Abstract

In this paper we explore the question of whether collaborative virtual environments can be regarded as social systems. One criteria of social systems is the presence of social conventions which serve as a basis for common communication. We conducted empirical research in three different on-line environments in order to identify explicit and implicit social conventions. We found evidence that a number of social conventions exist and that the nature of their expression was different depending on the media and functionality available in the environment.

Keywords: collaborative virtual environments, social conventions, empirical studies, avatars

1. Theoretical Background: Social conventions

In several articles on new media [8, 15, 17, 20], the question has been discussed whether collaborative virtual environments (CVEs) can be regarded as social systems comparable to those in “real-life”. Our paper contributes to this discussion by focusing on a specific aspect. Referring to sociological and social philosophical research on social systems like groups and communities [10, 16, 21], we assume that social conventions and rules are a fundamental precondition for the stability, efficiency, and inner coherence of a social system.

Especially in social philosophy, social conventions have been described as normative rules of conduct which are based on implicit ethic imperatives [12, 24]. According to this, social conventions are accepted by group or community members even if they have the opportunity to behave in a different way. Social conventions not only determine how to behave within a group, but furthermore, they define some behaviour as incorrect. Following this, they guarantee the stability and consistency of a social
system. Normally, a distinction between implicit and explicit social conventions has been made in social philosophical discussions [14, 23]. Some social conventions are articulated by explicit agreements, or even laws, which have been established by institutions or responsible persons. However, more often, social conventions are implicit. They determine the behaviour of members of a social system without being codified or formulated. Therefore, to describe how a specific social system is functioning and how the members of such a system are acting, it seems to be a good starting point to explore and identify such underlying social conventions. We assume that an investigation about the use of such implicit social conventions would give insight into the social practice of a social system, i.e. demonstrating the way how people behave and act [27]. Furthermore, social rules are the underlying preconditions of communication [25, 28], because the way how people communicate with each other is embedded in social practice and specific life styles, which are determined by implicit social conventions. According to this, social rules function not only for comprehension but also for coherence within such a system by establishing a common context [24] and a common normative fundament.

Social philosophical research has pointed out that new members of a specific social system have to become aware of these implicit social conventions [24, 25]. By learning and accepting them, they will be integrated in such a group or community. Furthermore, one will only be able to communicate with and understand a partner if one has gained some experience about these implicit conventions. Thus, if we regard collaborative virtual environments as specific forms of social systems, it seems to be a successful research strategy to explore the implicit and explicit social conventions as a first step toward gaining an insight into the particular social practice within such environments.

Other empirical studies have addressed social behaviors in virtual environments, such as the nature of turn-taking and avatar movement [4], dynamics of virtual meetings [5], movement in the virtual world [11], experiences from a mixed-reality environment [2], identity construction [1, 7], cultural formations [20], communication in on-line communities [15], and observations in text-based MUDs [6].

2. Methodological Approach and Research Setting
Ethnomethodology [9] is a methodological approach whereby the social conventions which guide the behaviour and attitudes of members of a social system can be grasped by observation. In ethnomethodology, social systems are regarded as a net of meaningful behaviour, which cannot only be traced back to formal rules and explicit conventions, but which are guided more often by implicit conventions which are to some extent open, contingent, and flexible. Following this, the empirical facts which have been found in observation cannot be explained by identifying global structures or formulating general laws, but are more seizable by the description of single phenomena. Therefore, we concentrated our research on social conventions in CVEs by a detailed description of conventions in communication to get some insight into the social practice of these environments.

Three different on-line environments were chosen in which to study the existence of social conventions: Active Worlds\(^1\) (AW), Onlive Traveler\(^2\) (OT), and LambdaMOO\(^3\) (LM). All environments are accessible from the Internet. These environments were chosen primarily since they have been in existence for some time and offered different functionality for communication and representation, and thus, we expect, for the expression of social conventions. The main differences are that LM is purely text based for both representation and communication, OT has graphical 3D representations and offers text and audio for communication, and AW has graphical 3D representations and offers only text for communication. The basic functionality available for the representations and communication is:

AW: full-bodied avatars can walk and exhibit movements of waving, jumping, and dancing, activated by mouseclicks. Avatars can move in six dimensions by using the arrow keys. Communication between people is text-based by typing on the keyboard. All public messages appear in a scrollable window and also above the avatar head with the avatar name for 30 seconds or until the next typed message appears.

---

\(^1\) Copyright © 1997 COF inc.
\(^3\) Designed by Pavel Curtis at Rank Xerox Research Centre, 1991.
OT: the avatars are heads, and have four different emotions that one can activate by a mouseclick: happy, sad, surprise, and anger. The avatars exhibit what appear to be random blinks. Movement (three dimensions plus rotation in four directions—left, right, forward, and backward) is made by using arrow keys. Communication is audio (outgoing audio is activated by pressing down the control key and speaking into a microphone) or text-based (pulling down a menu, selecting an avatar, and typing into a message which appears on the screen). The text is limited to two lines.

LM: all representation of users and communication is text-based. Different commands are used for communication (e.g. say, whisper, emote), manipulation (e.g. get/take, move), information (e.g. look, who, etc.), and creation (e.g. dig, create), as well as others.

Three different researchers spent time observing three different on-line environments. Approximately 59 hours were spent in total observation time: 21 hours in AW, 20 hours in OT, and 18 hours in LM. Each observer was primarily responsible for making observations in one particular environment, but all observers also spent time in each of the other environments to become familiar with them. Although the on-line characters adopted by the researchers were varied somewhat, most of the time the same on-line characters were used during the time spent in the environments. The observation was performed during May-June, and October, 1997 for LM, and September-October, 1997 for OT and AW. The observers took notes and recorded behavioral observations under assigned categories of social behaviors, described in the next section. The observers met periodically and compared observations to make sure that the categories were being coded consistently. On-line recording and logging was not performed due to privacy concerns.

3. The Conventions Investigated

Based on the goals of our study, we now describe the conventions that we set out to observe. We first present our motivations for investigating the particular conventions. We then describe the possible actions that one can take in order to perform the different social behaviors by listing the available functionality. Since LM offers only text-based communication,
the functionality for LM will thus only be described if a particular feature of the text functionality is relevant for the social behavior.

3.1 Contacting others: greeting and acknowledging

Greeting is an important convention that serves to initiate contact and can influence subsequent interaction. Acknowledgment functions to express the degree of willingness to communicate. The following behaviors were coded: *greetings* (What types of expressions or rituals are used to greet?), *acknowledgment* (How is the presence of others acknowledged, e.g. role of eye contact/body position?).

*Possible actions: (AW): enter into another avatar's visual field, click to wave or other movement, send text message (must be one of 12 closest avatars to see the message) avatar can turn to face another, mouseclick for movement, mouseclick to look at avatar at eye-level. (OT): enter into an avatar's visual field, can speak using audio, show emotion, e.g. smiling, send text message regardless of whether being in their visual field, avatar can turn to face another. (LM) can send public (say command), private (whisper command) or behavioral (emote command plus an action) messages.*

3.2 Leaving

Rituals used for leaving can indicate the degree of engagement in a conversation, and social binding. The following behaviors were coded: *leaving* (What rituals/expressions can be observed in leaving?).

*Possible actions: (AW): can send text message, can use arrow keys to move away (or leave avatar's field of view), disappear (logoff, teleport to another location), fly away by depressing plus key. (OT): can use audio, can send text message, can use arrow keys to move away, disappear (logoff, teleporting).*

3.3 Group interaction: establishing group membership and indicating privacy

Through social conventions, group membership can be defined; conventions serve also as a means for the group to communicate
information about itself, e.g. whether the group is open for new members [19]. The following behaviors were coded: group membership (What conventions are used to welcome/exclude others from a group?, showing/disturbing the intimacy of others in a group (What signals do actors give others to indicate intimacy? What happens when actors are disturbed?).

**Possible actions:** (AW): can create formations, can add avatar to contact list, can send private message, can perform simultaneous actions, e.g. dancing, avatar can move away when approached. (OT): can create formations, can send text, avatars can ignore another, avatars can position themselves in close proximity, or can move away when approached.

### 3.4 Communication: verbal and nonverbal expression and conversation flow

A social system requires a common language and expression for members to communicate and understand each other. The following behaviors were coded: use of facial expressions (Are facial expressions used in the environment? When are they used?), language conventions (What language conventions can be identified? When is audio and when is text used, for OT?).

**Possible actions:** (AW): no visual facial expressions available, happiness and anger shown through body language (jumping up and down, hands on hips), can describe expression with text. (OT): can click on icon to show happy, surprise, sad, and angry expressions, can describe expression with text. (LM): emote command plus emoticons (i.e. graphical representations of emotions).

### 3.5 Social positioning and intimacy

Here we were interested to what extent the notion of space in the physical environment, and the corresponding behavioral expectations, transfer into a virtual environment [13]. The following behavior was coded: social distance (Is a social distance kept? What happens when this distance is violated?)
Possible actions: (AW): arrow keys move avatar, (OT): double clicking on another's avatar brings one face-to-face with them at a distinct distance, selecting the “go near” command brings one close to avatar anywhere in the space, avatars can move backward and forward to adjust distance.

3.6 Sanctions on behavior

In this category we were interested to explore whether implicit or explicit conventions are not only fulfilled, but whether there exists a mechanism to insure that they are followed. The following behaviors were coded: reacting to offensive behavior (How do others react to distasteful comments or behavior?).

Possible actions: (AW): can reply with text, can click on offensive avatar and use mute option (only you cannot hear them), can leave visual space or turn avatar around, can click on angry emotion button. (OT): can reply with audio or text, can use ignore command, can leave visual space or turn avatar around, can click to show angry emotion.

4. Results

In reporting the results, the observer in the environment will be designated as 0. We use the term figure to denote both the visual and text-based representations. If we are referring only to representations in AW or OT, we use the term avatar. In general, in all three environments, the expression of social conventions were influenced by the available functionality for the representations and communication. In OT, the movement and spatial positioning of the avatars played a strong role in expressing the conventions. In AW, conventions were expressed more strongly by text, similar to the expression of conventions in LM. These results will be explained in each of the categories as follows.

4.2 Contacting others: greeting and acknowledging

In all environments, informal greetings (i.e. hello, hi) are generally used to initiate conversation with another figure. The main difference lies in that greetings in AW and LM are first made as more of a public greeting, to all in the room (LM) or space, and then private greetings may be made
to individuals (using the avatar name in AW, or the whisper command in LM, which allows private communication). Greetings are usually made by the person at the time the figure joins the location. In contrast, in OT, greetings are usually directed at individuals, or to a specific group, since the avatars in OT are spatially separated and no mechanism exists to address avatars who are spatially distributed. The greeting is usually audio-based, and the initiator of the greeting repositions the avatar to face the other. Greetings are not made when one first enters the world but it is often observed that avatars initially scan the scene (by rotating or moving around). The avatars’ graphical positions give information about who they are interacting with. The avatar then navigates to a position close to another before it initiates a greeting.

It is also a convention to acknowledge another’s greeting. In OT, if the O’s avatar is not already positioned directly in front of another avatar, the other will turn to face O. In fact, sometimes considerable trouble is taken to reposition the avatar. Avatars first respond with audio, when the audio is working and quality is good. Once when a person took a long time to respond, he apologized saying he was overwhelmed with text messages. If an audio message is not received, then a text message is generally sent to confirm whether another can hear them. In AW, the response to a greeting is not from those avatars in closest proximity but from anyone within this group of 12 avatars, generally 2 or 3. In LM, acknowledgment is made also only by a few in the room. In LM, for newcomers, a convention is used, following a description in the tutorial, that one announces “Hello, I am new here”. People often offer their assistance in response.

Because in AW it is possible to greet and communicate with individuals by using the figure’s name, the avatars do not change their position very often when new contacts are made. Rather, using an avatar’s name in the greeting signals that the message is for a specified avatar. In OT, new contacts are made by moving the avatar to face another and addressing another with audio. When an avatar is spatially very far away, they are generally not approached by other avatars. This was observed with other avatars and tested by the O who positioned herself far away. The O received several text messages in greeting, but was never approached by an avatar.
Text descriptions of facial expressions and body gestures in LM are sometimes used as greetings (i.e. emote smile). These are often acknowledged by others. In contrast, in OT, even though the O’s avatar may be positioned directly in front of another, a smile has no effect in initiating conversation (nor is it reciprocated). In AW, the greeting of the O’s wave was returned a few times, but the O never received a wave from another as greeting.

4.3 Leaving

In all three environments, avatars rarely leave without saying goodbye. (In LM, there exists a written convention that it is impolite to leave without saying goodbye). Avatars sometimes disappear unexpectedly, but this could be due to technical problems, which occur. In AW and OT, the convention was observed that people almost always include an explanation for why they are leaving. This was observed less often in LM. O once said goodbye without an explanation. The other avatar responded:

Wait a minute, you haven't said why you're leaving

The average number of conversation turns was measured, and for OT and AW, the average was four turns. For LM, the average number was two turns. There is almost always a confirmation of the good-bye as though to establish a common ground, i.e. that both parties understand that a good-bye is being said, and what the reason is. In general, in AW, three sets of goodbyes often occur: goodbye to the person you are last speaking with, goodbye to others you have spoken with, or friends, and then goodbye to everyone in the space. In contrast, in OT, because avatars are repositioned to form groups, and because there is spatial audio perception, goodbye is only said to the last person or group members that one spoke with.

4.4 Group interaction: establishing group membership and indicating privacy

The main difference in the environments observed is that in OT, people make use of the visual information available to determine group membership. In OT, avatars generally welcome one into the group by repositioning themselves to form a circle thereby including the new member. Thus, one sees by scanning the environment, who is interacting with whom, and the size of the group interaction. In AW, this is less often
the case. Avatars reposition themselves to form a circle roughly about 30% of the time. Most of the time, group membership is determined by the text flow in the scrolling window, i.e. who is talking with whom. In LM, group membership is often unclear. Only by observing the text dialogues can one ascertain who is talking with whom to get an insight into the interaction structures within the environment. However, if people are conversing with the whisper command, group membership is completely unknown (which can also happen in OT and AW when private messages are sent).

In both LM and AW, it is more often the case that groups tend to break up into one-to-one conversation, rather than remaining in larger groups. In AW, this may be due to the typing; it is difficult to keep track of different conversations, since the server is slow. In OT, when groups form, they appear relatively stable.

In OT, it is rude to simply barge into the middle of an existing group. When one approaches a group, the person generally rotates their avatar around to see whether they are blocking another. Sometimes one will pull their avatar far back to see the complete positioning of the group members, a way to compensate for the lack of peripheral vision of the avatars. New visitors to OT (confirmed by asking them) are often characterized by coming into the middle of the group and not looking around. When O or others did this, it provoked annoying reactions. In all environments, when avatars are engaged in a private conversation, the reaction to any attempt to enter into the conversation is to simply ignore the outside party.

In OT, while the avatar positioning indicates that people are speaking in a group, other visual information can also signal that this conversation is private. For example, two avatars turned completely upside down signal that they are having a private conversation (with their own common perspective). This was confirmed when O (who was right side up) approached them, tried to join in, and was not acknowledged. Joint movement can also indicate privacy, e.g. moving below the floor to a semi-hidden location. Absence of lip movement in two avatars facing each other indicates they are having a text conversation, and this is often an indication that it is private, since, O was generally not acknowledged. Two avatars conversing far off in the distance from the main arena also
signals a private conversation. In AW, avatar positioning is sometimes also used to indicate a private conversation. This was observed when two avatars were positioned face-to-face very close together. In LM, because no visual information about dialogue situations exists, people can create their own private spaces without being seen by others. Thus, the possibility to disturb their intimacy is not so noticeable.

4.5 Communication: verbal and nonverbal expression and conversation flow

In AW and OT, the body gestures and facial expressions are seldom used. In AW, emotions are rather communicated with text: e.g. :o) or *blushing*. In OT, emotions are expressed either via speech (e.g. laughing or utterance) or text. One user replied, “when you laugh (i.e. with audio), that says a lot”. In LM, emotions are expressed both with the emote command and with the say command. In OT, when the avatar facial emotions were used, they appeared to be used in appropriate contexts, e.g. sadness when saying goodbye, smiling in response to a joke, or showing anger when disturbed.

There is a general convention of politeness to maintain the conversation flow in all environments. In OT, silences during conversation are generally explained after or beforehand, e.g.: Excuse me I have to take a telephone call, I'll be back in a minute. In AW, people usually did not give reasons for their absence, but did give indicators that they were leaving (BRB=be right back) or returning (I’m back).

4.6 Social positioning and intimacy

Social positioning can be expressed through boundaries in physical space. In OT and AW this is expressed by the positioning of the avatars in the space. In LM, this is expressed through text, e.g. emote: comes close. In both AW and OT, coming too close to an avatar provokes responses that suggest that people feel their social distance is being violated, which suggests that a social distance exists. Sometimes avatars in OT move very far away quickly in the distance as a response or turn to the side. For example, in OT and AW we observed such responses as:

*You're too close to me, I can't breathe*

*[message to everyone]: Moonbeam (O's avatar) is touching me*
I come from the Midwest. In the Midwest, we maintain a social distance of 28 inches.

In one case in OT, an avatar did not move away. O (female) asked avatar (male) if he minded her being so close. His answer: No I like it.

The reactions to closeness could also be due to blocking an avatar. To test this hypothesis, in OT, O moved close to others, but stood to the side, so as not to block the view. The same reactions still occurred. In AW, the text above the avatars overlaps when the avatars are too close, but text also appears in the window below. The comments suggest that it is not the text overlap that people are annoyed about since their comments have more to do with a social distance being violated.

4.7 Sanctions on behavior

Only one case of rude behavior was reported in OT. A female was subjected to harassing language by a male, and the male avatar began to ram itself into the female avatar. The female reported that a friend of hers was close by who promptly got a support person who disconnected this avatar. This experience was quite emotional for her “like a nightmare”. Other avatars reported that support people are not lenient about swearing and disconnect people for this. In LM, rude behaviour was never observed. But it has been reported that people who feel offended may ask others for help. Verbal comments are successful for dealing with offensive behavior. Normally in LM, participants apologize at once for their incorrect behaviour. In AW, abusive language was seldom observed, but when it was, people reported that they used the mute button to ignore the person. In one case, a person broadcast a public message in response to swearing: greetings all, this is your one and only warning about unacceptable language, thank you. The offensive person soon left.

5. Discussion

After observing the environments and comparing our results, we are left with the notion that the difference between virtual and physical environments seems to be less clear than we had first imagined. We found that many social behaviors in the virtual environments mirror those in the physical environment; however, the nature of the expression was different depending on what media and functionality was available.
The greeting conventions in all environments that we observed serve a function as a way to announce oneself to a group or to an individual. How a greeting is made can be an important determiner of whether another person finds the greeter attractive, and thus whether interaction will continue. This is especially true in virtual environments where social cues of a person are few [18]. Presenting oneself in an attractive way so that one will be contacted is especially important in the early selection process of interaction.

We observed two interesting aspects concerning leaving others in the environments. First, leaving others in OT, even for a short time (once a bilateral communication was established) was generally accompanied by an explanation. Second, a convention to give an extended goodbye existed more in OT and AW than in LM, as measured by number of conversational turns. This suggests to us that people were trying to establish a shared understanding [25] about their parting; they wanted to insure that the other knew they were leaving for a reason, and not simply being rude. In OT, the ease of using audio would explain an extended goodbye, but this behavior was also observed in AW. Whether the existence of a visual embodiment might lead to a closer bond among the people is an interesting hypothesis to consider.

The expression of a social distance was observed in OT and AW, despite the fact that avatars are used. This suggests an identification to some extent of the physical body with the graphical representation. It also suggests that the space in the virtual environment is understood and translated as a space similar to that in the physical world, one which contains a particular set of behavioral expectations [13]. Maintaining an appropriate social distance is an implicit convention in the physical environment; the same notion is also applied in virtual space.

The similarity in expression of social behaviors in real and virtual worlds is also illustrated in a feeling of shyness that people report feeling in the virtual environments. Most people do report having the same degree of shyness behavior as in real life. In AW and LM, where communication is text-based, shyness is reported by those when language is a concern, e.g. when one’s native language is not English. However, in OT, more people reported being shy than in the other environments:
The first time I was afraid to talk. I was afraid of saying something stupid. I don’t know why, no one knows who you are anyways.
I was real shy here at first. I'm real shy in real life. I stay in my room a lot. [OT] is a learning experience.

For some, though, the same physical world conventions did not carry over:

There are certain things that you can do in an on-line world, certain freedoms that you can’t do otherwise, for example going up and talking to strangers (OT).
I’m not shy because I don’t know what the wrong thing is (AW).

5.1 The expression of emotions by verbal and nonverbal communication

The substitution of emoticons for facial expressions used in LM and AW is due to the lack of visual representation for the embodiment. Language (textual descriptions) is the only means for communicating any form of expression; therefore language is used to describe events, almost in a comic book-like fashion [26] e.g. (LM) say “it’s just great, hmmmhm!!!” or (AW): *hmpfh*. This is combined with the textual description of body movement, gestures, and facial expression, to substitute for the visual representations of the body [1].

In OT, several reasons may exist for the lack of using emotional expressions with the avatars. First, it is an overhead to click on the icon. Second, clicking on the icon is an indirect method of showing emotion. Since audio is available, one can laugh, and express emotions through speech, which is more direct. Third, emotions may be used sparingly to make an impact when they are used. One support person who was interviewed reported that she believed that since the avatar has a lot of animation anyways, the visual emotions are used only when a unique emotion is to be expressed. Lastly, the movement of the avatars is also a way to express emotions: e.g. dislike for another’s behavior by moving away.

5.2 The role of awareness in the expression of social conventions
The awareness information about others’ social activities and behaviors played a role in influencing who contacts whom, and this was especially observed in OT, where the awareness information is visual. In OT, the visual placement of the avatars show group membership, group privacy, who wants to be alone and undisturbed (by their spatial separation), who is willing to meet others, and who is interested in you. The face-to-face positioning during interaction is a convention also found by Bowers et al., (1996) in a virtual environment where audio was used.

In AW, although visual information is present, it is used less for showing awareness information than OT. Because the avatars do not move to face each other when they speak, the text is a stronger determiner of group membership. With text, the signals are less clear for who wants to hold a private conversation or remain undisturbed. This could also explain the greater fluidity in group membership in AW, since people often enter into each others’ conversations.

In LM, the only way for others to know of one’s presence (in addition to the who command which provides a listing) is for someone to write a message (say or emote commands). In the case of entering an environment, the greeting appears to function to announce one’s presence and to gain the attention of others. It is difficult to become aware of others’ group interactions, and individual social interests. In LM, before bilateral communication is established, one feels as though they lack a partner. One is speaking to what can be compared to an empty space [1, 7], as an actor on stage facing an audience in a darkened room.

5.3 Social conventions and social sanctions

Social conventions and methods of sanctioning incorrect behaviour exist in all three environments. They may serve to form a kind of communication context which enables the understanding and inner coherence of the virtual spaces. In LM, both explicit and implicit social conventions can be found: In the beginning, a general “rule” is formulated, according to which people should behave in a correct and friendly way, otherwise they might be disconnected. People who are entering this world have to accept this convention before participating.
In all environments, people seem to be very sensitive about the fulfillment of conventions [22]. We found this sensitivity with violations of social distance and with offensive behavior. So, being confronted with rude behaviour, they react at once by making verbal comments, by showing angry emoticons, or by moving away. It is also typical that people, who feel attacked, can ask for the help of other participants and usually, they will find assistance. In AW and LM, there exist not only comparable social conventions concerning behaviour, but furthermore, language conventions have to be fulfilled, as with not using abusive language or by not using capital letters, which AW users reacted to at once as screaming behavior. When people violate these conventions, they are pointed out at once by other participants.

In general, social conventions play an important role regarding the inner coherence of these environments. People who are coming for the first time to these virtual spaces have to become aware of these conventions and follow them to be accepted by the others. In fact, people report being uncomfortable by their lack of knowledge of the conventions:

...cause I didn’t know the “in-jokes” and the current word games etc. etc.
I felt stupid a lot when I was new.

6. Conclusion

In the beginning of this paper, we raised the question of whether collaborative virtual environments can be considered as social systems. As we assumed, social conventions are a basis for the common understanding of interactive behaviors, and thus can serve to maintain the consistency of a social system. To explore this question, we looked for the presence of social conventions in several on-line environments. We did find evidence that a variety of conventions exist and many reflect those used in the physical environment. Therefore, our results are a first step in suggesting that virtual environments show characteristics that we know to exist in social systems. We hope that our results can motivate further study in this direction.

Acknowledgments
We thank Phillip Jeffrey for his valuable help in making this study possible.
7. References