Computer Supported Cooperative Work (2007) 16:265–305
DOI 10.1007/s10606-006-9021-4 

# Doing Virtually Nothing: Awareness and Accountability in Massively Multiplayer Online Worlds

ROBERT J. MOORE, NICOLAS DUCHENEAUT & ERIC NICKELL
*Palo Alto Research Center, 3333 Coyote Hill Road, Palo Alto, CA, 94304, USA (Phone: +1-650-8124613; Fax: +1-650-8124470; E-mail: bobmoore@parc.com)*

**Abstract.** To date the most popular and sophisticated types of virtual worlds can be found in the area of video gaming, especially in the genre of Massively Multiplayer Online Role Playing Games (MMORPG). Game developers have made great strides in achieving game worlds that look and feel increasingly realistic. However, despite these achievements in the visual realism of virtual game worlds, they are much less sophisticated when it comes to modeling face-to-face interaction. In face-to-face, ordinary social activities are "accountable," that is, people use a variety of kinds of observational information about what others are doing in order to make sense of others' actions and to tightly coordinate their own actions with others. Such information includes: (1) the real-time unfolding of turns-at-talk; (2) the observability of embodied activities; and (3) the direction of eye gaze for the purpose of gesturing. But despite the fact that today's games provide virtual bodies, or "avatars," for players to control, these avatars display much less information about players' current state than real bodies do. In this paper, we discuss the impact of the lack of each type of information on players' ability to tightly coordinate their activities and offer guidelines for improving coordination and, ultimately, the players' social experience.

**Key words:** collaborative virtual environments, conversation analysis, ethnomethodology, game design, Massively Multiplayer Online Games, virtual worlds

## 1. Introduction

Although "virtual worlds" – persistent 3D digital environments populated by thousands of users – have been envisioned since at least the 1980s in science fiction novels such as William Gibson's *Neuromancer* (1984) and later Neal Stephenson's *Snow Crash* (1992), it is only recently that they are beginning to be realized with ubiquity of powerful PC graphics cards and the increasing availability of broadband internet access. To date the most popular type of virtual worlds can be found in the area of video games, especially Massively

"They come here to talk turkey with suits from around the world, and they consider it just as good as a face-to-face. They more or less ignore what is being said—a lot gets lost in translation, after all. They pay attention to the facial expressions and body language of the people they are talking to. And that's how they know what's going on inside a person's head—by condensing fact from the vapor of nuance." —Neal Stephenson, *Snow Crash*, 1992

Multiplayer Online Role Playing Games (MMORPG). Such games include *Ultima Online, EverQuest, Lineage, Star Wars Galaxies, City of Heroes, World of Warcraft* just to name a few. Other similar massively multiplayer virtual worlds that emphasize socializing over "gaming" include *Second Life* and *There*. Today millions of people interact on a daily basis in these virtual worlds (Woodcock, 2005). On average, players spend 20 or more hours in these environments per week (Yee, 2005).

Player-to-player interaction is central to the experience of virtual worlds. What sets them apart from other computer games is that players can adventure in a world that is inhabited by other players. Game play is fundamentally social in nature, and game worlds are designed to encourage and facilitate social interaction among players. The social interactions that players have and the social networks they build up are what make virtual worlds "sticky" and are believed to be what enables game worlds to retain players long-term (Bartle, 2003).

Unlike the earlier text-based Multi-User Dungeons (MUDs; see Curtis, 1992; Bartle, 2003), today's massively multiplayer virtual worlds combine rich 3D environments and "avatars" (or virtual bodies) with text-chat systems (similar to Internet Relay Chat or Instant Messaging) which players use to "talk" to one another. The result is a form of virtual social interaction that has been purposefully designed to resemble face-to-face conversation in many ways (see Figure 1). Avatars can approach each other, face each other, gesture to each other, in some cases exchange facial expressions, and more.


*Figure 1.* Avatar-mediated interaction in Star Wars Galaxies (left) and EverQuest II (right).

A major trend in the game industry overall is the push toward increasing realism in virtual game worlds.[1] Game developers have made great strides in achieving game worlds that look and feel increasingly like the real world. They are doing this by modeling environments with realistic landscapes and objects, by building physics engines that simulate physical forces, by simulating lighting and shadow effects, by modeling avatars after the human anatomy, by capturing and transferring the motion of real human bodies to avatars, and more. This is not to say that the goal is to make everything realistic. Of course, much of the fun comes from environments, characters, objects, and actions that are fantastic. As in live-action fantasy and science fiction movies (e.g., the *Star Wars* or *Lord of the Rings* movies), the goal is to create an experience of a world that is both familiar and fantastic. For example, Luke Skywalker can move objects with his mind and Frodo Baggins can become invisible, yet they both walk and talk pretty much like ordinary real people.

But despite their ever-increasing visual realism, today's virtual game worlds are much less advanced in terms of their *interactional* sophistication. This lag in the sophistication of social interaction systems is no doubt due to the fact that developers have only started to grapple with questions of how human-to-human interaction works and how it should be modeled relatively recently with the emergence of MMORPG genre of games which emerged in the late 1990s. Avatars are currently rather crude in terms of the level of information they communicate to others about a player's current state.

Although past studies have cautioned that "the straightforward translation of human physical embodiments into CVEs [Collaborative Virtual Environments] (however detailed) are likely to be unsuccessful, and indeed misleading..." (Hindmarsh et al., 2001, p. 135), game developers are nonetheless aiming for social interaction systems modeled after the real world. In virtual worlds built for entertainment, simulating features of the real world can be critical for the aesthetic experience in ways that are perhaps less relevant in virtual environment build for work. We see this renewed effort to create some level of "realism" in interaction as an exciting challenge for research and design. It raises many interesting questions: how can the affordances of face-to-face interaction be simulated in virtual worlds in such a way that their coordination and enjoyment are increased? Which features of real life can be adapted? Which ones cannot? And how can the rules be bent in order to create a player "experience of realism" without necessarily following the real world exactly? Furthermore, as virtual worlds are used more and more as simulations for training real-world skills,[2] solving the technical challenges in simulating face-to-face interaction will become increasingly important.

Success in creating an avatar interaction system that looks, feels, and behaves realistically to players will require a technical understanding of real-life face-to-face. To this end, the fields of conversation analysis and ethnomethodology (Garfinkel, 1967; Sacks et al., 1974; Garcia and Jacobs, 1999), should be particularly useful in guiding virtual world design because they uncover the mechanics of real-life social interaction empirically and at the right level of granularity, that is, the level of embodied actions on a turn-by-turn basis. In this paper we examine the current state-of-the-art of social interaction systems in massively multiplayer virtual worlds and offer design guidelines for increasing their effectiveness.

## 2. Methods

This study employs the methods of video-based conversation analysis (Sacks et al., 1974) that are grounded in virtual ethnography (Lyman and Wakeford, 1999; Hine, 2000; Miller and Slater, 2000; Mason, 1999; Rutter and Smith, 2005). We began by gaining a player's perspective through participant observation in several virtual game worlds including: *Star Wars Galaxies* (SWG), *World of Warcraft* (WoW), *EverQuest II* (EQ2), *EverQuest Online Adventures* (EQOA), *Second Life* (SL), and *There*. What each of these systems has in common is that they are massively multiplayer persistent worlds in which players interact with each other primarily by using 3D avatars and text chat. We played the games on a regular basis (sometimes for as much as 12 hours a week), learned the game lingo and practices, observed active public spaces, grouped with other players, formed online friendships, joined and started player guilds, and more. As with most virtual ethnographies, we studied the players through the personae they project into those worlds, rather than as their real-life personae. Mason (1999, p. 63) describes the approach:

> A virtual ethnography is one that fully immerses the ethnographer into the consensual reality experienced by groups of people who use computer-mediated communication as their primary, and often only, means of communication. As such, the online or virtual personae of the participants are the main focus of the ethnographer. Generally, researchers have wanted to focus on the person at the keyboard; a virtual ethnography reverses this and works instead with the persona that has been projected into cyberspace by the typist.

Just as we have adapted the use of ethnographic methods to the study of virtual worlds, so too have we adapted the methods of ethnomethodology and conversation analysis. These fields are concerned with how people organize and achieve recognizable social activities, such as "taking turns in conversation" or "standing in a service line," through their concrete embodied actions, and especially through their talk (Garfinkel, 1967, 1988; Heritage, 1984). The aim of analysis is twofold: to identify the generic systematic features of social action and interaction and to explain how these features are uniquely accomplished on particular occasions (Heritage, 1984; Schegloff, 1987; Zimmerman, 1988).

Ethnomethodology and conversation analysis have been used to understand face-to-face interaction (Goodwin, 1979; Schegloff, 1998) as well as technologically mediated interaction including the telephone (Schegloff, 1968; Whalen et al., 1988), video conferencing (Ruhleder and Jordan 2001), push-to-talk telephones (Szymanski et al., forthcoming), internet relay chat

(Garcia and Jacobs, 1999; Vallis, 1999), and avatar-based online interaction (Bowers et al., 1996; Hindmarsh et al., 1998, 2001).

Both ethnomethodology and conversation analysis use detailed transcripts of recordings of naturally occurring activities in order to discover their endogenous, local organizations. The transcription conventions used were initially developed by Gail Jefferson to capture many details of talk, especially the timing of turns-at-talk relative to one another. Later others elaborated the system to include the capturing of gesture and embodied conduct (Goodwin, 1979, 1986; Streeck, 1993; Hindmarsh and Heath, 1999). For interaction in virtual worlds, transcription is somewhat different. All of the interaction systems examined in this paper use primarily text chat instead of voice. Because the phenomenon itself is textual, the chat does not need to be "transcribed" *per se*. In fact, some game clients conveniently enable the user to capture the chat log, complete with the timestamps for each posted message, to a file. However, automatically generated system logs never capture all of the features of the interaction that are relevant for analysis. For example, as Garcia and Jacobs (1999) point out for IRC, the chat logs fail to capture the temporal dimension of composing a chat message. The log merely records the final message that is posted to the server. Furthermore, current chat logs do not capture most of the avatar actions, especially movement through space, or user interface actions. To capture these details, system logs must be manually annotated with additional detail by using screen-capture video recordings.

For this study, we recorded over 200 hours of screen-capture video. We collected video from our own characters' perspective as we conducted the ethnographies. In most cases, other players were not aware that we were researchers. Conversely, in all cases, we were not aware of the real-world identities of the players, only their in-game pseudonyms and personae. Gaming sessions were recorded with a Mini-DV camera connected to the video-out port of the PC's graphics card. Tapes were later converted as MPEG-2 video. We used this method, rather than capturing the video directly to the hard disk because it enabled us to capture multiple hours of video easily without having to worry about quickly filling up hard drives. The videos were reviewed and collections of salient phenomena were made following conversation analytic practice. Phenomena of interest included the mechanics of avatar interaction (e.g., chat, gestures, facing, eye gaze, facial expression), grouping sequences, coordinated travel and combat, player-hosted events, "role playing" techniques, and more. While our methods confirm the recurrence of many of the phenomena observed, it does not allow us to make precise statements about generalizability since relative frequencies were not computed nor a systematic sample drawn (Schegloff, 1987).

## 3. Background

Research on social interaction in a variety of settings – face-to-face, Internet Relay Chat, Collaborative Virtual Environments, and mediated communication more generally – is relevant for understanding social interaction in virtual worlds. Many observations and findings from these settings apply to virtual worlds, yet the latter also represents a unique domain in its own right. Although there is a growing literature on virtual worlds (Castronova, 2001; Bartle, 2003; Taylor, 2003), few studies focus on the mechanics of the social interaction systems in these virtual worlds (Brown and Bell, 2004).

### 3.1. COLLABORATIVE VIRTUAL ENVIRONMENTS: BLOCKIES ARE GROWING UP

Early research on "Collaborative Virtual Environments," or CVEs (Benford et al., 1994, 1997; Bowers et al., 1996; Hindmarsh et al., 1998, 2001) examined a new kind of three-dimensional social space with new systems of social interaction. CVEs were designed primarily for work and involved multiple users (up to 64 according to Robinson et al., 2001), crude avatars dubbed "blockies," and voice communication. A decade later MMORPGs have exploded onto the video game scene. Unlike early CVEs, MMORPGs are designed primarily for play and of course can support many more simultaneous users (up to several thousand players per server). In addition, the avatars are much more sophisticated in appearance, often with animations produced through the motion capture of human bodies, and tend to use text chat instead of voice for conversation. Because millions of gamers and online socializers have been participating in these worlds for several years, complex cultures concerning in-world conduct have emerged.

Drawing on insights from the analysis of face-to-face interaction, the CVE literature identified several fundamental questions regarding social interaction in virtual worlds. They include: how is turn-taking in voice conversation organized? How are references to virtual objects in the environment accomplished (e.g., through avatar pointing)? How are avatar orientation and gesturing used communicatively? How can users determine where another user is looking in the environment? How are technical failures and interruptions from the real world interpreted and managed? In each area the CVE literature begins to articulate differences between social interaction in real life and in certain virtual environments, identifying obstacles to interaction in the latter. A central aim of this study is to address the question, how is life in virtual worlds changing as "blockies" are growing up? What has changed in the new massively multiplayer environments and what issues remain?

One key observation of early CVEs was that the "blockies" were poor at conveying one's on-going activities to others (Benford et al., 1997).

Consequently, "...it is problematic for individuals to design and tailor their actions for co-participants, as they have little sense of how they are engaged in, or orienting to, the ongoing activity" (Hindmarsh et al., 1998, p. 8). For example, in one environment it was not possible for an on-looker to distinguish a user placing a virtual lamp in a particular spot from temporarily setting it down in the course of moving it (due to limitations of the interface; Hindmarsh et al., 1998, p. 7). Yet in face-to-face, an actor can "display the temporary nature of its setting-down through their visual conduct and demeanor" (Hindmarsh et al., 1998, p. 7). Conveying such on-going activity not only enables others to make sense of that activity but also to coordinate their actions more seamlessly. Information about a user's current activity can give clues about the individual's "availability for interaction. The aim here is to convey some sense of how busy or interruptable a person is" (Benford et al., 1994, p. 658). In this paper we explore and expand this idea of "conveying information about activity" in the context of entertainment-related massively multiplayer virtual worlds. We begin to answer the question, to what level of granularity do player activities need to be represented publicly in order to adequately provide for their accountability and tight coordination?

### 3.2. ETHNOMETHODOLOGICAL ACCOUNTABILITY

As the CVE literature has demonstrated, ethnomethodology and conversation analysis (Bowers et al., 1996; Hindmarsh et al., 1998, 2001) provide many useful concepts and findings in understanding avatar-mediated interaction. A major concept in ethnomethodological theory is that of "accountability." Garfinkel (1967, p. 1) writes, the central recommendation of ethnomethodological studies "...is that the activities whereby members produce and manage settings of organized everyday affairs are identical with members' procedures for making those settings 'account-able.'" In other words, a key feature of the shared methods for enacting ordinary social settings is that they also provide for the recognizability of those settings to other members on the scene. Lynch (1993, pp. 14–15) elaborates this idea by breaking ethnomethodological accountability down into six sub-claims: (1) that social activities are orderly and that this orderliness is; (2) observable; (3) ordinary; (4) oriented; (5) rational; and (6) describable. While we will not review each of these points, there are two that are especially useful for understanding avatar interaction. First, the orderliness in real-life social activities is "observable" (#2): "The orderliness of social activities is public; its production can be witnessed and is intelligible rather than being an exclusively private affair" (Lynch, 1993, p. 14). In other words when members merely observe the actions of other members, they can recognize their sense, to varying degrees. Second, this sense or orderliness is "oriented" (#4):

"Participants in orderly social activities orient to the sense of one another's activities, and while doing so they contribute to the temporal development of those activities" (Lynch, 1993, pp. 14–15). That is, participants orient their own actions to the publicly observable sense of the other participants in the setting. It is this publicly displayed sense of one's actions that matters for interaction and not what one privately intends to be doing. This is not to say that this sense-making and the resulting actions are never problematic. Rather as Garfinkel (1967, p. 1) argues that social settings are "an endless, ongoing, contingent accomplishment": at any step troubles in understanding can become evident and eventually repaired (see Schegloff, 1992). In addition, an observer's ability to understand an activity depends in part on his or her competence in the practice in question. For this reason, our direct participation in the game worlds and communities as part of the virtual ethnography was crucial.

Conversation analysis (Sacks et al., 1974) deepens the concept of "accountability" with the concept of "projectability." It shows that the accountability of real-life social activities enables participants to achieve *tight coordination* in interaction. For example, Sacks et al. (1974, pp. 707–708) observe that in ordinary English conversation, at the place where turn transition tends to occur, there tends to be minimal overlapping talk and minimal gaps between turns. In other words, the participants are good at starting their next turns at just the right time. How is this possible? Sacks et al. (1974, pp. 707–708) show that such tight coordination in turn-taking is possible because a variety of features of turns-at-talk make them "projectable." That is, by observing early parts of an unfolding turn-at-talk, other participants can see *who* (if anyone) the turn selects as the preferred next speaker and *where* it is likely to end based on "turn constructional units," such as, words, phrases or clauses. Thus, turns-at-talk are not only accountable in a general sense, but they are also projectable in a more fine-grained sense.

Not only are turns-at-talk projectable but so too are nonverbal gestures. For example, by observing the early stages of a pointing gesture, recipients in the situation can begin to predict in advance in what general direction it will go and what the likely referent will be and thereby begin to design their responses and deliver them with precision timing. Similarly Streeck (1995, p. 104) shows for iconic, or descriptive, gestures that "they *preface* speech units and *prefigure* the concepts communicated by them. They thus enable recipients – in varying degrees and depending upon contextual information provided by prior talk – to *anticipate* conceptual profiles of subsequent talk." Thus the real-time unfolding of gestures is as crucial to that of talk in enabling participants to achieve tight coordination in interaction.

So ethnomethodology and conversation analysis provide some concepts for understanding how intersubjectivity and tight coordination are possible in real-life face-to-face interaction. In subsequent sections, we will argue that intersubjectivity and tight coordination in virtual worlds likewise depend on the public observability of on-going player actions.

### 3.3. MEDIATED COMMUNICATION AND AWARENESS

While what participants are doing with their bodies in a face-to-face social setting is observable and therefore accountable, this is not the case in mediated communication. In forms of social interaction that are mediated by a technology (e.g., telephone, mobile phone, email, instant messaging), participants have varying degrees of access to what the other is doing in his or her physical environment. So for example with telephone interaction, a speaker can hear a recipient's voice in real time, but cannot see what else the recipient might be doing or who else he might be interacting with. In addition, in deciding to make the call, the caller generally cannot see if the desired recipient is currently busy with another activity (e.g., eating dinner) in order to determine to what degree he or she is available. Hence when face-to-face with the other, a speaker can better interpret what a recipient is currently doing and can minimize interruptions or mitigate their impacts. "Ethnographic studies of co-located cooperative work show that people tacitly and unobtrusively align and integrate their activities in a seamless and highly sophisticated manner without interrupting each other" (Schimdt, 2002, p. 292; see also Heath and Luff, 1992). However, in mediated situations this is harder or in many cases impossible.

Research on computer-supported cooperative work (CSCW) shows that "awareness" information is critical for the success of systems that support remote collaboration (Dourish and Bellotti, 1992). In this context, "awareness is the understanding of the activities of others, which provides a context for your own activity" (Dourish and Bellotti, 1992). For example, joint video feeds in distributed locations can enable one to see who is around and what they are doing (Dourish and Bly, 1992; Bly et al., 1993), multi-user text editors can enable one to see who is currently editing a document and how they are changing it (Dourish and Bellotti, 1992), and buddy lists in instant messaging systems can enable one to see who is "around," or online, and therefore possibly available for interaction (Nardi et al., 2000).

Social interaction in MMORPGs is of course an instance of computer-mediated communication and as such, it involves two different kinds of awareness information: (1) "real-world," or what a player is currently doing in the physical world and (2) "in-game," or what a player is currently doing in the virtual world. For example, when initiating a chat conversation with another player, one may be interrupting a real-world conversation the player is having with a family member *or* an in-game conversation the player is having with a third player. Ideally it would be useful for a player who is about to initiate a conversation to know if the recipient is currently busy in the real world or in the game world. However, the strategies for providing these two types of awareness information are very different. Real-world awareness information, because it is external to the system, requires the use of sensors to monitor the player's physical environment. On the other hand, in-game awareness information is internal to the system itself and so sensing is not an issue, only how to present player-activity information to other players. In today's virtual worlds, like in earlier CVEs (Bowers et al., 1996), little of either type of awareness information is usually made available to players.

In the following sections, we put aside the issue of providing *real-world* awareness cues and instead show why relaying *in-game* awareness information about a fellow player's current on-going activities in the game world is vital for managing player-to-player interaction. We argue that several particular types of in-game awareness information are crucial for making fellow players' actions publicly accountable. In the absence of this information, a fellow player's actions can be difficult to make sense of in a fine-grained way and tight coordination with his actions can at times be impossible.

## 4. Findings

In real-life face-to-face, participants use a variety of kinds of observational information about what others are doing in order to interpret others' actions and design appropriate responses. In other words, accountability and tight coordination depend on participants' access to these kinds of observational information. Three important types of such information are: (1) the real-time unfolding of turns-at-talk; (2) the observability of embodied activities; and (3) the direction of eye gaze for the purpose of gesturing. But despite the fact that today's virtual worlds provide virtual bodies, or "avatars," for players to control, these avatars display much less information to other players than real bodies do. For example, they only partially display the three kinds of information listed above or not at all. Below we discuss the impact of the lack of each type of information on players' ability to tightly coordinate their activities.

### 4.1. REAL-TIME UNFOLDING OF TURNS-AT-CHAT

In face-to-face conversation, you can hear the talk of a co-participant as it unfolds in real time. Consequently, you can do things like wait until the turn has finished, anticipate precisely when it will end and start your turn there, see where the turn is going, and correct it mid-course. In other words, turns-at-talk are "projectable" (Sacks et al., 1974). For example, the following transcript (from one of the author's other studies) is of a real-life face-to-face interaction over the service counter at a quick-print shop in which a customer places an order for a bound document.

**(1) Real-Life Face-to-face: Quick-Print Shop**

```
01   Emp:     Can I help you?
02            (0.2)
03   Cus:     Yeh.
04            (0.4)
05   Cus:     I wan' to uh:
06            (1.2) ((puts down stack & removes top sheets))
07   Cus:     thro::w [(0.4) this into a binder.
08   Cus:             [((flips thru top sheets
09            (0.3)
10   Emp:     M'kay?
11            (0.5)
12   Emp:     What type uh:,
13            (0.2)
14   Cus:     Whadda ya [got.]
15   Emp:               [Jst ]uh (0.2) just a regular
16            [binder?]
17   Cus:     [Yeah. S]omething tha' just uh:
18   Emp:     Like uh three ho:le,=
19   Cus:     =Well actually I was thinkin'of- just
20   Cus:     sorta [uh]
21   Emp:           [or] binding.
22            (0.1)
23   Cus:     Yeah binding. [°Some sort of] binding.°
24   Emp:                   [    Okay.    ]
```

Conversation analysis has shown that a generic feature of real-life social interaction is that gaps and overlaps between turns are minimal (Sacks et al., 1974). Short gaps of around 0.2 seconds are common (lines 02, 09, 13 and 22); no gap occasionally occurs (lines 18–19); and small overlaps are not uncommon (lines 14–15, 15–16, 20–21 and 23–24). In fact, in real life, one second is considered the "maximum standard" length of a pause within conversational sequences (Jefferson, 1989; this does not apply *between* sequences as in the case of "continuing states of incipient talk" Schegloff and Sacks, 1973). In general the ordinary pace of a conversation is marked by pauses between turns of less than one second. When pauses last longer, they become noticeable as possible indications that something else is going on. For example, the 1.2-second pause at line 06 is noticeable for its length, yet is accountable as a delay caused by the speaker manipulating his document which he begins to flip through precisely when he says, "this" (lines 08–09). In other words, turn-taking in real-life conversation tends to be tightly coordinated and fast, which is made possible by the projectability of turns-at-talk. By hearing a turn unfold in real-time, a hearer can anticipate where the turn is likely to end and can start his or her turn promptly at the next "transition relevance place" (Sacks et al., 1974). In addition, because one can hear that a current speaker is taking a turn on a particular topic, he or she can refrain from talking at the same time and introducing a new topic until the current topic is closed. In other words, speakers tend to achieve one speaker at a time (Sacks et al., 1974) and one topic at a time.

In virtual worlds this kind of projectability and tight coordination among turns-at-chat is usually not possible. Most virtual worlds employ a text-chat system very similar to IRC or IM in which you cannot see a player's turn unfold in real time (Garcia and Jacobs, 1999; Herring, 1999; and also Cherny, 1999). In IRC the construction of chat messages is *private* and only becomes public all at once when the player presses the ENTER key. This simple feature makes the organization of turns and sequences dramatically different from voice interaction as Garcia and Jacobs (1999) demonstrate: you cannot achieve "one speaker at a time," pairs of actions (e.g., question-answer, greeting-greeting) become interspersed with other pairs, lengthy pauses develop between pairs, turns are not repaired by others mid-course, and more. Because they employ IRC-style messaging systems, chat in today's virtual worlds exhibits all of these organizational characteristics.

Pavel Curtis (1997) describes some of these features of conversation in MUDs. First, the pace of chat conversations is slower than that of face-to-face conversation. "The most obvious difference between MUD conversations and those in real life is that the utterances must be typed rather than simply spoken. This introduces significant delays into the interaction and, like nature, MUD society abhors a vacuum" (Curtis, 1997). Such delays are not simply caused by slow typing, but also by the fact that you cannot see another player constructing his or her turn. Second, unlike face-to-face conversation, chat conversations tend to have multiple topics or threads progressing simultaneously within the same conversation.

> Even when there are only two participants in a MUD conversation, it is very rare for there to be only one thread of discussion; during the pause while one player is typing a response, the other player commonly thinks of something else to say and does so, introducing at least another level to the conversation, if not a completely new topic. These multi-topic conversations are a bit disorienting and bewildering to the uninitiated, but it appears that most players quickly become accustomed to them and handle the multiple levels smoothly (Curtis, 1997).

Each of these features of chat conversations are demonstrated in the following example from Star Wars Galaxies, in which Atac and Ka'Bob meet at a shuttleport (Figure 2, Excerpt 2).[3]


*Figure 2.* Message-by-message text chat.

**(2) Star Wars Galaxies: Meeting Ka'Bob (Atac's perspective)**

```
01   00:40  Ka'Bob:  sup
02                   (4.1)
03   04:48  Atac:    nada
04   04:48  ((Ka'Bob turns head toward Atac))
05   10:16  ((Ka'Bob turns to face Atac))
06                   (2.1)
07   12:24  Atac:    I like your name
08   13:34  ((Atac zooms in))
09   15:24  ((Atac pans right))
10                   (3.1)
11   15:34  Ka'Bob:  hehehe ok
12                   (2.2)
13 →17:50   Ka'Bob:  why thx
14                   (6.1)
15   23:58  Atac:    hey can you train me in medic?
16                   (5.1)
17   29:18  Ka'Bob:  you ever had one lol
18          ((Ka'Bob's avatar laughs automatically))
19                   (4.0)
20 →33:16   Ka'Bob:  sure
21   36:08  Ka'bob invites you to join a group.
22                   (4.0)
23   37:12  Atac:    no
24   38:14  ((Atac accepts invitation by clicking yes))
25   39:14  You have joined the group.
26                   (2.0)
27   41:04  Ka'Bob:  lol
28          ((Ka'Bob's avatar laughs automatically))
29                   (7.4)
30   48:40  Ka'Bob:  ok i have not too
31          ((Ka'Bob's avatar shakes head automatically))
32                   (6.1)
33   54:54  Atac:    I should be ready for master
34                   (0.5)
35 →55:46   Atac:    hehe
36   55:46  ((Training window appears))
```

The first thing that is striking about this conversation is that it has long pauses and gaps between turns-at-chat. They range in length from 0.5 to 7.4 seconds with an average of about 4 seconds. This is compared to 0.0- to 0.5-second silences in face-to-face conversation, such as that in Excerpt 1. Thus, even the normal pauses in such text chat are considerably longer than the 1-second "maximum standard silence" (Jefferson, 1989) in real-life conversation. One obvious reason for this is that, as Curtis (1997) points out, is that the turns are typed rather than spoken. However, it is also slower because recipients cannot read a turn as it unfolds. They must wait until the very end before they can begin reading it. This takes longer than in a face-to-face conversation in which a recipient can listen to the turn as it unfolds and be ready to jump in with a response exactly when it ends.

In addition to producing long pauses and gaps, the players in Excerpt 2 also display difficulties in coordinating sequences of chat. As Curtis (1997) and Herring (1999) point out, chat conversations tend to develop multiple simultaneous topical sequences or "threads." In this excerpt there are three. One involving a greeting (lines 01, 03 and 11), a compliment of a name (lines 07, 13, 17, 23, 27, 30 and 35), and a request for training (lines 15, 20 and 33). These sequences are distinct in terms of topical continuity, but they are interlaced with each other. As Garcia and Jacobs (1999) nicely demonstrate, with "quasi-synchronous" or IRC-style chat, users cannot predictably place responses to prior turns adjacent to those turns as is characteristic real-time conversation. For example, Ka'Bob's turn "why thx" (line 13) does not respond to the immediately prior turn in which he says, "hehehe ok" (line 11). Instead it is a response to the earlier compliment by Atac, "I like your name" (line 07). Similarly, Ka'Bob's turn, "sure" (line 20), does not respond to the immediately prior turn in which he says, "you ever had one lol" (line 17), but rather to Atac's earlier request, "hey can you train me in medic" (line 15). Finally, Atac's laugh, "hehe" (line 35) does not respond to the immediately prior turn in which she says, "I should be ready for master" (line 33), but to Ka'Bob's earlier admission, "ok i have not too" (line 30; i.e., have not had a ka'bob). In each case, these "late" turns (arrows) result from the fact that just prior (at lines 07 and 11; 15 and 17; 30 and 33) both players are typing simultaneously on different topical trajectories but do not know it. Thus, in this kind of system, players cannot tightly coordinate their turns to predictably achieve "one speaker at a time" or "one topic at a time" as is characteristic of real-time conversation. This is not to say that such chat is unintelligible. As Curtis (1997) suggests, players quickly become accustomed to this alternative kind of turn-taking system.

Only one massively multiplayer virtual world we know of implements the near-real-time posting of chat messages. Unlike the other chat systems, the one in *There* posts messages on a word-by-word basis instead of a message-by-message basis. As a result, players can see each others' turns unfold one word at a time. Turns-at-chat in this system are therefore more projectable unlike those in other systems. The resulting interactions are much more tightly coordinated, less fragmented in terms of topics, and more like face-to-face conversation, as we can see in the following interaction from *There* (Figure 3, Excerpt 3).


*Figure 3.* Word-by-word text chat in *There* (left to right: A, B and C).

**(3) There: Local Tour Guide (A's perspective)**

```
01   B:   once she is done I'll take us where we can try vehicles and
02   B:   see
03   B:   the puppy [ and  stuff  like  that  ]
04   A:             [so are you the local tour] guide bret?
05        (0.2)
06   B:   yes
07        (2.0)
08   A:   do they pay you?
09        (9.6)
10        (1.0) ((dressing screen appears in front of B))
11        (1.0)
12   B:   no, they used [  to  pay  ] Mentors, but no- not anymore
13   A:                 [they should]
14        (1.8)
15   A:   hmmm
16        (5.0)
17   A:   ((clicks on B, selects "look at" & turns head toward B))
18        (3.0)
19   B:   I was working to (4.0) qualify for being a paid Mentor
20        when they
21        (2.3)
22   B:   discontinued that program
23        (1.3)
24   B:   they used to pay
25        (2.3)
26   A:   too [ bad ]
27   B:       [house] owners for throwing events
28        (2.3)
29   B:   also, but did
30        (2.3)
31   B:   away with that too :(
32        (4.0)
33   A:   cutback huh?
```

From this annotated log, we can see that the pace of the conversation is faster than the chat in Excerpt 2, with many pauses of only 0.2–3.0 seconds between turns (lines 05, 07, 14, 23, 25 and 28) and also overlapping turns (lines 03–04, 12–13 and 26–27). As others have pointed out about chat in *There*, ‘‘Overlapping talk affords replying to chat as it is produced, rather than waiting to the end of a turn, hastening conversation’’ (Brown and Bell, 2004, p. 354). Like in real-life face-to-face, players can place a turn precisely where a prior turn is likely to end. While A initiates a next topic, ‘‘so are you the local tour guide bret?’’ (line 04) in overlap with B’s current turn (line 03), she comes in at a ‘‘transition relevance place’’ (Sacks et al., 1974), a point at which the current turn is possibly complete. B’s turn appears possibly complete after the word ‘‘puppy’’ (line 03) although he is not in fact finished. Similarly at line 13, A responds in overlap with B’s current turn (line 12), but she produces her turn just after the transition relevance place after the word ‘‘no.’’ Thus turn taking in *There* is much closer to that in real-life face-to-face than it is in the chat systems of other massively multiplayer virtual worlds.

In addition, in this kind of system, a pause like the 9.6-second pause at line 09 becomes noticeable as ‘‘long,’’ unlike in IRC or MUDs. It is considered a ‘‘pause’’ rather than a ‘‘gap’’ (line 07 or 16) because in the immediately prior turn, A selected B to speak next by asking a question (line 08). However, instead of immediately typing a response, B appears to do nothing for a long 9.6 seconds. At the end of this pause, a dressing screen appears for 1.0 second in front of B (line 10) and when it disappears, he is wearing a different shirt. After another 1.0 second (line 11), B finally answers A’s question (line 12). Thus, in this kind of system, pauses become more informative. When there is a long pause, other players can see that it is *not* due to the player typing a long message and therefore can begin attribute it to some other activity. Although A can rule out ‘‘taking a turn’’ as a possible cause of the delay, she cannot see what else he is doing. It’s not until the dressing screen appears that she can infer that he was browsing his inventory for a shirt during that 9.6 seconds.

Even more strikingly, unlike Excerpt 2, the conversation in Excerpt 3 is not characterized by multiple simultaneous topical sequences. A initiates a topic shift about B’s ‘‘tour guide’’ behavior (line 04), and the players sustain this single topical trajectory throughout the remainder of the excerpt. Because the players can see *when* the other is taking a turn and see *what* topic it is addressing, they can (if they choose) refrain from introducing new topics until prior topics are properly closed as people tend to do in real-life conversation. However, because they are using persistent text rather than voice, they may choose to pursue multiple topics if they wish.

Thus hiding turn construction from the public eye versus revealing it has a major impact on the sequential organization of chat interactions. Although IRC- or MUD-style chat systems have proven to be powerful forms of communication in their own right (see O’Day et al., 1998; Grinter and Palen, 2002), we argue that near-real-time chat systems, such as that in *There*, are better suited *as a substitute for talk in the avatar interaction systems* of 3D virtual worlds. This is based on two reasons. First, *if* the world designers' goal is to create a user experience that feels more or less like a face-to-face, then MUD-style chat is not the best choice given the "dead periods" between messages and the other sequential implications of hiding turn construction. Second, and perhaps more importantly, in a virtual 3D world, hidden turn construction makes it more difficult for users to tightly coordinate their turns-at-chat with other in-game activities. For example, in the following annotated log (Figure 4, Excerpt 4) from EverQuest II, Ataca and Rattington are looking for animals to hunt in order to gain experience points.


*Figure 4*. "Late" turns-at-chat and joint activities.

**(4) EverQuest II: Hunting Armadillos (Ataca's perspective)**

```
01 07:38:16  ((Ataca stops running))
02 07:38:54  ((Ataca targets an armadillo))
03 07:41:38  ((Rattington stops next to Ataca))
04 07:41:38  Ataca points at a banded armadillo.
05           ((Ataca's avatar points))
06 07:43:06  ((Ataca rearranges icons in her toolbar))
07 07:50:00  ((Ataca mouses over an attack spell))
08 07:51:00  ((Ataca mouses over armadillo))
09 07:54:16  ((Ataca initiates an attack by clicking a
10            spell icon in her toolbar & her avatar begins
11            a spell-casting animation))
12 07:57:42  You surround A young armadillo with arcane chains!
13 07:58:18→ Rattington says to the group, "I see two that
14           are grouped but I think we could take them."
15 08:00:10  ((Ataca clicks another spell icon))
16 08:01:58  Rattington hits banded armadillo for 4 points of
17           crushing damage.
18 08:03:34  You strike A banded armadillo with a storm of lightning!
19 08:04:08  Rattington hits a banded armadillo for 3 points of
20           crushing damage.
21 08:05:02  Ataca says, "hehe"
```

At the opening of this encounter, Ataca stops running (line 01) and targets a pair of ‘‘banded armadillos’’ by clicking on them (line 02). She then points to one, a public action that Rattington can see through a pointing animation of the avatar (line 05) and a text emote in the chat log (line 04). In this context, such pointing can be seen as a possible proposal to attack the armadillos. Ataca then spends about 8 seconds rearranging icons on her toolbar (lines 06), an activity that Rattington cannot see. After modifying her toolbar, she initiates an attack on the armadillos by clicking a spell icon on her toolbar (lines 09–11). This action is observable to Rattington by a spell-casting animation of Atac’s avatar, a sound, and eventually a text message after the spell is cast (line 12). But 4 seconds *after* Ataca initiates the attack, Rattington says, ‘‘I see two that are grouped but I think we could take them’’ (lines 13–14) referring to the two armadillos. In other words, he offers his assessment of the armadillos as suitable targets, but his turn appears too late to be consequential. Rattington obviously started typing his assessment *before* Ataca initiated the attack, but she could not see this typing, and therefore did not know to wait to see what her companion had to say before attacking. Rattington eventually joins in the battle (line 16), while Atac chuckles at their apparent lapse in coordination (line 21). Thus, with these kinds of late turns-at-chat, making sense of the text itself is not necessarily problematic, but coordinating them with the initiation of other joint activities is. We have found that such slippages in coordination are common in massively multiplayer virtual worlds.

In sum making the composition of turns-at-chat public, rather than private, enables a kind of text-based conversation that is organized much more like face-to-face than the kind of text chat that is used in almost all virtual worlds today. Seeing *that* the other is taking a turn and seeing *what* he or she is doing with that turn is a critical type of awareness information that enables players to tightly coordinate turns-at-chat with each other and, perhaps most importantly, with other joint activities.

### 4.2. OBSERVABILITY OF EMBODIED ACTIVITIES

Although avatars are currently very good at providing real-time feedback to one’s fellow players regarding their orientation (i.e., facing) and movement through space, they are very poor at giving off other kinds of awareness information about what a player is doing. One feature of human bodies, which has an enormous social impact, is that you can see what a person is doing, to varying degrees, simply by looking at them. For example, any competent member of the culture can recognize at a glance what the various people in Figure 5 are doing at a gross level.


*Figure 5*. Human bodies conducting ordinary activities.

Looking through a bag, consulting a map, reading a book, and talking on a phone involve particular configurations of the body with particular objects that are recognizable by competent members. Thus, such activities are publicly accountable merely through visual inspection of the body. In Goffman's (1959, p. 2) terms, human bodies "give off" all kinds of information even when the actor is not intending to communicate.

In contrast, the avatars in massively multiplayer virtual worlds do not "give off" nearly as much information about the player's current state. Players must explicitly communicate such information through commands or text. For example, although players routinely look through bags, consult maps, read "books," chat to remote friends, and the like, avatars simply stand still and do nothing (see Figure 6). Whether due to oversight or cost considerations, game developers do not currently tend to provide avatar animations for this class of "personal" player activities. Thus, when engaged in them, players do not automatically "give off" information about what they are doing as they would in real life.

*Figure 6*. Avatars remain idle when players are active.

The effects on the player experience of keeping such activities private are very similar to those of concealing turn construction in chat: (1) interaction is marked by intermittent dead periods; (2) idle avatars give a false sense of availability; and (3) slippages arise when trying to tightly coordinate joint activities. We discuss each of these effects in turn.

First, avatars often appear like lifeless zombies when they stand and do nothing. In real-world public places, standing still and doing nothing is a surprisingly unnatural thing to do. As part of his sociology courses, McGrane (1994, chapter 2) instructs his undergraduate students to go to any public place and to stand still and do nothing for five minutes. He explains that by "do nothing," he means refrain from doing "waiting" by visibly glancing at one's watch or reading a magazine and refrain from doing "looking for someone" by repeatedly glancing off into the distance and the like. The students are instructed to stand motionless with their arms at their sides. The results of this exercise are that it is surprisingly difficult to deviate from the shared practice of making one's business accountable to others on the scene and that it opens students' eyes to a subtle layer of mundane public self-presentation (Goffman, 1959) that we generally take for granted. By not animating avatars for many routine activities, game developers have to some extent simulated a world populated by McGrane's deviant sociology students.

Second, private activities can give a false sense of one's interactional availability. Consequently, interactional sequences can get initiated when recipients are in the middle of another activity. Also, delays and noticeably absent responses by recipients become difficult to interpret. For example, in EverQuest Online Adventures (EQOA) (Figure 7, Excerpt 5), a group of three players, A, B and C, has just arrived at the bustling coach house in the city of Highpass Hold in order to find a high-level character to help them with a difficult quest.

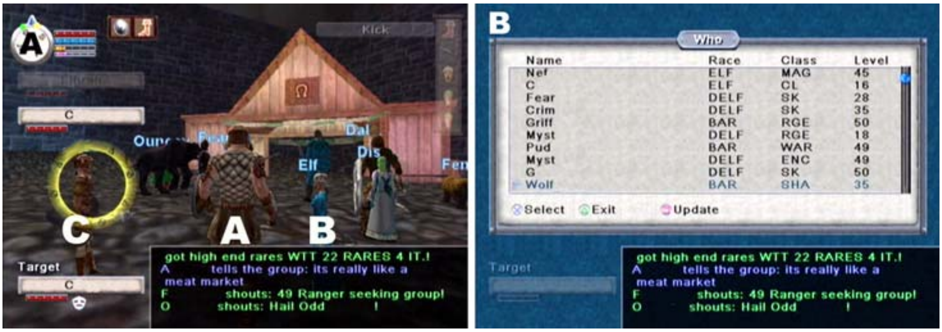


*Figure 7.* Searching a directory privately (left, A's view; right, B's view).

**(5) EverQuest Online Adventures: Meat Market (B's perspective)**

```
01   00:07:22   ((B opens her main menu))
02   00:09:19   ((B opens her player directory))
03 →00:19:18   A tells the group: its really like a meat market
04   00:22:23   ((B scrolls 21 names down list))
05   00:50:22   ((B selects Sven))
06   00:53:14   ((B selects Tell command))
07   00:57:14   ((B begins typing a private message to Sven))=
08   01:14:01   =C tells B: i have a bite
09   01:14:07   =((B continues typing message))
10   01:30:12   B tells Sven: hi, could u by any chance give my
11              group a little help with LVL 17 runnyeye quest?
```

At the beginning of the encounter, B and C are standing near A and appear to be doing nothing. Their avatars stand motionless except for a periodic "idle animation" (not controlled by the player) in which it looks from side to side. Thus from A's perspective, B and C appear to be available for interaction. A then says to the group, "it's really like a meat market" (line 03), remarking on the unusually large volume of spam messages from other players looking to form a group. Ordinarily for this tight-knit group, such a comment would have received at the very least an "lol" (laugh out loud) and more likely a subsequent joke that builds off of it. However, in this case, B and C fail to give any response. Why B and C do not respond at that moment is not immediately discernable from A's perspective. However, from B's perspective, we can see that she was in fact busy doing something. At the moment A made the comment, B was busy accessing a private system menu (lines 01–02 and 04–06), the "who list," a directory of all characters in the vicinity. B finds a particular player, Sven (line 05), and writes him a private message asking him if he will help with the quest (lines 06–07 and 09–11). In order to respond to A, B would have had to cancel what she was doing and navigate back to the main view, as *EQOA* does not allow the player to chat and access system menus simultaneously. As a result, the system does not allow players to interleave talk and UI activities. This sharply contrasts with face-to-face conversation in which talk and activity can be interleaved with great precision, as we saw in Excerpt 1.

During B's work at trying to solicit help, none of which is visible to the other group members, C sends a message to the group, "i have a bite" (line 08), revealing that he too has been soliciting help and has already received a response. From this, A can guess why C did not respond earlier, but at this point (over a minute later) he has no visible clues as to why B did not respond. In addition, while the design of most systems would have allowed B to respond to A without canceling her private menu activities, none of the systems we've seen would have shown A that B was busy browsing the "who list."

Such noticeably absent responses are not of course entirely unaccountable to players. They know from experience the variety of reasons that another player might not be able to respond – accessing one of many system menus, chatting privately with another player, missing the message in the flood of public spam, going briefly away from the keyboard, and more. They also know that it could be an intentional ignoring or a snub. But while they can account for such absent responses in a very general way, they cannot on most occasions determine the *particular* reason for the absence unless the other player offers a textual account.

Third, the fact that players routinely cannot see *that* a fellow player is engaged in a particular activity nor see *what* kind of activity it is makes it more difficult for players to tightly coordinate their actions. The coordination of joint activities requires adequate feedback on the other's current state. For example (Figure 8, Excerpt 6), two players of Star Wars Galaxies (SWG) are doing a set of missions together to gain access to Jabba the Hutt's palace. They have just completed one objective, killing a kaadu, and are about to travel on their speeders to the next.

In this case, Atac is trying to coordinate two different activities simultaneously: replenishing the stimpacks (used for healing in combat) in her toolbar and traveling with Nike to their next destination. When Nike appears to be about to leave the scene on his speeder (line 06), Atac types "wait" (line 07) and then opens her inventory about 1.5 seconds later (line 09). However, due to the system design, there is no public signal that Atac has opened her inventory. As a result, Nike cannot see what he is waiting for. In the absence of this awareness information, he takes a guess, "did you get it?" (line 11) referring to a quest item from the dead kaadu. Nike then begins to drive off (lines 12–13) as Atac requests clarification by saying "get what?" (line 14), displaying that, that was not in fact the reason for the waiting. Nike drives off (line 15) and Atac selects some stimpacks and drags them to her toolbar (lines 16–19). Thus Atac fails in getting Nike to wait for her while she browses her inventory and replenishes her toolbar.


*Figure 8*. Searching an inventory privately.

**(6) Star Wars Galaxies: Jabba's Missions (Atac's perspective)**

```
01  01:38   ((Nike runs past Atac toward speeders))
02  02:44   Atac: is that it?
03  03:10   ((Atac turns around and looks in Nike's direction))
04  04:16   ((Atac begins to follow Nike))
05  09:34   ((Atac inspects an empty slot in her toolbar))
06  12:06   ((Nike mounts his speeder and begins turning it))
07  13:46 →Atac: wait
08  13:52   ((Nike stops turning speeder))
09  15:15 →((Atac opens inventory))
10  17:24   ((Atac opens backpack within inventory))
11  20:52   [GroupChat] nike: did you get it?
12  28:48   ((Nike turns speeder))
13  29:46   ((Nike begins to move forward))
14  30:24   [GroupChat] atac: get what?
15  31:56   ((Nike passes Atac on speeder))
16  32:16   ((Atac mouses over and inspects some stimpacks))
17  37:28   ((Atac drags stimpacks))
18  38:54   ((Atac drops stimpacks onto empty slot in toolbar))
19  40:18 →((Atac closes backpack))
20  41:00   ((Atac turns around to look for Nike))
21  42:12   [GroupChat] nike: the kaadu egg
```

We see then that keeping players' UI activities private makes coordination with other players more difficult, gives a false sense of one's interactional availability, and more generally creates an "unnatural" public atmosphere in which avatars stand idle. However, through our ethnographic observation, we discovered an uncommon practice for mitigating the coordination troubles that result from private menus. Sometimes "role players"[4] use the custom emote feature to write text descriptions of private actions for others to see. For example, (see Figure 9, Excerpt 7), one player[5] uses this practice to reveal to another player the fact that she is reading in-game email, which is otherwise a private activity. Atac and Galen are waiting for a player-hosted event, a “live market” or swap meet, to start and have been having a conversation for several minutes while they wait.

*Figure 9*. Revealing private activities with custom emotes.

**(7) Star Wars Galaxies: Datapad (Atac’s perspective)**

```
01   38:07  Galen: looking for dancers for the club still....it's new
02   38:15  Atac [shouts,playfully]: Hey Sis
03   38:18  Atac [playfully]: over here
04   38:28  Atac [playfully]: Ella's my sister
05   38:34  Galen: ah
06   38:46  Atac [playfully]: ooh I think she just hopped on the shuttle
07   38:52  Galen: looks that way
08   38:58  Atac [playfully]: oh well
09 →39:15  (playful) Atac looks at her datapad.
10   39:20  ((Atac opens email and selects message A))
11   39:22  Ella tells you: ahh crap, i didn't see it till too late
12   39:32  You reply to ella, 'no problem'.
13   39:39  You reply to ella, 'coming to the market?'.
14   39:43  ((Atac selects email message B and scrolls))
15   39:43  Ella tells you: i'll catch up with you in a bit :)
16   39:47  Ella tells you: yeah i should be through there
17   39:50  You reply to ella, 'ok'.
18   39:56  ((Atac selects email message C and scrolls))
19   40:03  ((Atac selects email message A again and scrolls))
20   40:09  ((Atac selects email message B again and scrolls))
21   40:16  ((Atac selects email message D and scrolls))
22   40:29  ((Atac closes email))
23 →40:39  (playful) Atac puts datapad away.
24   41:20  Galen: hrm....not many here yet
25   41:25  Atac [playfully]: hehe
26   41:29  Atac [playfully]: you can say that again
```

During her conversation with Galen, Atac notices a friend, Ella (her role-playing “sister”), walk to a nearby shuttle stop. Atac tries to greet her by shouting "Hey Sis" (line 02) and saying "over here" (line 03), but receives no response. Atac and Galen then conclude that Ella must have boarded the shuttle and left (lines 06–08).

After a 17-second lapse in the conversation, Atac opens her in-game email and begins to look at a message announcing the "live market" (line 10). But instead of merely opening her private email window (the usual practice), she first types a custom emote "Atac looks at her datapad" (line 09) for Galen to see. She then begins to read her email (line 14; see email window in Figure 9 on right side of screen). Then, after 7 seconds, she receives the overdue response from Ella via a private message or "tell" (line 11), also invisible to Galen, confirming that she indeed missed the earlier greeting because she was getting on the shuttle. Atac then has a brief conversation with Ella (lines 12–13 and 15–17). In addition to this private conversation, Atac continues to search through four email messages regarding the live market looking for some bit of information (lines 14 and 18–21). The system provides no cues to Galen that Atac is either browsing her email or chatting remotely. However, because of Atac's custom emote (line 09), Galen refrains from initiating conversation. Finally, Atac closes her email (line 22) and then announces this fact to Galen by typing a second custom emote, "Atac puts datapad away" (line 23). Less than a minute later, Galen can read that Atac is no longer busy with her datapad, and he initiates conversation again (line 24).

Thus a minority of "role players" in SWG have developed a workaround for the fact that the avatar system conceals certain activities from public view. As Excerpt 7 demonstrates, this practice of externalizing private activities through custom text emotes can be effective in coordinating personal activities with joint activities, such as reading email and having a conversation. However, the main drawback of this practice is that it requires quite a bit of extra work on the part of the player since he or she must type a relatively long text description manually each time. But this player practice could easily be automated: each time the player opens the datapad, an emote automatically appears saying, "Atac looks at her datapad." This is just what we did.

To test the utility of externalizing private activities, we created a "mod" or "addon" to the game, World of Warcraft (WoW). WoW provides a tool to players for modifying the user interface of the game. The tool enables players to customize the 2D user interface, but not to modify the 3D world. Using this tool, we wrote a simple addon that provides *heightened awareness* of what other players are doing. That is, it automatically produces public text emotes each time the player opens or closes a particular private menu. For example, the player opens the inventory and it says, "Player rummages through her bags"; the player opens the quest log and it says, "Player checks her To-Do list"; the player closes a trade window and it says, "Player completes a trade with X"; the player begins to type a message and it says, "Player says, "uh..."" ; and the like. We created 34 unique messages for the opening and closing of 17 different private menus.

So far, the addon has proved to be quite effective. We have tested it informally amongst ourselves and with over 20 ordinary players. Certain types of coordination between players, especially around joint travel and combat, are significantly increased. For example, in one case (Figure 10, Excerpt 8), a party of five players is questing. At one point, they must "share

*Figure 10.* Seeing others "share a quest".

**(8) World of Warcraft: Heightened Awareness Addon (Atac's perspective)**

```
01 35:36:42   Burbl says, "uh..."
02 35:38:14   Burbl says, "uh..."
03 35:42:46   ((Grr initiates an attack on a Trogg))
04 35:45:38   [Party][Burbl]: anybody shared quests w/ Noob btw?
05 35:47:32   Yuk says, "uh..."
06 35:49:16   [Party][Yuk]: spit it out, burbl
07 35:50:58   [Party][Atac]: no
08 35:52:30→  Yuk checks his To-Do list.
09 36:07:34   ((party defeats Trogg & stands around corpse))
10 36:11:10   Burbl says, "uh..."
11 36:13:04   [Party][Noob]: ty
12 36:14:42   Your share of the loot is 15 Copper.
13 36:20:18   Grr rummages through his bags.
14 36:24:30   [Party][Burbl]: again, the addon is fantastic -
15            I saw Yuk was about to share so I did not bother
16 36:26:08   Grr begins to trade.
17 36:26:08   Noob begins to trade.
18 36:31:26   Atac says, "uh.."
19 36:32:22   [Party][Atac]: me too!
20 36:33:10   Noob completes trade.
21 36:33:10   Grr completes trade.
22 36:37:36→  Yuk crumples up his To-Do list.
```

a quest" with one of the members, which involves opening the Quest Log, finding the quests for that zone, and clicking "share."

As the party is battling a Trogg (line 03), Burbl asks the group if anyone has shared the quests they are working on with Noob, the most recent member (line 04). One member, Atac, who knows the most about Noob replies "no" (line 07). A second and a half later, a text emote automatically generated by the addon appears that says, "Yuk checks his To-Do list" (line 08), the first action one would take in sharing quests with another player. After the battle ends (line 09), Noob says, "ty" or thank you (line 11), which can be seen as referring to the sharing of the quest, and Burbl comments on the addon itself, "again, the addon is fantastic – I saw Yuk was about to share so I did not bother" (lines 14–15). Atac then says, "me too!" (line 19) indicating that she likewise saw that Yuk was starting to share the quest and so did not go to the trouble of opening her quest log and looking for the relevant quests. In addition, in the post-combat period, the addon enables the party to see some of what each other is doing – browsing an inventory (line 13) and trading items with each other (lines 16–17 and 20–21). This is particularly useful in determining when the party members are ready to begin the next battle.

Despite the usefulness of the addon in facilitating player-to-player coordination, it has its drawbacks. The primary drawback is that it uses the text chat channel to communicate the public information, and as a result, it can at times flood this channel causing chat messages get lost. We chose the chat channel because it was the easiest to implement and because we could not customize avatar animations with the tool. Yet despite these limitations, it does seem to support the idea that increasing public awareness information about what fellow players are doing in the game world makes coordination easier for players.

### 4.3. DIRECTION OF EYE GAZE FOR THE PURPOSE OF GESTURING

In face-to-face interaction, people communicate not only with their mouths but also with their hands and bodies. They can use their bodies to create an infinite number of unique gestures, just as they can put words together in an infinite number of unique combinations. People use gestures in at least four distinct ways (see Kendon, 1980; McNeill, 1992): (1) to perform an action through the gesture alone (i.e., "emblems"), such as a greeting with a wave or an agreement with a nod; (2) to refer to objects by pointing (i.e., "deictics"); (3) to emphasize particular words (i.e., "beats"); and (4) to describe objects (i.e., "iconics") by simulating their shape, spatial relationships or motions. A key factor in the production of gestures, regardless of the type, is that they must be *coordinated with the eye gaze of the intended recipient*. Because gesture is a visual mode of communication, it doesn't count if the recipient does not see it. A basic resource in this kind of coordination is "holding" (Kendon, 1980; Hindmarsh and Heath, 1999) the gesture until you can see that the recipient has seen and understood it. This length of time will vary from one occasion to the next based primarily on what the recipient is doing at the moment.

In massively multiplayer virtual worlds, the systems of gesture tend to work somewhat differently. Players perform them by selecting from a list of pre-defined gestures, usually by typing "/wave," "/shrug," or the like. In some cases, these lists are short (e.g., EQOA has only 4 gestures) and in other cases, they are quite lengthy (e.g., SWG has over 350 "socials"). In addition, each of these pre-defined gestures has a fixed duration: every "point" lasts the same length of time independently of what is going on in the situation. In most systems, gestures cannot be tightly coordinated with words in the chat because gesture and chat must be done as separate commands. And in most systems, a gesture consists of two parts, an animation in the avatar and a text emote describing the gesture.

In these avatar gesture systems, most of the pre-defined gestures are ones that can be used to perform an action on their own (i.e., "emblems"). These work relatively well in part because they do not require any coordination with the chat. Pointing gestures (i.e., "deictics") are somewhat more difficult to perform because they usually rely on some text chat for their full meaning (e.g., "let's attack *that one* over *there*"), yet the chat must be typed immediately before or immediately after the gesture instead of simultaneously. This can be awkward to accomplish. Gestures that emphasize particular words (i.e., "beats") tend to be lacking since they cannot be tied to particular words, although beat-like gestures are sometimes activated by entire messages or by exclamation points. And finally, descriptive gestures (i.e., "iconics") are generally impossible because players cannot create novel gestures on the fly to mimic whatever particular object or motion they might want to describe. Furthermore, although stand-alone and pointing gestures work fairly well, still neither can be held for varying lengths of time and until the particular recipient appears to see and understand them. We demonstrate some of these issues with the coordination of gesture in virtual worlds in the cases below.

The consequences of fixed-duration gestures are nicely demonstrated in EverQuest Online Adventures (EQOA). The gesture system in EQOA is similar to that of face-to-face in some ways but very different in others. It is similar in that the orientation of the avatar determines what is displayed on the player's screen. In other words, there is more or less a one-to-one correspondence between the direction of the avatar's "eye gaze" and the player's view of the world. When interacting with others, you know that if your avatar is behind the back of another, it is impossible for the other player see your avatar and any of its gestures. Furthermore, unlike most other systems, gestures consist entirely of avatar animations and do not have a corresponding text emote. As a result, players move their avatars into other avatars' field of view and try to maintain such visibility during interaction just as people do in the physical world.


*Figure 11.* Coordinating a bow with another player's view.

**(9) EverQuest Online Adventures: Meeting Vindar (Doriel's perspective)**

01 09:48 *Doriel: 9 Ranger seeking group!*
02 12:18 ((Vindar approaches and stops right in front of Doriel: Frame 1))
03 13:50 ((Doriel targets Vindar))
04 15:52 ((Vindar's profile appears: Level 16))
05 20:04 ((Doriel opens QuickChat menu))
06 22:06 ((Vindar begins to run off to Doriel's left: Frame 2))
07 22:42 ((Doriel selects "Communicate" from QuickChat menu))
08 23:00 ((Vindar begins to back up))
09 →23:20 ((Doriel aborts the QuickChat menu))
10 24:16 ((Vindar runs off the left side of the screen))
11 24:24 ((Doriel begins to pan the camera left))
12 25:20 ((Doriel opens QuickChat menu as Vindar is at a distance in front of her and slightly past her centerline to the right))
13 25:42 ((Doriel pans the camera right as Vindar is about to disappear off the right edge of the screen))
14 27:12 ((Doriel selects "Communicate" from the QuickChat menu just after Vindar passes her center line to the left))
15 27:34 ((Vindar begins to move toward Doriel and to her left))
16 29:10 ((Doriel pans camera to her left and achieves alignment with Vindar: both are facing each other though at a slight distance))
17 29:12 ((Doriel selects "Action" from the QuickChat menu as Vindar continues to approach head on))
18 →29:58 ((Doriel selects "Bow" from the QuickChat menu as Vindar is close and still facing her: Frame 3))
19 30:04 ((Doriel begins to bow as Vindar begins to run to her right))
20 30:44 ((Vindar stops and turns toward Doriel))
21 30:56 ((Doriel reaches the lowest point of her bow))
22 31:26 ((Doriel completes her bow))
23 31:46 ((Vindar turns to face Doriel))
24 32:10 ((Doriel turns right to face Vindar))
25 32:42 ((Vindar begins to bow))
26 33:20 ((Doriel stops turning))
27 33:54 ((Vindar reaches the lowest point of his bow: Frame 4))
28 34:26 ((Vindar completes his bow))

However, like all systems we've observed, gestures in EQOA have a fixed duration. This can make timing the extent, or "stroke" (Kendon, 1980), of the gesture with the "gaze" of the recipient's avatar challenging at times. For example, in one case (Figure 11, Excerpt 9), Doriel experiences some slippages when trying to bow to Vindar (circled in red).

This encounter begins with Doriel issuing a looking for group, or "LFG," command which appears in the public chat channel (line 01). Vindar promptly approaches Doriel and faces her (line 02; frame 1) thus creating the optimal configuration to exchange gestures and indicating a possible interest in grouping. Instead of gesturing or chatting with Vindar immediately, Doriel first takes a moment to check his level, via a private profile menu (lines 03–04), to see if they are a good match (they are not). She then begins to initiate the "bow" command, which requires navigating a series of four "QuickChat" menu options: Open QuickChat menu, select Communicate, select Action, select Bow. But Doriel gets only half way through this sequence (lines 05 and 07) when Vindar begins to runs off to Doriel's left (frame 2, line 06). From Vindar's perspective, he has waited 8.6 seconds while Doriel appears to do nothing and thus might appear disinterested in grouping with him.

Now since Doriel can see that Vindar is no longer facing her, and thus is unlikely to be able to see a bowing gesture, she aborts it (line 09). She then attempts to re-establish a facing configuration with him as he runs around the area (lines 10–16). Eventually she initiates a second bow just as Vindar is again facing her (frame 3, line 18), but then he runs off again to her right (line 19). Because the player cannot control the duration of the bow, Doriel completes it (line 22) *before* Vindar completely stops and faces her (line 23). The timing is therefore slightly off, yet it is just adequate to enable Vindar to see it. Seeing now that Doriel is interested in interaction, he returns the bow as the two re-establish a facing configuration (frame 4, lines 23–28).

This encounter nicely demonstrates the difficulty for players in timing a gesture when they cannot control the duration of it. Had Doriel been able to "hold" the bow for an arbitrary length of time, she could have simply maintained it until Vindar noticed it rather than aborting it (line 09). Instead the player had to work unnecessarily hard, navigating four menus, to time the 1.5-second bow animation so that at least part of it would be visible to her moving target.

However, a one-to-one correspondence between avatar orientation and player view in EQOA is not the norm among virtual game worlds today. In the majority of systems, there is an *independence between the player's view and the avatar's orientation*, in other words, the orientation of the avatar does not always determine what the player can see on his or her screen. Player's can manipulate their view independently of the avatar's orientation, zooming in and out and panning 360-degrees. As a result, you can never be certain what another player can see. This kind of system is sometimes called "virtual cameras" (Hindmarsh et al., 1998).

In addition, in these same systems (e.g., SWG, EQ2 and WoW), gesture animations have corresponding text descriptions of the gesture. For example, when the player types "/wave," the avatar waves *and* the message, "Atac waves to Sin'thea" appears in public chat. As a result, recipients can know that someone has gestured to them either by noticing the animation or the text emote. In these systems, gesturing works very differently than in EQOA. For example, one encounter in SWG (Figure 12, Excerpt 10), Atac is chatting privately with Sin'thea, who is somewhere in the player city, while she sits face-to-face with Zeroks. The player has maneuvered the camera view so that he can see Zeroks better while up against a wall (frame 1), but so that it


*Figure 12.* Not seeing another avatar's approach.

**(10) Star Wars Galaxies: Not Looking (Atac's perspective)**

```
01   05:02:52  You reply to Sin’thea, 'the live market went
02             pretty well, huh?'.
03   05:03:14  Zeroks hums a tune to himself.
04   05:04:42  Sin'thea tells you: yep seems so - (love was in the air)
05   05:10:50  Sin'thea tells you: yeah I really enjoyed it again
06   05:16:44  Sin'thea tells you: should be a hit going forward
07   05:24:28  Zeroks: ((you have LFG above your head, FYI))
08   05:33:00  Atac: hehe
09   05:33:50  Atac: oh
10   05:49:28  You reply to Sin'thea, 'I hope so'.
11   05:50:04  Sin'thea tells you: yeah i haven't talkted to you since
12             then huh
13   05:59:24  You reply to Sin'thea, 'I haven't been around'.
14   06:05:56  ((Atac opens Community menu))
15   06:07:17  ((Atac selects Character tab))
16 →06:07:40  Sin'thea smiles at you.
17   06:08:44  ((Atac unchecks "Looking For Group"))
18   06:09:56  ((Atac closes Community menu))
19 →06:10:20  Sin'thea [, all bubbly]: there you are!
20   06:10:56  ((Atac moves cursor to text box))
21 →06:12:28  ((Atac changes her view to “looking straight ahead”))
22   06:13:40  ((Atac clicks on Sin'thea's avatar))
23   06:16:06  You wink suggestively at Sin'thea.
24   06:16:46  Sin'thea [, all bubbly]: didn't even see you sneak in!
25             ((Sin'thea's avatar waves automatically))
26   06:21:38  Atac: hey chicka
```

does not correspond with the avatar's orientation: there is almost a 90-degree difference.

Atac and Sin'thea have been chatting privately for sometime and begin to discuss a player-hosted event, "the live market" (lines 01–02), while Zeroks waits patiently (line 03). As they chat (lines 04–06), Zeroks indicates that she has her "LFG" tag on (line 07), and Atac acknowledges him (lines 08–09). Atac and Sin'thea continue to chat privately (lines 10–13) before Atac decides to turn off her "LFG" tag. She opens the Community menu and selects the Character tab (lines 14–15). Almost simultaneously, a text emote appears that says, "Sin'thea smiles at you" (line 16). This indicates that Sin'thea, with whom Atac was not previously co-present, is now in the room. However, Atac does not appear to notice this text emote immediately and her offset camera view (not the Community menu) prevent her from seeing Sin'thea's avatar. Atac closes the Community menu (line 18) about the same time that Sin'thea says, "there you are!" (line 19) in spatial chat instead of the private chat channel they had been using. Atac moves her cursor to the text box, presumably noticing the spatial chat and the gesture emote (line 20) and then quickly corrects her camera view (line 21; frame 2). When she does this, we can see that Sin'thea is standing right in front of her avatar. They then continue the conversation in face-to-face mode, using gestures and spatial chat (lines 22–26). Thus, in this case, Atac misses the approach of Sin'thea's avatar because she changed the camera view to one that no longer corresponded to the direction of the avatar's eye gaze. Sin'thea has no way of knowing that Atac cannot see her avatar. All she can see is that Atac's response (line 23) is noticeably delayed.

Now the unpredictability of players' views in games with virtual cameras in mitigated to a degree by the text emotes that accompany gestures. Players may miss a gesture animation of an avatar but still notice the text emote. For example, on a different day, in the same player city, Atac notices Sin'thea's gestures without actually seeing them (Figure 13, Excerpt 11).

Atac is walking around the side of the town's player association hall (line 01; frame 1), when two text emotes appear, "Sin'thea waves to Teli Tubbi" and "Sin'thea nods at Kimon Calari" (lines 02–03). These descriptions of two gestures alert Atac to the presence of her friend Sin'thea even though she cannot see Sin'thea's avatar or the gesture animations themselves. Atac continues walking toward the entrance of the hall and she catches sight of her friend's avatar (line 04). Although she is most likely outside Sin'thea's field of view, she nonetheless initiates a "wave" since her friend will be able to see the text emote (line 07; frame 3). She continues up the steps of the hall, approaching Sin'thea's avatar from behind (line 09). Sin'thea finally returns the wave (line 10; frame 4) and the two continue their greetings (lines 14–18). At what point Sin'thea could actually see Atac's avatar, we cannot tell, but it is clear that each of them first noticed the other through the text descriptions


*Figure 13*. Noticing a “wave” without seeing it.

**(11) Star Wars Galaxies: Unseen Wave (Atac’s perspective)**

00:00 ((Atac walks toward the entrance; frame 1))
00:00 *Sin'thea waves to Teli Tubbi.*
01:04 *Sin'thea nods at Kimon Calari.*
05:21 ((Atac mouses over Sin'thea's avatar))
07:02 ((Atac selects sin'thea's avatar))
07:14 Sin'thea: is that Teli I see?
→10:12 *You wave to her.* ((frame 2))
16:20 *Kimon Calari nods at Sin'thea.*
18:00 ((Atac approaches Sin'thea from behind; frame 3))
→19:30 *Sin'thea waves to you.* ((frame 4))
((Sin'thea's avatar automatically spins around
and waves))
21:06 *Goldtooth greets you.*
22:24 Sin'thea: ATAC!!!
22:34 Atac: hiya
((Atac's avatar waves automatically))
23:22 *Sin'thea gives you a hug.*
((Sin'thea's avatar automatically gives half a hug))

of their gestures rather than through the avatars and their gestures. Thus, text emotes for gestures could be seen as a solution to the problem of coordinating the production of avatar animations with the unpredictable eye gaze of the other player; however, they appear to do so at the price of rendering the avatars irrelevant. Furthermore, of "virtual camera" systems, such as the one in SWG, Hindmarsh et al. (1998, p. 133) predict: "In some ways, then, the most critical benefits of using an avatar may be undermined; that is, if the position of the avatar completely belies the nature of the other's view of the world, the problem of establishing (for all practical purposes) what another is seeing would be exacerbated" (Hindmarsh et al., 1998, p. 133). Our data from SWG further support this claim.

Thus, we can see that the gaze system (or "virtual camera" system) creates complications for the gesture system in SWG (and most other virtual worlds), in terms of seeing the avatar animations. Despite the impressive sophistication of the avatars and the extensive library of animated gestures in SWG, players often rely more on the textual emotes instead. The result is a player experience that is more like a text-based MUD, than an avatar-based MMORPG. Rather than simulating face-to-face, what such a gesture system simulates is a world in which every time someone performs a visual gesture they at the same time announce it verbally (e.g., "I wave to you").

## 5. Discussion: Design considerations

In the preceding analyses we showed how various kinds of private player activities can hinder players' ability to recognize each others' actions and to achieve tight coordination. On the basis of these analyses, we offer a set of considerations and guidelines for designing social interaction systems that better provide for the "accountability" (Garfinkel, 1967) of player activities. In general, *entirely private player activities should be avoided.* Players can better coordinate their actions when they can see what the other is doing.

In terms of text chat systems (if they continue to be used as the primary form of "talk" instead of voice), other game companies should probably follow *There's* lead and implement word-by-word (or even character-by-character) posting of chat messages. Such systems produce a turn-taking system that is more like that in face-to-face, and they better facilitate the coordination of turns-at-chat with each other and with other joint game activities. Because avatars move in real time and players act in real time, they need a "voice" that responds in real time (or nearer to real time). Otherwise their utterances in the world run the risk of arriving late.

In terms of embodied activity, game developers are aware that avatars that stand motionless appear lifeless. The typical strategy for remedying this appears to be to create animations for autonomic behaviors such as breathing, blinking, glancing around, and shifting posture. Such animations are known as "idles." While this technique certainly helps make avatars appear more alive, it is only partially successful. Avatars still act like McGrane's deviant sociology students: alive but zoned out. Developers can make avatars come to life even more simply by animating more of the activities that players are already performing, such as rummaging through bags, consulting maps, or chatting privately.

More generally, we want to emphasize that in virtual worlds the presentation of information becomes more than a user interface problem: it becomes a *social interaction problem* as well. Whenever users interact with the interface, they may also at the same time be engaged in an interaction with another player. Therefore, their interaction with the system should be made observable publicly. Again *There* is ahead of other systems because it systematically implements an avatar animation for when players have opened a browser window (i.e., it wears a visor) and an IM window (i.e., it wears a headset). While this is a good start, it does not go far enough. Providing more fine-grained cues about what kind of activity the player is engaged is more powerful, as our addon to WoW demonstrates. For example, the party members in Excerpt 8 could not only see that Yuk had opened a window, but that he had opened his "quest log" and therefore was about to share quests with Noob. A better system would provide a different avatar animation for each of UI activity and could perhaps provide more fine-grained information.

Increasing the level of transparency regarding player activity raises some interesting issues for players and for designers. For example, in testing our *heightened awareness* addon to WoW, we discovered that while revealing more player activities facilitated coordination, it also revealed new kinds of clues about player competence. One player worried about "looking stupid" in front of his group mates because they could see that he was repeatedly consulting his map. On the one hand, this player might have liked to hide his actions in this instance. Indeed players have become accustomed to the odd sense of *public privacy* that virtual worlds enable. On the other hand, the fact that he was unfamiliar with the area could be useful for his group mates to know in prompting them to offer help or forgive wrong turns. How much control should players have in terms of keeping their actions private? Ultimately designers must consider the trade-offs between privacy and transparency and decide what degree of each is most appropriate given the kind of gaming experience they wish to create.

In addition, avatar gesture systems would be improved by enabling *user-controlled gesture duration*. For example, holding a "bow" would mean that the avatar remains in the most bent position until the user releases it. Or holding a "wave" would mean that the avatar continues to wave its hand back and forth. The easiest way to implement this feature would be that the user holds down the ENTER key upon typing a gesture command in order to "hold" the gesture until the key is released. When users can control the duration of a gesture, they can "hold" the gesture until they can see that the recipient has seen it and understood it, as in real life. User-controlled gesture duration would also enable players to be more expressive with a limited set of pre-defined gestures they already have. Despite the increasing sophistication of avatar animations, achieved through motion capture techniques, the fixed duration of such gestures nonetheless still gives them a mechanical feel.

Finally the management of players' screen views is notoriously difficult in virtual worlds. Problems include the fact that screens are not wide enough or curved enough to easily support peripheral vision (Hindmarsh et al., 2001); in third-person view, a player's avatar often occludes his or her view of others' avatars; chat bubbles over avatars' heads can occlude the view of other characters' chat bubbles; enclosed spaces (such as houses) can interfere with third-person views, etc.

To manage the visual display of avatars and chat bubbles on the screen, different systems use different solutions. Most systems (e.g., SWG, EQII and SL) use a user-controlled virtual camera. This solution is good in that it enables users to fix any problems in their view themselves. An alternative solution is the avatar-rearrangement approach in *There*. In *There* players can join a "conversation" group and when they do so, the system rearranges each avatar in the group into a semi-circular formation. As a result, avatars and chat bubbles do not overlap on the user's screen. However, the disadvantage of this approach is that it totally prohibits the natural arrangements and movement of avatars during conversation.

A better solution may simply be to enable virtual cameras, but with the addition of an indicator that makes the *player's view predictable from the avatar*. When the player's view becomes too offset from the avatar's orientation, this state is represented in the avatar (e.g., dark glasses or a blindfold appear on the avatar's face, avatar closes its eyes, the avatar looks up and around, etc.). In other words, an "I'm not looking" mode with a public avatar animation of some kind.

## 6. Conclusion

As "blockies" have been growing up over the past 15 years, we have seen the greatest advances in terms of the scale of virtual worlds, or the number of simultaneous players (i.e., from dozens to thousands), and in terms of the visual sophistication of avatars achieved through vastly more detailed anatomical models, motion capture of human bodies and the simulation of lighting and physics. Where virtual worlds have made somewhat less progress is in the area of social interaction systems. Despite the early work on CVEs and IRC and advancements in a minority of systems (i.e., *There*), game companies are slow to learn from them, and cumbersome interaction systems continue to be implemented. In general, slippages tend to occur when features of a player's activities (even seemingly insignificant ones) are hidden from public view. What enables fluidity in face-to-face interaction is the observability of all manner of activities, which in turn provides for the *accountability* (Garfinkel, 1967) of social action.

In this study, we have used qualitative interaction analysis to try to demonstrate, in a fine-grained way, the impact of certain features of avatar and chat systems on the players' experience of virtual social interaction and more specifically on their ability to tightly coordinate their activities. Concealing what players are typing, what menus they are accessing, and where they are looking from fellow players leads to delays and slippages in interaction. Revealing such features publicly creates a system that provides for better coordination between players. In general we caution designers in creating any kind of player activities that are entirely private. When designing any player activity, game developers should ask, "how might the player at that moment need to coordinate the activity with those of others?"

Beyond the system features discussed above, there is much room for further refinement of social interaction systems in virtual worlds. For example, despite the current predominance of text-based conversation in massively multiplayer virtual worlds, voice promises to be implemented more and more as a mode of conversation in the future. Already voice is the predominant mode of conversation in the first-person shooter genre of video games in which quick communication is vital for game play. It is also currently used by players, via third party VoIP applications (e.g. TeamSpeak or Ventrilo), for "raid" and "player-vs.-player" (PvP) activity in MMORPGs such as SWG and WoW.

Another major area of future development is to enable *free gesticulation* instead of a pre-specified library of avatar gestures. Just as the free composition of chat messages enables players to say any textual utterance they like, free gesticulation would enable players to perform any gesture they might perform with a physical body. Avatars can only become fully functional vehicles of expression when players can do this. The challenge with this of course is how the player will control such an avatar. Perhaps the most promising current approach is that of "real-time motion capture" (Richard Marks, Sony Computer Entertainment America) in which a camera and other sensors are used to track a player's body motions in real time and transfer them to their avatar. Players could then use their own bodies and faces as joysticks in puppeteering their avatars.

Thus the industry and the player base is already realizing that most current massively multiplayer worlds place too much burden on the players' hands as they must be used to "talk," walk, gesture, interact with objects and access menus. As Hindmarsh et al. (2001, p. 137) predict, when input to the system is too complicated, users will have to concentrate on "working their avatar" instead of collaborating with others, and thus become alienated from the very activities the environment was intended to support. This is a fair characterization of today's MMORPGs as well. Although players can become very skilled at chatting and "working their avatars" quickly, it is nonetheless cumbersome and detracts from the overall immersive experience. Therefore, systems in which players can use their voice to talk, their bodies to gesture, and their faces to emote would dramatically reduce the current workload on the hands and increase the ease and fluidity of expression. As these and other developments occur, it will continue to be critical to use a deep understanding of the mechanics of face-to-face interaction in order to guide design.

Although the study focuses on ''games,'' our primary interest is in avatar-mediated interaction and virtual worlds more generally. It just happens that games constitute the most popular form of virtual worlds at the moment. However, the virtual worlds in video games are already inspiring other more ''serious'' applications. For example, Forterra Systems Inc. has recently adapted the avatar engine of *There* to create, *Asymmetric Warfare Environment*, a simulation for training U.S. soldiers in how to communicate with people of the local cultures they encounter. In addition, Education Arcade has adapted the MMORPG, *NeverWinter Nights*, to the study of American history with their game, *Revolution*. And it is only a matter of time before online retailers such as Amazon enable customers to shop together in a virtual 3D version of their website. We believe that as the applications of virtual worlds diversify, basic questions of how to simulate the dynamics of face-to-face conversation effectively, especially the tight coordination of people's actions, will become increasingly important.

## Notes

1. An exception to this trend is Blizzard's World of Warcraft which uses somewhat blocky graphics no doubt in order to improve online game performance.
2. For example, Forterra Systems Inc. is using the avatar-interaction engine of *There* to create *Asymmetric Warfare Environment* for training U.S. soldiers in how to communicate more effectively with people of the local culture. Also, the University of California at Davis is using SL to create disaster-response simulations for training purposes. In addition, Education Arcade is adapting the *NeverWinter Nights* game engine to teach American history in their game, *Revolution*.
3. Our annotated logs use the following notations, some of which are adapted from conversation analysis (Sacks et al., 1974): line numbers appear in the leftmost column; time codes for messages appear in the next column to the right; descriptions of player and avatar actions appear in double parentheses; and system-generated messages appear in italics. In Excerpt 2, we also added timed pauses and gaps *in addition to the time codes* in order to highlight their length and make it easier for the reader to compare them with those in the real-time and near-real-time conversations in Excerpts 1 and 3.
4. ''Role players'' constitute a minority group of players who take their character's persona seriously by saying and doing only things the character might do in the fictitious world.
5. Although Atac is in fact one of the researchers, we did not invent this practice. We observed it initially on a handful of other occasions among the role players we were studying, but did not capture any of these on videotape. We did, however, incorporate the practice in our own behavior around role players and captured a few of these cases.

## References

Bartle, Richard. A. (2003): *Designing Virtual Worlds*. New Riders: Indianapolis.

Benford, S., J. Bowers, L.E. Fahlén and C. Greenhalgh (1994): Managing Mutual Awareness in Collaborative Virtual Environments. In G. Singh, S.K. Feiner and D. Thalmann (eds): *Proceedings of Virtual Reality Software and Technology*., Singapore: World Scientific Publishing Co. Pte. Ltd.

Benford, S.D., J.M. Bowers, L.E. Fahlen, C.M. Greenhalgh and D.N. Snowdon (1997): Embodiments, Avatars, Clones and Agents for Multi-user, Multi-sensory Virtual Worlds. *Multimedia Systems*, vol. 5, no. 2, pp. 93–104 Springer-Verlag.

Bly, Sara, Steve R. Harrison and Susan Irwin (1993): Media Spaces: Bringing People Together in a Video, Audio, and Computing Environment. *Communications of the ACM*, vol. 36, no. 1, pp. 28–47 January.

Bowers, John, James Pycock and Jon O'Brien (1996): Talk and Embodiment in Collaborative Virtual Environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Common Ground*. New York: ACM Press, pp. 58–65.

Brown, Barry and Marek Bell (2004): CSCW at Play: 'There' as a Collaborative Virtual Environment. *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work*. NY, USA: ACM Press, pp. 350–359.

Castronova, Edward (2001): Virtual Worlds: A First-Hand Account of Market and Society on the Cyberian Frontier, (December 2001). *CESifo Working Paper Series No. 618*. Available at SSRN: http://ssrn.com/abstract = 294828.

Cherny, Lynn (1999): *Conversation and Community: Chat in a Virtual World*. Palo Alto, CA: CSLI Publications.

Curtis, Pavel (1997): Mudding: Social Phenomena in Text-Based Virtual Realities. In S. Kiesler (ed.): *Culture of the Internet*. Hillsdale, NJ: Erlbaum, pp. 121–142.

Dourish, Paul and Victoria Bellotti (1992): Awareness and Coordination in Shared Workspaces. In *Proceedings of The ACM conference on Computer-Supported Cooperative Work, Toronto, Ontario, Canada*. NY, USA: ACM Press, pp. 107–114.

Garcia, Angela and J. Jacobs (1999): The Eyes of the Beholder: Understanding the Turn-taking System in Quasi-synchronous CMC. *Research on Language and Social Interaction*, vol. 32, no. 4, pp. 337–367.

Garfinkel, Harold (1967): *Studies in Ethnomethodology*. Englewood Cliffs N.J: Prentice-Hall.

Garfinkel, Harold (1988): Evidence for Locally Produced, Naturally Accountable Phenomena of Order, Logic, Reason, Meaning, Method, etc. in and as of the Essential Quiddity of Immortal Ordinary Society (I of IV): An Announcement of Studies. *Sociological Theory*, vol. 6, pp. 10–39.

Grinter Rebecca E. and Leysia Palen (2002): Instant Messaging in Teen Life. *Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work, CSCW 2002, New Orleans, LA*. New York, pp. 21–30.

Gibson, William (1984): *Neuromancer (Remembering Tomorrow)*. New York: Ace Books.

Goffman, Erving (1959): *The Presentation of Self in Everyday Life*. New York: Doubleday.

Goodwin, Charles (1979): The Interactive Construction of a Sentence in Natural Conversation. In George Psathas (ed.): *Everyday Language: Studies in Ethnomethodology*. New York: Irvington Publisher, pp. 97–121.

Goodwin, Charles (1986): Gesture as a Resource for the Organization of Mutual Orientation. *Semiotica*, vol. 62, pp. 29–49.

Heath, Christian and Paul Luff (1992): Collaboration and Control: Crisis Management and Multimedia Technology in London Underground Line Control Rooms. *Computer Supported Cooperative Work*, vol. 1, pp. 69–94.

Heritage, John (1984): *Garfinkel and Ethnomethodology*. Cambridge: Polity Press.

Herring, S. (1999): Interactional Coherence in CMC. *Journal of Computer-Mediated Communication [On-Line]*, vol. 4, no. 4. http://www.ascusc.org/jcmc/vol4/issue4/herring.html (30/1/03).

Hindmarsh, Jon, Mike Fraser, Christian Heath, Steve Benford and Chris Greenhalgh (1998): Fragmented Interaction: Establishing Mutual Orientation in Virtual Environments. In *Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work*. New York: ACM Press, pp. 217–226.

Hindmarsh, Jon, Mike Fraser, Christian Heath and Steve Benford (2001): Virtually Missing the Point: Configuring CVEs for Object-Focused Interaction. In *Collaborative Virtual Environments*, Springer, September 2001, pp. 115–133.

Hindmarsh, Jon and Christian Heath (1999): Embodied Reference: A Study of Deixis in Workplace Interaction. *Journal of Pragmatics*, vol. 32, pp. 1855–1878.

Hine, Christine (2000): *Virtual Ethnography*. London: Sage.

Jefferson, Gail (1989): Preliminary Notes on a Possible Metric which Provides for a 'Standard Maximum' Silence of Approximately One Second in Conversation. In D. Roger and P. Bull (eds): *Conversation: An Interdisciplinary Perspective*. Clevedon: Multilingual Matters, pp. 166–196.

Kendon, Adam (1980): Gesticulation and Speech: Two Aspects of the Process of Utterance. In M.R. Key (ed.): *The Relation between Verbal and Nonverbal Communication*. Mouton, pp. 207–227.

Lyman, P. and N. Wakeford (1999): Going into the (Virtual) Field. *American Behavioral Scientist*, vol. 43, no. 3, pp. 359–376.

Lynch, Michael (1993): *Scientific Practice and Ordinary Action: Ethnomethodology and Social Studies of Science*. New York: Cambridge University Press.

Mason, Bruce (1999): Issues in Virtual Ethnography. In K. Buckner (ed.): *Ethnographic Studies in Real and Virtual Environments: Inhabited Information Spaces and Connected Communities. Proceedings of Esprit i3 Workshop, Queen Margaret College, Edinburgh, January 1999*. pp. 61–69.

McGrane, Bernard (1994): *The Un-TV and the 10 MPH Car: Experiments in Personal Freedom and Everyday Life*. Fort Bragg CA: The Small Press.

McNeill, David (1992): *Hand and Mind: What Gestures Reveal About Thought*. Chicago: Chicago University Press.

Miller, D and D. Slater (2000): *The Internet: An Ethnographic Approach*. Oxford and New York: Berg Publishers.

Nardi, Bonnie A., Steve Whittaker, and Erin Bradner (2000): Interaction and Outeraction: Instant Messaging in Action. In *Proceedings of CSCW 2000, Philadelphia, PA. New York: ACM*, pp. 1–10.

O'Day, V., D.G. Bobrow and M. Shirley (1998): Network Community Design: A Social-Technical Design Circle. *The Journal of Collaborative Computing*, vol. 7, no. 3–4, pp. 315–337.

Robinson, Mike, Samuli Pekkola, Jonni Korhonen, Saku Hujala, Tero Toivonen and Markku-Juhani O. Saarinen (2001): Extending the Limits of Collaborative Virtual Environments. In Churchill, Snowdon and Munro (eds.): *Collaborative Virtual Environments*. Springer-Verlag, pp. 21–42.

Ruhleder, Karen and Brigitte Jordan (2001): Co-Constructing Non-Mutual Realities: Delay-Generated Trouble in Distributed Interaction. *Journal of Computer Supported Cooperative Work*, vol. 10, no. 1, pp. 113–138.

Rutter, Jason and G.W.H. Smith (2005): Ethnographic Presence in a Nebulous Setting. In Christine Hine (ed.): *Virtual Methods: Issues in Social Research on the Internet*. Berg, pp. 81–92.

Sacks, Harvey, Emanuel A. Schegloff and Gail Jefferson (1974): A Simplest Systematics for the Organization of Turn-taking for Conversation. *Language*, vol. 50, pp. 696–735.

Schegloff, Emanual A. (1968): Sequencing in Conversational Openings. *American Anthropologist*, vol. 70, pp. 1075–1095.

Schegloff, Emanual A. (1987): Analyzing Single Episodes of Interaction: An Exercise in Conversation Analysis. *Social Psychology Quarterly*, vol. 50, pp. 101–114.

Schegloff, Emanual A. (1992): Repair After Next Turn: The Last Structurally Provided Defense of Intersubjectivity in Conversation. *American Journal of Sociology*, vol. 98, pp. 1295–1345.

Schegloff, Emanual A. (1998): Body Torque. *Social Research*, vol. 65, pp. 535–586.

Schegloff, Emanual A. and Harvey Sacks (1973): Opening Up Closings. *Semiotica*, vol. 8, no. 4, pp. 289–327.

Schmidt, Kjeld (2002): The Problem with 'Awareness'. *Computer Supported Cooperative Work*, vol. 11, pp. 285–298.

Stephenson, Neal (1992): *Snow Crash*. New York: Spectra.

Streeck, Jurgen (1993): Gesture as Communication I: Its Coordination with Gaze and Speech. *Communication Monographs*, vol. 60, no. 4, pp. 275–299.

Streeck, Jurgen (1995): On Projection. In E. Goody (ed.): *Interaction and Social Intelligence*. Cambridge: Cambridge University Press, pp. 84–110.

Szymanski, Margaret H., Paul Aoki, Erik Vinkhuyzen and Alison Woodruff (forthcoming): Organizing a Remote State of Incipient Talk: Push-to-talk Mobile Radio Interaction. *Language in Society*, vol. 35, no. 3.

Taylor, T.L. (2003): Power Gamers Just Want to Have Fun?: Instrumental Play in a MMOG. In *Proceedings of Level Up*. University of Utrecht, The Netherlands.

Vallis, Rhyll (1999): 'Members' Methods for Entering and Leaving #IRCbar: A Conversation Analytic Study of Internet Relay Chat. In K. Chalmers, S. Bogitini and P. Renshaw (eds): *Educational Research in New Times*. Flaxton: PostPressed, pp. 117–127.

Whalen, Jack, Don H. Zimmerman and Marilyn R. Whalen (1988): When Words Fail: A Single Case Analysis. *Social Problems*, vol. 35, pp. 333–362.

Woodcock, Bruce (2005): *An Analysis of MMOG Subscription Growth (version 7.0)*, http://www.mmogchart.com/.

Yee, Nick (2005): *The Daedalus Project: The Psychology of MMORPGs*, http://www.nickyee.com/daedalus/.

Zimmerman, Don H. (1988): On Conversation: The Conversation Analytic Perspective. In J. Anderson (ed.): *Communication Yearbook 11*. Beverly Hills: Sage, pp. 406–432.