a channel for communication, which can be used to complement, or supplement, traditional ways of working. This channel should transmit as much information as possible ... but should avoid specifying procedures or ways of doing things. ... In contrast, Scandinavian CSCW systems focus on providing frameworks for collaboration that will facilitate communication and decrease social distance between all those using the system, thereby improving the quality of working life. (Heaton 1998)

This article illustrates a point that is of interest to this study, too: cultural differences should be looked for even before the design process starts. They may be found already in the goals and priorities set in the research agenda of the institutions involved in research on virtual characters.

Dis-embodiment and Re-embodiment

Another question has not yet been discussed at all in the present report: it is the fascinating phenomenon that there seem to be diverging trends in Internet development. On the one hand, the disembodiment of information (Hayles 1999) and, consequently, of Internet users, has been underlined repeatedly as a prominent feature of "cyberculture" (Turkle 1998). On the other hand, avatars and synthetic actors point towards a "re-embodiment" even if this body is virtual: is this a contradiction or a reversal of trends? Does this development address completely different user groups or applications?

PART III

Agents as Actors

This part of the report is concerned with human-agent-interaction as social interaction. It can be seen as work that prepares the study of cross-cultural human-agent-interaction. Several steps are necessary to approach that subject. First, the notion of "believability", dear to the agent community, has to be replaced by a concept that better reflects the attitude of users (= human partners in the interaction), namely "agency". Second, a connection has to be made between converstion and action in order to analyse the phenonema in terms of interaction in the sociological sense of the word. Habermas' theory of communicative action (Habermas 1999, quoted as TKH I and TKH II in this paper) is introduced shortly to that purpose. It opens the way for the third step, namely to discuss the question of rationality in human-agent-interaction. Conflicts in inter-group or inter-cultural communication are, in most cases, no simple "misunderstandings" that could be solved by way of question and argument. Instead, they arise from behavior (most of it nonverbal) that is obviously outside the limits of communicative rationality and is grounded in the actors' lifeworlds. A concept of communicative rationality is therefore necessary to draw this borderline - which, as is shown at the end of this part,taking gestures as an example, lies not between verbal and nonverbal communication, but divides behavior in each of the "channels" of nonverbal expression.

On Believability

Nearly every researcher or developer in the field of embodied agents mentions as his/her goal the creation of "believable" characters.

Where does the concern with believability come from and what is, in the minds of agent developers, involved in it? Following Bates (1994), believability of lifelike agents is the extent to which users interacting with them come to believe that they are observing a sentient being with its own beliefs, desires, intentions, and personality (quoted by Lester et al. 2000, p. 127).

> There is a notion in the Arts of "believable character". It does not mean an honest or reliable character, but one that provides the illusion of life, and thus permits the audience's suspension of disbelief. The idea of believability has long been studied and explored in literature, theater, film, radio, drama, and other media. (Bates 1994), quoted in (Rizzo et al. 1999)

The goal of creating believable characters is rooted in the dramatic arts, and came into the field of embodied agents mainly via research projects on virtual theatre. Hayes-Roth (Hayes-Roth, van Gent, and Huber 1997) makes an important distinction between the agent as individual and the agent as actor.

> We are studying personality in the context of synthetic agents that function as actors. Like human actors, these agents assume prescribed roles. ... In this context, it is not the personality of the actor that interests us, but the personality of the character he or she portrays. Thus, when we say that an actor is "in character", we mean that the actor is behaving in accordance with a personality created by an author, shaped by a director, and assumed by an audience, for purposes of a particular performance. (Hayes-Roth, van Gent, and Huber 1997, p. 92)

It is the character as an artistic creation that raises the challenge of believability, not the actor (as a personality) as such. But it becomes a challenge only in specific settings that are clearly separated from "real life". The theatre and the cinema are such specific settings, where everyday life is suspended through architecture, decoration, and orchestration of the event. The challenge is to restore believability to characters inside and despite this special setting. Only when entering the theatre and the cinema, the audience's main attitude is disbelief that has to be overcome by the actor's performance. Our attitude in everyday life, however, is not one of disbelief.

On the contrary: our everyday attitude is one of belief. We do not ask ourselves, when we see people in the street, whether they are enacting a dramatic character. We take them for ordinary people, and we expect

them to be what they appear to be. If they start acting in a strange, inexplicable way, we will come up with all sorts of explanations (the person is mentally deranged, heavily drunk, has suffered a severe shock etc.) before we even consider that it might be a dramatic performance. We are ready to draw on any kind of experience and assumption to maintain or restore our belief. It should be - and actually is - much harder to suspend belief than to "suspend disbelief".

The issue should be completely different once we deal with artifacts. Artifacts - especially virtual characters that, up to now, exist only on small computer screens - should not be considered living beings. Therefore, designers would have to strive to make them believable.

Reeves and Nass (Reeves and Nass 1996) have shown that this is not so. People's interactions with computers are "fundamentally social and natural, just like interactions in real life" (p. 5). They have these interactions independently of age, knowledge, or concentration, and of what social group they belong to - children, college students, or technology experts. It is as if they could not help acting this way:

> People respond socially and naturally to media even though they believe it is not reasonable to do so, and even though they don't think that these responses characterize themselves. (p. 7)

People are not aware of their anthropomorphizing practice in interacting with media, maybe they would even deny it when asked openly. Evolution might provide us with an explanation for this behavior. Until very recently, all things that humans had to deal with were physical objects, and all persons real humans. Anything that acted socially really was a person and could not be anything else.

> Because these were absolute truths through virtually all of human evolution, the social and physical world encouraged automatic responses that were, and still are, the present-day bases for negotiating life. (p. 12)

Believability, then, is easy to achieve in an agent. One could (and many would) object that no machine so far has been able to pass Turing's imitation game (or Turing test), it is therefore still possible to distinguish between man and machine. But this is not the same claim that is being made when we consider a virtual character "believable". The Turing test probes into the capacity of a program to be mistaken for a human. This is a much stronger claim in some sense, while, in another sense, the setting of the test narrows down the criteria for "humanness" to believable processing and production of written language.

Judith S. Donath (Donath 2000) reports an "encounter" between a woman and "Julia", a chatterbot (see page 73). At first, the woman really takes Julia for another human member of the chatroom. But even after finding out that Julia is a program, she continues to communicate with the chatterbot in a "humanlike" fashion. In practice, then, the Turing test does not make a decisive difference in how humans interact with agents. Why? The simple answer is: the woman simply can't help herself communicating in any other way. The communication does not change radically once she stops believing that she is chatting with another human being. She has lost this belief - but obviously only to believe something else, at least that there is an object/being that will understand her utterances and give more or less reasonable answers. She is convinced that there is some "other" capable of communicating and reasoning in a rational way. And what else should or could the user confronted with an agent believe?

An agent is an agent ...

For the user to "believe", it does not make any substantial difference whether she is conversing with a talking animal, a robot, another human being or a software agent: there is something all of these have in common that can make them a valid counterpart in interaction. The concept that best describes what they have in common is "agency".

To talk of "agency" instead of agents makes it easier to consider it a property that can be attributed to an entity to a certain degree instead of an indivisible and inalienable characteristics of human beings (Rammert 1998). The clear distinction between the human being as the only possible agent - and therefore capable of becoming a subject, while all the "rest of the world" can have no more than object status - vanishes.

While this perspective allows us to attribute agency to non-humans like machines, robots and programs, it also puts into question the formerly untouchable agent status of human beings. On the level of everyday routine action, the level of autonomous action is restricted even for humans. The woman at the cash-desk of the supermarket or the newspaper vendor have not many degrees of freedom to plan, decide or act individually in their work. Division of labor into fragmented routine jobs is grounded in the belief that too much freedom would be detrimental to productivity. The domains where action has been successfully reduced to routine are exactly those that can be easily automated: the boundary separating human agents with only a tiny margin of agency from machines achieving more and more agency as they get "smarter" cannot be sharply drawn. "Human and technical agency are similar in that they both include repetitive characteristics and are only temporarily emergent" (Rammert 1998).

This conclusion may be sufficient for the field of Distributed AI or "Sozionik" as it is currently called in German, as it is not concerned with humans and machines as social "beings", but with "studying and modelling artificial sociality" (SPP-Sozionik n.d.). It focuses on multi-agent-systems and is interested in:

- creating software for intelligent collective problem solving following models of social coordination of action
- contributing to sociological knowledge by modelling artificial sociality
- exploring the new relationship between humans and technology, e.g. in "hybrid communities of cooperation".

The authors point out, however, that the third issue is only emerging as a field of research. "Macro-level" models are currently in the center of research efforts.

The micro-macro-problem

The micro-macro-problem of sociology can be formulated as follows: Are social structures the result of individual actions and strategies (micro perspective) or is individual action determined by social structures (macro perspective) - or are actions and structures mutually dependent? The field of Distributed Artificial Intelligence (DAI) clearly starts out from a structure orientation, especially where the modelling of societies is concerned, while the micro-level is considered mainly under the aspects of communication among agents in multi-agent-systems. Research on "hybrid communities", where necessarily individual agents communicate with individual users, is still in its infancy. While sociologists are aware that the micro-macro-problem exists in agent societies as well as in hybrid societies, and expects contributions from Artificial Intelligence to it in the future, the interdisciplinary field of DAI largely leaves out the micro-level at the moment.

The separation, however, goes deeper than that between different schools of social theory. DAI research on the one hand and research on embodied agents on the other replicate roughly the disciplinary divide between sociology and economics (as disciplines dealing mainly with collectives) on the one hand and psychology and linguistics (as disciplines dealing mainly with individuals) on the other.

From the viewpoint of the interaction between the individual user and the (embodied) agent, the conclusion about "partial agency" is therefore not entirely satisfying. It is a theoretical explanation that helps sociologists to overcome the separation between software and human agents for the purpose of studying artificial societies, but it has no explanatory power for the actual interaction taking place. We have (as yet) no awareness of nor skills in dealing with "partial agents". A software agent is an agent in the sociological sense when it is attributed

agency by the human user or partner. In order to avoid repeating this clumsy specification, I will use the term "actor" for the "agent in the sociological sense" (meaning humans as well as software agents) in what follows.

While "believability is in the eye of the observer" (Dautenhahn 1998), agency is in the mind of the subject. We attribute agency readily and to a lot of things, not only to pieces of software or robots. This certainly made sense for the survival of our ancestors: when you meet a lion, it is useful to think of it as a being that is able to act according to its own goals. Being able to imagine these goals and to foresee the possible and probable actions of the animal may have been of great advantage for the human. However, the event of being attacked and eaten by a lion would hardly count as an interaction.

Something more than agency, then, is needed. Mead (in *Mind, Self and Society*, 1934, quoted in TKH II) sees the essential difference in the ability of symbolic interaction. Symbolic interaction means, in very short terms, that both partners in an interaction share the meaning of symbols, which allows them not only to foresee the other's (re)action, but also to put themselves into the role of the other. At that moment, the relationship becomes one between two subjects instead of a subject-object-relationship. (To my knowledge, sociologists have never cared very much about the relationship with animals - maybe this would have provided us with insights on the relationship with non-human actors long before the advent of computers). "Taking the attitude of the other" means that the subject recognizes the externalised subjective in the other.

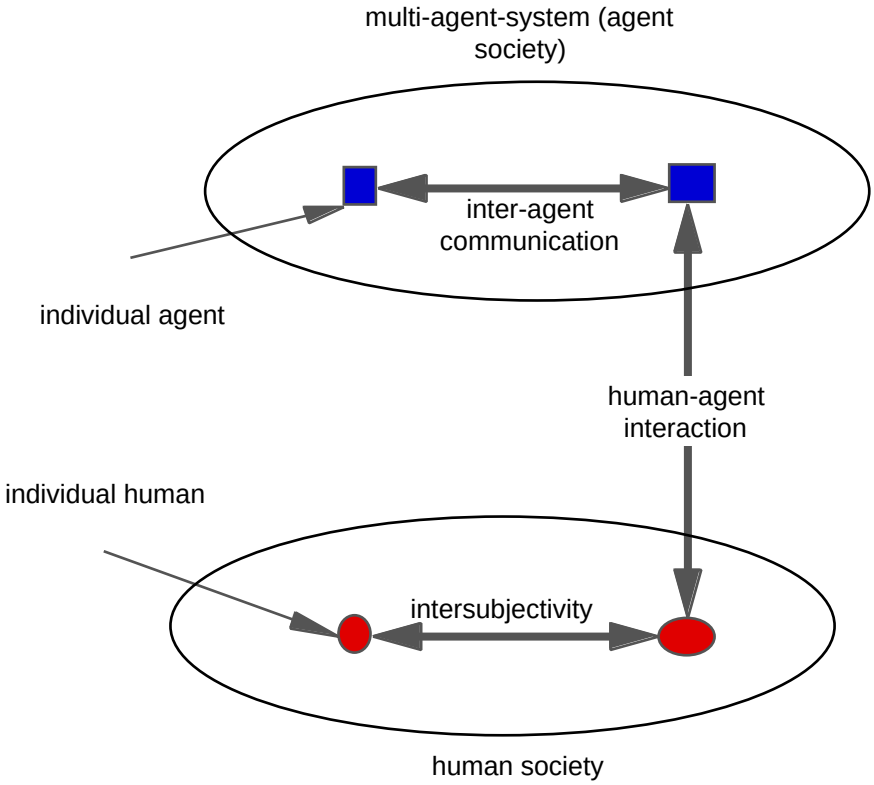
Mead's concept of intersubjectivity is at the basis of his theory of how social structures can emerge. For him, the point of articulation between the micro and the macro level is the "generalized Other": Through intersubjectivity, the individual is able to take over the norms and laws of society through the interaction with other individuals, e.g. when a child interacts with its parents. Finally, the norms do not appear as external (arbitrary) constraints and exercise of power anymore, but are internalized: the individual carries in itself the concept of a general

"other" which represents the normative power of society. "Over against the protection of our lives or property, we assume the attitutde of assent of all members in the community." (Mead.1934, quoted in TKH II, p. 63).

Habermas takes Mead as one of his points of departure for his "Theory of Communicative Action" (Habermas 1999) which also belongs to the "constitution theories" of society. He criticizes Symbolic Interactionism in several respects, e.g. in its traditionalistic view of norms: Mead's theory leads to a picture of a "conformist" society based on consensus. With each member having internalized the generalized Other, there would be no necessity of negotiating norms and no possibility of conflict except in the adolescent in the process of developing the generalized Other. The post-traditional view of norms is linked to the concept of communicative rationality which presupposes that the members have and pursue their own interests.

The order of interaction

The following schema is based on Rammert's attempt to structure the different types of interaction. (Rammert 1998)



He suggests the notion of "intertextuality" for inter-agent-communication, motivated by the fact that it is an interaction between programs, which are, basically, texts. However, defining software agents through their "substance" is a reduction that raises more problems than it solves. By reducing agents to "texts", the author would have to reopen the discussion on agency in order to establish how (much) agency can be attributed to them. It is, in my view, preferable to limit the definition to the level of the "agent" as an entity - there are enough problems involved here: a software agent can be (and most often is) a program written explicitly to do one specific task and nothing else. A complex agent, e.g. a virtual character acting as a pedagogic agent, can contain numerous software agents brought together in a single interface. It is itself a multi-agent-system or "society of agents" (in analogy to Minsky's "Society of Mind").

While each single agent in this system can count as a member of the agent society on the ground that it is able to interact with other agents (thus having "agency" with respect to other agents), only the compound interface agent can interact with the human user. In this paper I will not pursue the possible implications of such a layered concept of agency, but will concentrate on the layer that is of interest here, namely the human-agent-interaction. For this purpose, agency as attributed by humans in and through interaction is necessarily restricted to agents with an interface - be it "traditional" or via a virtual character.

The other problem that is made clear by this diagram is that it allows for a society only on either side of the screen, so to speak. For each of the societies, interactions are supposed to have "constitutive" character (while not ignoring that interaction equally reproduces society), while the interaction between the agent and the individual is without consequences. This reflects well the current preoccupations of research, where hybrid societies as yet play a marginal role.

At the moment, we have to leave the question open whether the micro-macro-distinction is at all appropriate as an approach to hybrid societies, especially where interface agents are concerned. Possibly, an approach based on a "meso-level" of various collectivities could be more fruitful. Goffman's "order of interaction" could offer such an approach, as it is concerned with face-to-face-interaction, but in the framework of customary situational settings that give the interaction structure and that are maintained and reproduced by the individuals (e.g. board meeting, funeral, queuing etc.) Goffman underlines that these orders of interaction, while being coupled to the social order at large via "interfaces", are neither simply constitutive of it nor fully pre-determined by it, but that they are a level of sociological study in its own right (Goffman 1994). This "meso-level" would certainly merit more in-depth study with regard to human-agent interaction. As the mention of constitutive theories and the micro-macro-problem were introduced in this paper in order to clarify the distinction between communication and action, it has to be left aside for the moment.

From communication to action

Communication between embodied agents and users is currently a prominent research topic. With the "embodiment" of agents, it becomes clear that not only verbal communication is needed, but also appropriate non-verbal communication.

Communication in embodied agents

Verbal communication

Natural language processing has been making slow but steady progress over the last 30 years. Speech synthesis is far from perfect, but possible, while speech recognition and understanding have been shown to work in restricted content domains (e.g. automated information on railway timetables). When it comes to keep up conversation even when the limits of "understanding" are reached, the "tricks" that Weizenbaum already used in ELIZA still work, e.g. answering a question with another question, requesting the user to tell more or to talk about herself, or changing the subject.

The so-called "chatbots" or "chatterbots" are conversational agents that first turned up in (text-based) chat-rooms. The most famous example is "Julia" (Foner 1999).

Julia functions in a MUD (multi-user domain), a text-based, multi-person virtual environment and acts like the human players. That means, for example, that she is always located in one particular room of the virtual world, and a player ("mudder") can only have a conversation with her when being in the same room. She enacts not only another player, but serves useful functions for the human players: the program keeps track of all players and can tell when they last logged in and where they are, she can relay messages to players who are not online, and she keeps track of rooms and visitors, being able to give directions and recommendations about where to go.

As to the language processing capabilites, Julia mostly relies on pattern-matching frequent commands and queries:

> What makes Julia's conversation skills so powerful, as 'bots go, is that these sets of patterns were very cleverly grouped into useful equivalence classes by her programmer. Hence, the most common possible utterances that Julia might hear were enumerated this way, grouped appropriately, and such clever but brute-force programming proves sufficient for almost all cases. (Foner 1999, p. 337)

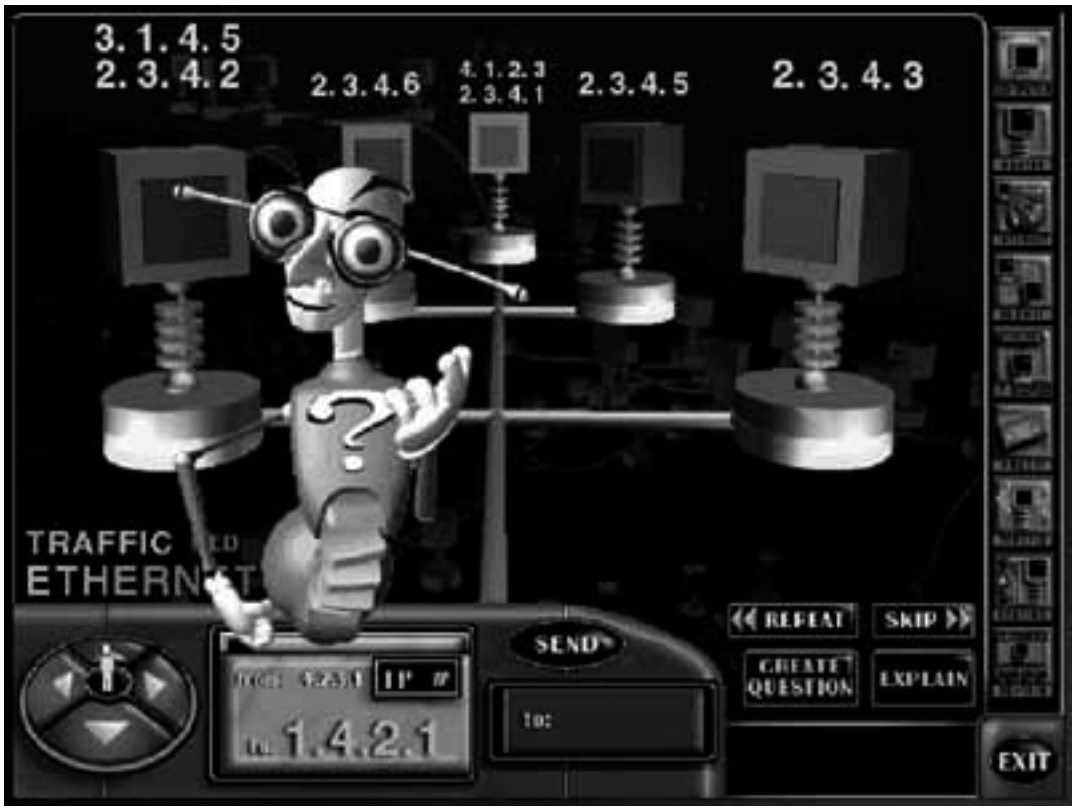
In those cases where this mechanism fails, Julia employs an activation network encoding some simple real-world knowledge. The pattern-matching and the equivalence classes can be fine-tuned by the programmer over time. For this purpose, the "fallback" responses by the chatbot - like "So?" or "Go on" as well as the customary excuse in MUDs "could you say that again, it scrolled off my screen" - are logged, so that more and more utterances can be "trapped" by the pattern matcher.

Julia's performance is convincing enough to leave human players unsure, at least for some time, whether she is a robot or a human. And even those who know that they are communicating with a robot cannot stop themselves from developing a social relationship with Julia (Donath 2000, Turkle 1998). The environment in which this robot operates certainly plays an important role in its "success" at passing as human: first, on the technical side, the text-based virtual world already abstracts away from any characteristics of appearance or voice. And second, on the level of discourse, MUDs are entertainment applications where players come together to chat without consequences. Human players use them to experiment with gender or role switching, to act out aspects of their personalities and behaviors that they suppress in real life. So "mudders" expect and tolerate behavior from other players that would appear to them strange or offending outside the MUD. There is therefore a wider margin for "strange" behaviour shown by the robot without it falling out of a valid enactment of the mudder's role. But, as we have shown before, it is not even necessary for Julia to be mistaken for a human to pass as a valid actor in her world.

Deixis

Deixis means reference to the situation in which the discourse takes place: both speaker and hearer share knowledge about place, time, objects and persons present in their situation so that they can refer to them (non-verbally) by pointing or (verbally) by indexicals like "over there", "this one", "now" or "you". Deixis covers the three dimensions of location, time, and identities.

Spatial deixis is a focus in Cosmo, a lifelike pedagogical agent that inhabits the Internet Advisor, a learning environment for the domain of Intrenet packet routing. Cosmo is a phantasy creature bearing antennas who hovers about the virtual world of routers and networks, providing advice to students as they decide how to ship packets through the network to specified destinations (Lester et al. 1999, 2000) .



Cosmo, the Internet Advisor
(http://www.csc.ncsu.edu/degrees/undergrad/Reports/jlelliot/thesis97.html)

Cosmo disposes of four "instruments" to achieve unambiguous and pedagogically sound deixis: gesture, locomotion, gaze, and speech. The referents to which they are applied are objects in the virtual world, e.g. routers or buttons. The deictic system goes through three phases to produce an utterance with coordinated nonverbal deictic elements.

- *Ambiguity Appraisal*: The two previous utterances are examined for occurence of the referent in question and for the number of other referents. Ambiguity or potential ambiguity is detected when the referent is either new or one among a number of referents mentioned previously.

- *Gesture and Locomotion Planning*: If (potential) ambiguity has been detected, the system will plan a pointing gesture. Several conditions are examined to decide whether the pointing gesture will be accompanied by a movement of the agent toward the object: if the object is small, if the object is near other objects that have been mentioned in those two previous utterances, or if the object is similar to other nearby objects of the same ontological category. Gaze is not used separately, as the agent will always look where it points to.

- *Utterance Planning and Coordination*: the mechanism used for appraising ambiguity also provides information about the expressions that will be used to refer to the object in question: a non-ambiguous referent will be referred to with "it". Data from the world representation and the results from the deictic planner lead to the decision between proximal and nonproximal demonstratives (this/that), while "this one" or "that one" are used when the agents points out one of a group of nearby objects of the same category. Verbal expressions and gestures/movements are timed so that the utterance sets on when the agent reaches the apex of its pointing gesture.

Nonverbal communication

A group of MIT researchers (Cassell, Bickmore et al. 2000) is currently developing an embodied conversational agent implementation called Rea (for "Real Estate Agent"). The agent plays the role of a real estate salesperson who interacts with users to determine their needs, shows them around virtual properties, and attempts to sell them a house. Rea has a fully articulated 3-D graphical body and communicates using both verbal and nonverbal modalities. She is able do describe features of a house using a combination of speech utterances and gestures, and can slo respond to users' verbal and nonverbal input. The system consists of a large projection screen on which Rea is displayed and in front of which the user stnads. Two cameras mounted on top ot the screen track the user's head and hand positions in space.

Rea, the Real Estate Agent. http://lcs.www.media.mit.edu/groups/gn/pubs/hi.htm

The architecture of Rea is based on the distinction between communicative goals, functions and behaviors: multiple communicative

goals (like initiation or turn-taking) are conveyed by communicative functions (e.g. taking or giving turn, reacting to a new person) which in turn can be expressed by communicative behavior in one or several modalities (speech, gaze, gesture etc. or a combination of these).

The input system detects gestures, onset, pauses and cessation of speech and contains a speech recognition module that transforms audio-input into text. Input is processed so that "interactional" behavior is distinguished from "propositional" behavior (both can be verbal or nonverbal).

The core of the system is the Decision Module, since all of its inputs are input discourse functions describing user actions, and its outputs are output discourse functions for Rea to execute.

> Once the speech act template has been selected and filled in [by the Understanding Module, SP], it is sent to the Decision Module that then needs to evaluate its effect and choose a response. The evaluation may update facts in the dynamic knowledge base and/or create an obligation that the agent needs to attend to. The agent can then perform simple plan reasoning to come up with one or more speech acts to achieve the obligation or communicative goal. ... When it is time to act, the relevant speech act template is filled out and handed to the Generation Module for realization, along with any interactional functions that need to be executed in order to contribute successfully to the conversation. (p. 51)

The principle of the architecture - distinction between communicative functions and behaviors - makes it possible to generate flexible ouput by leaving open, until realization, which type of behavior (in particular, verbal or nonverbal) is chosen to convey the contents and interactional components of an utterance. Gestures may stand in for verbal expressions or they can accompany them, making the output multimodal.

It is obvious that the realization of such an utterance, which has to take into account not only the verbal context, but also the current state of the agent and the virtual world, is in itself a complex system, especially

where "naturalistic" gestures synchronized with speech output is concerned.

The concept of speech act is used in theis project mainly to categorize input and to produce adequate output utterances. Smooth and seamless conversation are the central concern to which they are subordinated. However, conversation is not a goal *per se*.

The decisive step from conversation to action is to part with the belief that "utterances convey thoughts" (Cassell, Bickmore et al. 2000). If it were so, the salesgirl at the baker's could accept my statement "I would like four rolls and two doughnuts" with a simple nod - instead of springing into action to get the rolls and doughnuts, pack them, tell me the price, collect the money, and hand me the bag.

Communicative action

Speech acts

The classical way of looking at communication as action is speech act theory (Austin 1965; Searle 1969, 1979). The decisive innovation in speech act theory was to logically separate the proposition from the utterance. A proposition is an abstract statement with truth values, which is uttered with a certain illocutionary force, e.g. of promise, command, or assertion. Utterances like "I herewith close the meeting" show best its implicit character of "action": the utterance has indeed the consequence that the meeting is closed. Such an utterance does not have truth conditions, but "conditions of satisfaction" and is neither true nor false, but can be felicitous or not. In the example, the utterance must indeed be made in a meeting, and the speaker must have the authority - most probably as the chair of the meeting - in order for the speech act to be felicitous. The crucial point is that it is not merely an utterance, but brings about a change in the state of the world: the meeting is closed. It can therefore legitimately be called an action.

While Austin tied the typology of speech acts to the types of verbs used to express their illocutionary force, Searle moved on to a classification on the basis of the goals and intentions that the speaker pursues independent of the linguistic form in which these are expressed. His list of speech acts contains:

- constatives (or representative, statement, assertion)
- commissive (declaration of intention, promise)
- directive (rule, command, request, invitation, question)
- declarative (declaration of war, dismissal, nomination)
- expressive (expression of joy, compassion etc.)

The theoretical motivation of the categorization is the difference of "direction of fit" between statements and facts in the "world" (where "world" is "everything that is the case"): either a world-to-word-direction (an action to "make" the facts fit to the words, as in commands) or a word-to-world-direction (an action to make words fit to the facts, as in assertions).

Speech act theory has often been criticized, especially by sociologists, that it takes the perspective of the speaker alone.

> Allerdings beschränkt er [Searle] sich auf die Perspektive des Sprechers und läßt die Dynamik der Verhandlung und intersubjektiven Anerkennung von Geltungsansprüchen, also die Konsensbildung, außer acht. Für die intersubjektive Beziehung zwischen Kommunikationsteilnehmern, die sich miteinander über etwas in der Welt verständigen, läßt das Modell der zwei sprachlich vermittelten Beziehungen eines einsamen Aktors zur einen, objektiven Welt keinen Platz. (TKH I, p. 433)

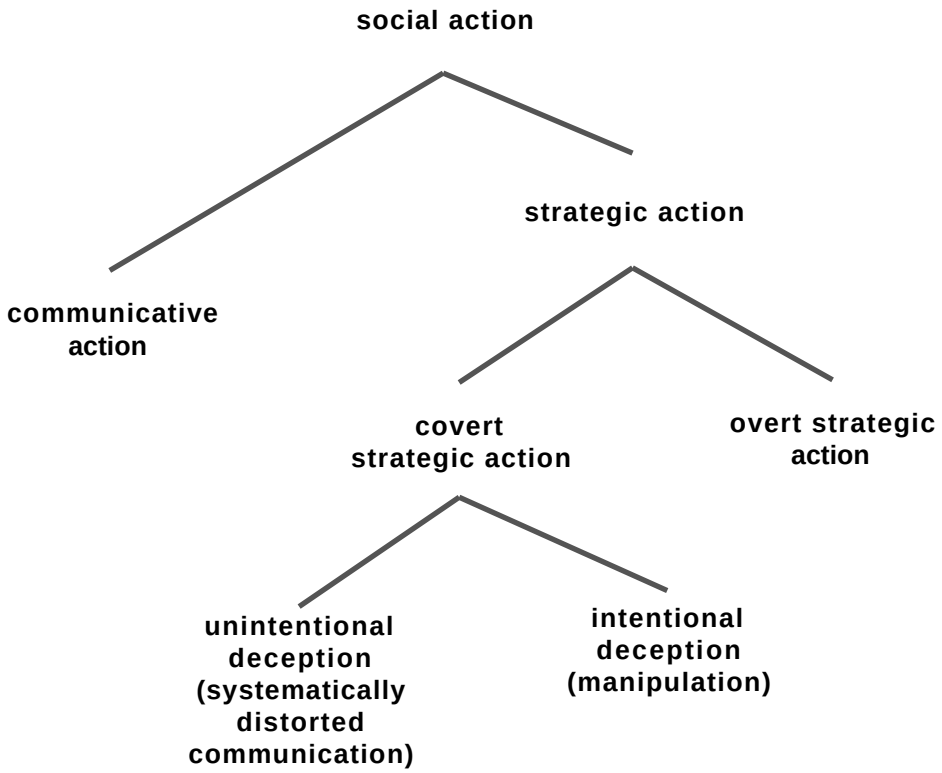
Habermas (TKH I, p. 429 - 435) criticizes Searle's theory in two further respects. First, he remarks that the conditions of satisfaction of a speech act that concerns some future action or event must not depend on the

realization of that event: I can make a valid promise (sincere, and regarding a future action that I am capable of undertaking) and still be unable to fulfill it for a number of reasons. The commissive speech act, however, is felicitous already at the moment I make the utterance. Its validity does not depend on my actual future action. Habermas claims that there has to be a distinction between the conditions in which an utterance is *valid* and those in which it is *successful*. As a consequence, he replaces conditions of satisfaction by "validity claims" as a distinctive layer between the speech act and the action.

The second criticism concerns the unsatisfactory distinction between directives and declaratives. For a directive to be successful, the obedience of the hearer is required. For such a speech act to be valid, therefore, the speaker must be sure that the hearer will have to obey: he must have power enough of any kind to be reasonably sure that the directive will be executed. Declaratives, on the other hand, are not grounded in personal (physical, psychological) power, but in social norms. The speaker must be in a role and a situation that allows him to make a valid declaration (see the example above: any other member of the meeting could utter "I close the meeting", but without consequences; the same is true for the chairperson who makes this utterance at the "wrong" occasion, e.g. at the dinner table). For Habermas, the speaker does not refer to anything in the objective world when making a declarative, but to the social world. Likewise, an expressive does not refer to either the objective nor the social world, but to the subjective world.

The three concepts - intersubjectivity, validity claims and reference to three worlds - are the cornerstones to Habermas' theory of communicative action. It is an attempt to go from speech act theory which offers a first link between communication and action to an action theory based on communication which has "constitutive power", e.g. offers an explanation for the establishment of society out of the interaction between individuals.

"Communicative action" is not a characteristic of any action, but one among several types of actions that have to be distinguished (TKH I, 446)



Social action itself has to be distinguished from instrumental action, i.e. an action that affects objects, as any action that concerns other subjects. Intersubjectivity is then the precondition for any type of action shown in the diagram. The different types of action can best be explained by first defining the central concept of communicative action.

Communicative action takes place between at least two subjects endowed with agency who enter an interpersonal relationship. The actors seek understanding about the situation in order to coordinate their plans and consequently their actions in a consensual way. Language is the medium for communicative action, but social action must not be reduced to speech nor interaction to conversation (or, to transfer this statement to our subject: life is not a chat-room, not even in the Internet). Verbal communication is only the mechanism by which actions are coordinated and by which the plans and purposes of the participants are brought together in the interaction. (TKH I, 141-143).

With each utterance, an actor who is interested in achieving mutual understanding makes three "validity claims" corresponding to the three "worlds of reference", i.e. the objective world, the social world, and the subjective world:

- the claim that the statement is true regarding its propositional content (objective world)
- the claim that the speech act is correct with regard to the normative context (social world)
- the claim that the speaker's intention is sincere (subjective world)

The speaker claims truth for statements, correctness for actions that are regulated by social norms and for their normative context, and sincerity for the expression of subjective experiences. (TKH I, 149).

Communicative rationality

Habermas' goal in postulating the three validity claims is to establish a concept of rationality that matches communicative action. His problem is that, traditionally, the prototype of rational action is teleological action: the actor realizes a purpose and/or effects a desirable state by choosing, in a given situation, those means that appear adequate, and by applying them. The central concept in this view of rationality is the decision between alternative actions which is guided by a purpose and by principles and based on the interpretation of the situation. (TKH I, 126-127). Teleological action becomes strategic action as soon as the actions of another actor have to be taken into account. On the whole, this is the model of action that dominates in economy and sociology, where the purpose of the actor is the maximization of his benefits or expectations of benefit.

This model of rationality is insufficient for communicative action, as it leaves out the intersubjectivity of coordinated action. Even in strategic action, where other actors come into play, they are regarded as external factors in the lonely actor's calculus. Rationality in communicative action

is based on the idea that the utterance can, in principle, be challenged by the other subjects on the ground of each of the three validity claims. In fact, the possibility of being challenged is taken into account by the actor who qualifies his utterance accordingly. The intersubjective acceptance of the other's validity claims is a precondition for understanding:

> Verständigung funktioniert als handlungskoordinierender Mechanismus nur in der Weise, daß sich die Interaktionsteilnehmer über die beanspruchte Gültigkeit ihrer Äußerungen einigen, d. h. Geltungsansprüche, die sie reziprok erheben, intersubjektiv anerkennen. (TKH I, 148)

One remark is necessary to avoid the frequent misunderstanding that Habermas' theory is useless because, in practice, communicative action is not as ideal as he defines it, i.e. the presupposition that actors really seek understanding without influencing or dominating (= strategic action) is wrong. Habermas underlines himself that communicative action is indeed a "counterfactual assumption" which is, however, necessary in order to analyse all other types of social actions as "systematically distorted communication". (TKH I, 445).

Overt strategic action takes place when an actor considers other persons (or agents) as factors in his/her planning of teleogocial action, but not as subjects. This would be the case of a manager disposing of his "human resources" as assets that have to be used in an optimal way to ensure the company's success. In covert strategic action, on the other hand, one of the participants in the interaction is deceived. In the case of a manipulating utterance, it is clearly the hearer who mistakenly believes that she is witnessing communicative action while being "used" as an object in the speaker's action. In the case of unintentional deception, the speaker herself believes that she is acting communicatively while subjecting the other to her unconscious purposes.

The problem with these qualifications is that they are not transparent to at least one of the parties involved. That is, seen from the standpoint of the "believer", communicative action does indeed take place. Communicative action is a theoretical construct that can only be

assumed but never observed: there is always the possibility of someone betraying - himself or others. However, a tendency to assume communicative action in the absence of evidence to the contrary seems to be dominant in everyday life. The "default" expectation is not one of being constantly deceived or manipulated, maybe because it costs too much effort to imagine and forestall the thousands of possible ways in which an utterance could be used if it were not straightforward.

Communicative inter-action

Does the interaction with agents respect the requirements of communicative rationality? "chat-bots" have to be left out of this discussion: their goal is to make conversation and nothing else. Communicative action must not be reduced to conversation - the question of rationality only arises when the agent is endowed with action goals. They can be those of the user that the agent takes over and pursues in the user's place, or it can be the goal to assist the user. In the third case which is most similar to human-human interaction, the agent pursues a goal that is different from the human's at the departure.

In order to discuss communicative rationality of agents, we first have to show how the concept of "three worlds" could apply to a software agent.

Objective world: as in communication between human actors, the "world" is the section of the totality of "what is the case" that is currently at issue (the situation). The world is, in this case, the application (virtual world, game or learning environment etc.) inside which the agent is implemented and of which it has "knowledge" in the form of representations. In some papers on the subject, this world is called the "lifeworld" of the agent. However, this concept ("Lebenswelt") has a distinct and complex meaning in sociology, where it designates the background of the situation and the part of the world that cannot be brought into rational discourse, the "horizon" of communicative action. (TKH I, 123, TKH II, 171ff.)

Subjective world: this could be interpreted in such a way that the agent's utterances indeed reflects its current state. This could be the state of the current plan to reach a goal or subgoal, it could be the current state of beliefs, desires and intentions, and/or the current value of the agent's emotional parameters. It would lead us into a long philosophical debate if we discuss whether the expression of internal states of the agents by using the same words as humans use to express their emotional state is legitimate. In principle, the subjective world of a human being is her own "world" to which she has privileged access. Expressions of inner states can be challenged, but are not open to examination and argumentation. In this sense, this validity claim rests on weaker grounds than the others in Habermas' theory (Baumgartner 1993). An agent software that displays, on request, its current parameter settings in its emotional model would be more transparent and "rational" than a human actor.

Social world: Although there is as yet no hybrid society with its norms and institutions that both user and agent share intersubjectively (or interactively), we can safely speak of norms that are at work in the interaction. The most obvious case are the social roles that are played by the user and the agent. In the case of a pedagogic agent, for example, the user is in the role of the student, while the agent acts as a teacher, tutor or learning companion (coach) (Baumgartner and Payr 2000). Both accept their roles and act accordingly. The same is true for any "assistant agent" that is expected by the user to be helpful, sincere, and obedient simply on the ground of its role. A detailed analysis of language processing and generating mechanisms in conversational agents would reveal a whole range of other, more fine-grained references to the social world.

The key principle of communicative rationality is that the validity claims can, in principle, be challenged and criticized by the hearer and justified by the speaker. This is certainly the case or at least feasible for references of the agent to the objective and the subjective world that both are represented explicitly in the program. The agent could, for example (even if only in principle) expose its emotional model and make visible the process by which it came to be in the state it currently is in.

References to the social world, however, are mostly implicit in the design of the agent and its application. The agent, contrary to the human, *is* in fact the role character that it enacts. It hardly makes sense to challenge a pedagogic agent's hint or instruction on the ground that you don't accept it as a tutor: it cannot be anything else and therefore cannot take the distance from its role that would be necessary to justify its reference to the social world.

Every communicative action presupposes an intersubjectively shared definition of the situation in which it takes place. When A gives an order to B, and B obeys it, they both share the view that it is the right place, time, and combination of persons to carry out the action, and the situation will probably remain implicit. Whenever they have diverging definitions of the situation in any of these aspects, they would have to re-negotiate their definitions in order to make them coincide sufficiently for the communicative action to be successful. The situation is the section of the lifeworld that is brought into focus by the issue at hand (TKH II, 187).

A distinction has to be made again between the goals of DAI research and research in "assistant agents". Where the modelling of human society is the issue, agents indeed seem to be too rational:

> Das Konzept der intentionalen Agenten geht gemessen am Vergleichsmaßstab menschlicher Akteure von idealisierten Anfangsannahmen aus, ... Zu Recht kann man daher das Konzept der intentionalen Agenten als "over-cooperative", "hyper-cognitive" und "over-rational" bezeichnen.

Designers of assistant agents seem indeed to start out from the counterfactual assumption of communicative action. Altruistic agents (as distinguished from egoistic agents who, in our terms, act teleologically, e.g. in multi-agent-systems modeled on game theory) are based on the assumption that the eagerness to cooperate and to help are mental properties (SPP-Sozionik n.d.). Egoistic agents, on the other hand, also rest on an unrealistic assumption, namely that humans act strictly teleologically and on the basis of instrumental rationality. Both are

extremes - a good model would probably lie somewhere in between. Modelling human behavior faithfully is certainly a necessity for the simulation of human social processes, but is it necessary or desirable when building assistant agents?

Is non-verbal communication rational?

Habermas' theory has a "linguistic bias": Language is the first if not only medium of communication. The theory has to adopt the "principle of expressibility" that "Everything that can be meant can be said" (Searle 1969, TKH I, 443) as a basic assumption, else the whole concept of communicative rationality be useless. However, nonverbal cues (body language, behavior, etc.) typically form somewhere between 65 and 95 percent of human communication (Picard 1999). There are two questions to be raised at this point: do non-verbal cues communicate anything with which a validity claim is made? If not, they do not enter into communicative action, and the question of their rationality does not arise. If yes, the second question arises: can all nonverbal cues be expressed verbally? If yes, the validity claims made nonverbally can be challenged and justified, and nonverbal communication is indeed rational in the sense of the theory of communicative action.

Obviously, the answers cannot be given so easily nor so globally for all of nonverbal communication. A closer look at the different "media" or channels, forms and functions of nonverbal communication will be necessary: voice, gaze, gesture, facial expression, motion, posture, space and time. In this first approach, gestures will be taken as an example.

Traditional classifications of gestures, e.g. in the "Dictionary of Worldwide Gestures" (Bäuml and Bäuml 1997) are purely descriptive and based on form and location of gestures. The fact that gestures are isolated from the verbal expressions that they accompany or complement a) excludes co-verbal gestures and b) gives us no possibility to study them in the context of communication. The following categorization of gestures

attempts a functional approach and combines it with a distinction between conscious and unconscious production (Cassell 2000):

- *Emblematic gestures* are gestures that are used or can be used instead of a verbal expression, replacing a whole utterance, e.g. "thumbs up". These gestures differ from culture to culture, also in the number of different gestures and the frequency of use. They are easy to observe and to interpret, and on the whole well known, studied and described, e.g. in the various "dictionaries of gestures". However, in few cultures they appear to constitute more than ten percent of the gestures produced by speakers.

- In *propositional gestures*, the gesture replaces a grammatical constituent for which a demonstrative pronoun appears in the verbal utterance, as in "the box was this big" (where the size is shown with the hands) or "bring me that" (while pointing at the object that should be brought). They are found mainly in communicative situations where the physical world is the topic.

Both emblematic and propositional gestures are classified as conscious gestures by Cassell, and contrasted with spontaneous gestures of which there are four kinds: iconic, metaphoric, deictic and beat gestures.

- *Iconic gestures* depict by their form some (physical) feature of the action or event being described, e.g. showing a structure or pantomiming a process.

- *Metaphoric gestures*, by contrast, represent a concept that has no physical form. The gesture takes form from a common metaphor, e.g. the "conduit metaphor" for the communication process, where information is represented as a concrete object that is passed on to the hearer.

- *Deictic gestures* "spatialize, or locate in the physical space in front of the narrator, aspects of the discourse". One could, for example, outline a problem, issue etc. as a region in space and refer to it by pointing to this region.

- *Beat gestures*, finally, are "small batonlike movements that do not change in form with the content of the accompanying speech." They can be observed in particular in meta-communication (e.g. comments on one's own utterances, speech repairs, and reported speech).

The second group of gestures contains those types that occur co-verbally, i.e. accompanying speech, but not replacing it. In the first group, the criterion which separates emblematic from propositional gestures is grammatical - whether the gesture replaces a sentence or a constituent. In the second group the distinction is based on the contents of the gesture: a physical object - present in the case of deictics, absent in the case of iconics, a metaphor, or a metalinguistic contents. Whether this description does justice to the wide variety of gestures is a question that cannot be pursued here. In any case, it makes clear that all the gestures taken into account here are "propositional" , not in the sense of the word used in the chapter cited here, but in the sense used by the same author in the following chapter (Cassell, Bickmore et al. 2000), where propositional and interactional functions of conversation are distinguished. The propositional parts (whether verbal or nonverbal) are those that "represent the actual thought being conveyed, or propositional content", while interactional parts (equally verbal or nonverbal) "serve the sole purpose of regulating the interaction".

From the perspective of communicative rationality, propositional gestures (in the broader sense of gestures pertaining to propositional functions) refer to or contribute to referring to the objective world. We can imagine - even if it is rarely the case - that a hearer challenges the claim made with such a gesture.

Metaphorical gestures are certainly difficult to challenge, but so are metaphors expressed verbally. There are two "layers" on which they would have to be criticized: one is the metaphor itself, the other is the assertion that is made using the metaphor. An example from political discourse may illustrate this point: some years ago, right-wing political parties in European countries used the slogan "The boat is full" to denounce governmental immigration politics that they considered too

liberal. Whoever negated this metaphor by affirming that the boat was not full had already accepted the metaphor with all its implications and connotations: a (life?)boat already filled to its brim with people, alone out there on the ocean, in constant danger of sinking. It was essential to reject the metaphor as such, i.e. that comparing wealthy European countries with their high standard of living to an endangered boat was utterly wrong and misleading.

This example makes clear that metaphors (and metaphorical gestures) belong to the domain of covert strategic action. The "boat metaphor" is used with deceptive, manipulative intention. The "conduit metaphor" may be used by the speaker in good faith as the natural way to express an abstract concept, but can nevertheless distort communication as it becomes extremely difficult for the hearer to step outside the "picture" that the metaphor draws.

So, leaving metaphor(ical gestures) aside as a problematic domain, and beat gestures as probably "interactional", we are left with propositional gestures that make validity claims, especially to the objective world.

If gestures were used only in propositional or interactional functions, the discussion could be closed at this point. Behavior that is focused on regulating interaction does not enter into communicative action, but is a pre-condition to it. It is certainly crucial in setting up the situation in which communicative action takes place, but only a small part of it can be "made an issue" and "re-negotiated" explicitly. Interactional nonverbal behavior raises other questions and has to be analyzed on other grounds (cf. Merten 1996, Suwelack 1998).

However, to distinguish only between propositional and interactional functions of communication leaves out important aspects, such as illocutionary force. It is much more often expressed nonverbally than verbally, e.g. by typical verbs indicating the illocutionary role of a speech act. Nonverbal cues play a decisive role e.g. in marking a speech act as indirect and in giving hints for its interpretation:

Assertions like, for example

- "Your brother arrives today." are transformed into questions by intonation,
- "The report will be ready on Monday" are transformed into a promise by a steady, direct gaze at the hearer,
- "I don't want to see any more of this mess" are transformed into a command by intonation, facial expression, and maybe a warning gesture (hinting at the possible sanctions that back up the command).

Here, too, validity claims are made. References to the subjective world and, in part, to the social world are the predominant function of illocutionary nonverbal behavior, while speech and propositional gestures are the main instrument for claims regarding the objective world. In every case, the hearer has the possibility to challenge the claims made and to bring them into the verbal discourse. The promise made with a gaze, the menace meant by a gesture etc. can not only be rejected or ignored by action (or non-action), but can be directly addressed. This part of nonverbal communication, then, also falls inside communicative rationaltiy.

Expression of emotions is a spcial case, as has been mentioned before: as the speaker is the only one with access to her subjective world, these validity claims cannot be really criticized. A challenge like "You are not sincere" can neither be justified by the challenger nor effectively refuted by the challenged: in the end, it remains a (dis)belief.

We will conclude this discussion by roughly re-drawing the boundary of communicative action in a new way: while we will include propositional and illocutionary nonverbal communication, we will exclude expressive and interactional behaviors. They make up the domain where we should look for potential breakdowns in nonverbal communication, as will be done in subsequent work.

References

André, Elisabeth, Martin Klesen, Patrick Gebhard, Steve Allen, and Thomas Rist. 2000. Exploiting Models of Personality and Emotions to Control the Behavior of Animated Interface Agents. In *Proceedings of the workshop on "Achieving Human-Like Behavior in Interactive Animated Agents", Barcelona, June 2000*, edited by J. Rickel. Barcelona. 3-7.

André, Elisabeth, Thomas Rist, and J. Müller. 1999. Employing AI methods to control the behavior of animated interface agents. *Applied Artificial Intelligence* 13:415-448.

Austin, J. L. 1965. *How To Do Things With Words. The William James Lectures at Harvard University in 1955*. New York/Oxford: Oxford University Press.

Badler, Norman I., Barry D. Reich, and Bonnie L. Webber. 1997. Towards Personalities for Animated Agents with Reactive and Planning Behaviors. In *Creating Personalities for Synthetic Actors*, edited by R. Trappl and P. Petta. Berlin/Heidelberg: Springer. 43-57.

Bates, J. 1994. The role of emotion in believable agents. *Communications of the ACM* 3 (7):122-125.

Baumgartner, Peter. 1993. *Der Hintergrund des Wissens*. Klagenfurt: Kärntner Druck- und Verlagsgesellschaft.

Baumgartner, Peter, and Sabine Payr. 1999. *Lernen mit Software*. 2nd ed. Innsbruck: Studienverlag.

Baumgartner, Peter, and Sabine Payr. 2000. Erfinden lernen. In *Konstruktivismus und Kognitionswissenschaft*, edited by A. Müller, K. H. Müller and F. Stadler. Wien, New York: Springer. 89 - 106.

Bäuml, Betty J., and Franz H. Bäuml. 1997. *Dictionary of Worldwide Gestures*. 2 ed. Lanham (MD): Scarecrow Press.

Blache, Fabian. 2000. *History of Tomb Raider*. [cited 15 May 2000]. Available from http://www.gamespot.co.uk/pc.gamespot/features/ tombraider_hist/.

Bret, Michel. 2000. Virtual Living Beings. In *Virtual Worlds. 2nd International Conference, Paris, July 2000. Proceedings*, edited by J.-C. Heudin. Berlin/Heidelberg: Springer. 119-134.

Brusilovsky, Peter. 1999. Adaptive and Intelligent Technologies for Web-based Education. *Künstliche Intelligenz. Special Issue on Intelligent Systems and Teleteaching. http://www.contrib.andrew .cmu.edu/~plb/papers/Ki-review.html* (4):19-25.

Cassell, Justine. 2000. Nudge Nudge Wink Wink: Elements of Face-to-Face Conversation for Embodied Conversational Agents. In *Embodied Convrsational Agents*, edited by J. Cassell, J. Sullivan, S. Prevost and E. Churchill. Cambridge, MA: MIT Press. 1-27.

Cassell, Justine, Tim Bickmore, Lee Campbell, and Hannes Vilhjalmsson. 2000. Human Conversation as a System Framework: Designing embodied Conversational Agents. In *Embodied Conversational Agents*, edited by J. Cassell, J. Sullivan, S. Prevost and E. Churchill. Cambridge, MA: MIT Press. 29-63.

Cassell, Justine, Joseph Sullivan, Scott Prevost, and Elizabeth Churchill, eds. 2000. *Embodied Conversational Agents*. Cambridge, MA: MIT Press.

Cook, Linda, Tim Bickmore, Sara Bly, Elizabeth Churchill, Scott Prevost, and Joseph W. Sullivan. 1999. Autonomous Synthetic Computer Characters as Personal Representatives. In *Human Cognition and Social Agent Technology*, edited by K. Dautenhahn. Amsterdam/Philadelphia: John Benjamins. 85-111.

Damer, Bruce, Stuart Gold, Jan de Bruin, and Dirk-Jan de Bruin. 2000. Conferences and Trae Shows in Inhabited Virtual Worlds: A Case Study of Avatars98 & 99. In *Virtual Worlds. 2nd International Conference, Paris, July 2000. Proceedings*, edited by J.-C. Heudin. Berlin/Heidelberg: Springer. 1-11.

Dautenhahn, Kerstin. 1998. The Art of Designing Socially Intelligent Agents: Science, Fiction and the Human in the Loop. *Applied Artificial Intelligence* 12 (7-8):573-618.

del Galdo, Elisa M., and Jakob Nielsen, eds. 1996. *International User Interfaces*. New York: John Wiley & Sons.

Dillenbourg, Pierre, M. Hilario, P. Mendelsohn, D. Schneider, and B. Borcic. 1993. Intelligent Learning Environments. Report from the project "Les systèmes explorateurs intelligents". Geneva: TECFA. http://tecfa.unige.ch/tecfa/research/memolab/report93.

Donath, Judith S. 2000. Being Real. In *The Robot in the Garden. Telerobotics and Telepistemology in the Age of the Internet,* edited by K. Goldberg. Cambridge, MA: MIT Press.

Elliott, Clark, and Jacek Brzezinski. 1998. Autonomous Agents as Synthetic Characters. *AI Magazine* (Summer 1998):13-30.

Foner, Leonard N. 1999. Are We Having Fun Yet? Using Social Agents in Social Domains. In *Human Cognition and Social Agent Technology*, edited by K. Dautenhahn. Amsterdam/Philadelphia: John Benjamins. 323-348.

Furukawa, Takao, Jin Gu, WonSook Lee, and Nadia Magnenat-Thalmann. 2000. 3D Clothes Modeling from Photo Cloned Human Body. In *Virtual Worlds. 2nd International Conference, Paris, July 2000. Proceedings*, edited by J.-C. Heudin. Berlin/Heidelberg: Springer. 159-170.

Goffman, Erving. 1994. *Interaktion und Geschlecht. edited and introduced by Hubert A. Knoblauch*. Frankfurt a.M./New York: Campus.

Gottdiener, M. 1995. *Postmodern Semiotics*. Oxford: Balckwell.

Gumperz, John J. 1982. *Discourse Strategies*. Cambridge: Cambridge University Press.

Habermas, Jürgen. 1999. *Theorie des kommunikativen Handelns*. 3rd ed. 2 vols. Frankfurt a.M.: Suhrkamp. Original edition, 1981.

Hall, Edward T. 1981a. *Beyond Culture*. New York: Anchor Books.

———. 1981b. *The Silent Language*. New York: Anchor Books.

———. 1982. *The Hidden Dimension*. New York: Anchor Books.

Hayes-Roth, Barbara. 1998. Character-based interactive story systems. *IEEE Intelligent Systems*, 12-15.

Hayes-Roth, Barbara, Robert van Gent, and Daniel Huber. 1997. Acting in Character. In *Creating Personalities for Synthetic Actors*, edited by R. Trappl and P. Petta. Berlin: Springer. 92-112.

Hayles, N. Katherine. 1999. *How We Became Posthuman. Virtual Bodies in cybernetics, Literature, and Informatics*. Chicago/London: University of Chicago Press.

Heaton, Lorna. 1998. Talking Heads vs. Virtual Workspaces: A Comparison of Design Across Cultures. *The Edge: The E-Journal of Intercultural Relations.* *http://kumo.swcp.com/biz/ theedge/heaton.htm* 1 (4).

Heim, Michael. 1987. *Ectric Language: A Philosophical Study of Word Processing.* New Haven/London: Yale University Press.

———. 1998. *Virtual Realism*. Oxford/New York: Oxford University Press.

Hirsh, Haym. 1998. Interactive fiction. *IEEE Intelligent Systems*, 12-21.

Johnson, W. Lewis, Jeff W. Rickel, and James Lester, C. 2000. Animated Pedagogical Agents: Face-to-Face Interaction in Interactive Learning Environments. *International Journal of Artivicial Intelligence in Education*.

Lester, James C., Stuart G. Towns, Charles B. Callaway, Jennifer : Voerman, and Patrick J. FitzGerald. 2000. Deictic and Emotive Communication in Animated Pedagogical Agents. In *Embodied Conversational Agents*, edited by J. e. a. Cassell. Cambridge (MA): MIT Press. 123-154.

Magnenat-Thalmann, Nadia, and Pascal Volino. 1997. Dressing Virtual Humans. In *Creating Personalities for Synthetic Actors*, edited by R. Trappl and P. Petta. Berlin/Heidelberg: Springer. 9-24.

Merten, Jörg. 1996. *Affekte und die Regulation nonverbalen, interaktiven Verhaltens. Strukturelle Aspekte des mimisch-affektiven Verhaltens und die Integration von Affekten in Regulationsmodelle.* Frankfurt/M.: Peter Lang.

Picard, Rosalind W. 1999. Response to Sloman's Review of "Affective Computing". *AI Magazine* (Spring 1999):134-137.

Rammert, Werner. 1998. Giddens und die Gesellschaft der Heinzelmännchen. Zur Soziologie technischer Agenten und der Multi-Agenten-Systeme. http://userpage.fu-berlin.de/~rammert/articles/Multi.html.

Reeves, Byron, and Clifford Nass. 1996. *The Media Equation*. Cambridge: Cambridge University Press.

Rizzo, Paolo, Manuela M. Veloso, Maria Miceli, and Amedeo Cesta. 1999. Goal-based personalities and social behaviors in believable agents. *Applied Artificial Intelligence* 13 (3):239-271.

Schefe, Peter. 1997. Prolegomena zu einer Agentologie: Magie, Metapher oder Mache? In *HyperKult. Geschichte, Theorie und Kontext digitaler Medien*, edited by M. Warnke, W. Coy and G. C. Tholen. Basel/Frankfurt a.M.: Stroemfeld. 411-432.

Searle, John R. 1969. *Speech Acts. An Essay in the Philosophy of Language*. Cambridge: Cambridge.

———. 1979. *Expression and Meaning. Studies in the Theory of Speech Acts*. Cambridge: Cambridge.

———. 1983. *Intentionality. An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.

SPP-Sozionik. n.d. Sozionik: Erforschung und Modellierung künstlicher Sozialität. http://www.tu-harburg.de/tbg/SPP/spp-antrag.html.

Suwelack, Dorit M. 1998. *Die stumme Beziehungssprache der Geschlechter. Eine Mikroanalyse des nonverbalen Interaktionsverhaltens gegen- und gleichgeschlechtlicher Dyaden*. Frankfurt/M.: Peter Lang.

Turkle, Sherry. 1998. *Leben im Netz. Identität in Zeiten des Internet*. Reinbek b. Hamburg: Rowohlt.

Ventrella, Jeffrey. 2000. Avatar Physics and Genetics. In *Virtual Worlds. 2nd International Conference, Paris, July 2000. Proceedings*, edited by J.-C. Heudin. Berlin/Heidelberg: Springer. 107-118.

Wasson, Barbara. 1998. Identifying Coordination Agents for Collaborative Telelearning. *International Journal of Artificial Intelligence in Education* 9 (3-4).