(Im)materiality, virtual reality and interaction: grounding the 'virtual' in studies of technology in action

Jon Hindmarsh, Christian Heath and Mike Fraser

Abstract

This paper explores the organisation of social interaction amongst participants 'in' Virtual Reality. Despite the wide-ranging sociological interest in 'virtual' technologies, there is rather little detailed sociological investigation of user experiences of the virtual technology par excellence, namely multi-user Virtual Reality. Interestingly the discourses that underpin discussions of more mundane virtual technologies (eg email, the Web, mobile phones, etc.) tend to draw on design visions for Virtual Reality, such as the opportunities for social life freed from the constraints of the physical body. This paper contributes to a growing number of empirical studies that provide a critique of this view, but maybe more importantly, provides a detailed analysis of action and interaction in virtual worlds. It considers the organisation of interaction within VR with particular emphasis on the ways in which visual features of the digital domain are seen and shared by participants. The paper describes the ways in which the abilities to share views on the virtual world requires participants to overcome problems associated with the very material character of the VR interfaces. The study is based on the analysis of recordings of a Virtual Reality system that enables participants to talk to one another and see one another's actions within a virtual environment.

Introduction

> Today we call many things 'virtual.' . . . All sorts of hybrid social realities have sprung up on fax machines and computers, cellular telephones and communication satellites. Yet most of these 'virtual realities' are not, in the strict sense of the term, virtual reality. They are pale ghosts of virtual reality, invoking 'virtual' to mean anything based on computers. (Michael Heim, 1998: 47)

Virtual society, virtual communities, virtual selves, virtual bodies, virtual realities. In 1985, the sociology of technology was exploring how the refrigerator

The Sociological Review, 54:4 (2006)

got its hum (Mackenzie and Wacjman, 1985); today the concern is with all things 'virtual'. Indeed, Hughes *et al*. (2001) argue that 'the notion of the "virtual" bids fair to become one of the most over-used concepts of the decade as its use spreads as quickly as the growth of the Internet' (p. 49). Interestingly, however, whereas the phenomena that sociologists of the 'virtual' examine range from mobile team working to internet dating, the discourses underpinning the debates about cyberspace primarily derive from the design visions for Virtual Reality. Maybe unsurprisingly therefore there is a growing recognition that these discourses do not match the practical realities of the everyday instantiations of cyberspace. The problems seem particularly acute when considering sociological theories of the virtual. As Tim Jordan argues, '[a] significant number of theories concerning the nature of cyberspace or virtuality are being constructed with little regard for the empirical realities of online life.' (Jordan, 2001: 1).

Indeed there is a scarcity of studies that have looked at real-time use of the virtual technology par excellence, namely Virtual Reality (VR). Therefore this paper takes a central concern with the ways in which people interact 'in', or 'through', VR. In particular it is concerned with multi-user Virtual Reality, which provides opportunities for participants to communicate with one another across distances and to work or interact within digital data spaces. Indeed, one of the key advantages of VR over the telephone for example is that it has the potential to provide enhanced support for work and interaction between remote participants by delivering a common world of digital data and virtual objects for co-participants to discuss. For example, future virtual worlds may support architects and clients discussing an emerging design; or medical experts discussing and planning surgical techniques. Thus a key feature of VR is the digital domain that is inhabited. Therefore this paper takes a particular interest in the ways in which participants see and share features of the virtual world in interaction.

The paper draws on a methodological tradition from ethnomethodology and conversation analysis and focuses on the detailed analysis of a number of illustrative instances from recordings of VR in use. In doing so, it demonstrates how even the most 'virtual' of our era of virtual technologies is ultimately and intractably founded on the material constraints of the physical body in action. Indeed it turns out that the abilities of participants to share these immaterial worlds rests in part on their communicative skills in overcoming the physical constraints of the interfaces.

Virtual reality and 'collaborative virtual environments'

At its outset Virtual Reality as a technology (or suite of technologies) was heralded by its self-proclaimed 'prophets' as giving us opportunities to travel to new digital worlds. One of those key to the development of VR, Ivan Sutherland talked about an Ultimate Display in the following terms:

> The screen is a window through which one sees a virtual world. The challenge is to make that world look real, act real, sound real, feel real. (Sutherland 1965)

Similarly, the highly quotable Jaron Lanier announced:

> Virtual Reality creates a new objective level of reality . . . If you're ever confused about which reality you're in, you put your hands on your eyes and see if you're wearing Eye-phones or not. (Lanier, 2000)

In promoting VR, Lanier, Sutherland and others appealed to certain design visions, to hopes for future *potential* instantiations and experiences of using VR technologies. These compelling statements that heralded the emergence of VR have greatly influenced the development of interactive graphics technologies. Research has been driven by a commitment to build environments and interfaces through a complex array of technologies which aim to provide the illusion of, simulate, or are even indistinguishable from the physical world. Although such rhetoric has more recently been downplayed amongst the VR community, there persists an overriding concern with the factors that influence the sense of 'presence' in virtual worlds.[1] Although computing technologies are developing at incredible rates, their abilities to live up to these visions are very much in debate.

This rhetoric seems to have been adopted almost uncritically within many sociological accounts of 'virtual' technologies. If we focus on accounts of VR for example, sociologists have argued that a major appeal is the ability to 'park' the flawed human body, to transcend the limits of the flesh, to create and explore new identities, to be free the body from social and physical constraint and so on and so forth. Michael Heim, for example suggests that as all technologies extend our sense and our physical reach, VR does this to the maximum, as it transports our nervous system in to an electronic domain (Heim, 1995). Thus, everyday 'embodied' experiences are unable to compete with 'hyperreal', 'disembodied' experiences in cyberspace. More generally, others have suggested that we are freed from our physical bodies in cyberspace (McCormick and Leonard, 1996) or how Virtual Reality technologies can offer the prospect of disembodied experiences (Dale and Burrell, 2000). Social scientists have also considered the potential social and cultural impact of possible virtual futures – potential transformations of communities, urban spaces, geographies, organisations, the self and the body (eg Barnatt, 1995; Hillis, 1999). These studies juxtapose the physical and the digital; the material and the immaterial; the embodied and the disembodied.

However there is a serious danger of extrapolating the social consequences for future societies from the visions and aims of technologists or indeed the rhetoric of science fiction writers. As Nigel Thrift powerfully argues: 'The literature on electronic telecommunications technologies has been infected by the virus of new era thinking, a virus which is simply another variant of

technological determinism' (Thrift, 1996: 1463). There is a strong and uncritical aspect to many studies of the 'virtual', an aspect that engages in futurology or new era thinking.

So there is a clear need to consider the actual experiences of and participation in virtual environments. Indeed even typical descriptions of VR need to be clarified in relation to actual experience. For instance, Featherstone and Burrows describe Virtual Reality as:

> . . . a computer-generated visual, audible and tactile multi-media experience. Using stereo headphones, head-mounted stereo television goggles able to simulate three dimensions, wired gauntlets and computerized clothing, VR aims to surround the human body with an artificial sensorium of sight, sound and touch. VR systems are also truly interactive in the sense that produces the simulated environment in which a person is immersed, constantly reconfigures that environment in response to body movements. As yet, the technology is relatively crude. There is sometimes a lag between movements of the body and the reconfiguration of the environment, graphics resolution is low and many environments rely upon line drawings and/or cartoon-like iconic representations. Nevertheless all the indications are that the level of realism attainable will improve dramatically towards the end of the century. (Featherstone and Burrows, 1995: 6)

Today, beyond the end of the century to which Featherstone and Burrows allude, VR is not much further developed in terms of 'realism' (although the *graphical* realism has markedly improved). Again the description owes much to the technological vision heralded by Lanier and Sutherland rather than to accounts of existing VR systems. For example the array of technologies that may be included under the heading 'Virtual Reality' is much wider (and increasingly so) than the standard notion of 'immersive' VR that is emphasised in the above description. One aspect of VR in use is that the digital environments can be accessed and navigated through many different technological interfaces; interfaces that communicate variously via 2-D, 2.5-D and 3-D graphics, text, audio, touch, and so forth. So VR systems can be configured differently to allow individuals to look at the digital world through head-mounted eyeglasses, on desktop screens, in room-sized set-ups (called 'Computer Automated Virtual Environments' or CAVEs); on large projection displays or even on PDAs. Virtual Reality environments can be designed for individuals to explore digital landscapes (datascapes) or can be built to support multiple users communicating 'in', or maybe better, 'through', the virtual environment. Multi-user systems are called Collaborative Virtual Environments (CVEs),[2] which enable users to communicate with others using real-time text or simply by speaking to one another through an audio channel (microphone and headset or speakers); they can receive force-feedback through data gloves or vibrating mice; and so forth (see Heim, 1998 or Schroeder, 1996 for comprehensive descriptions of the various types of Virtual

Reality technology). Indeed, it is possible for different individuals to see the virtual world and communicate through that virtual world whilst accessing it through different types of interface – ranging from PDAs to fully immersive HMDs.

It is not just the issue of definition at stake however. Empirical sociological studies of VR in use are limited. Those that do undertake studies of VR have often focused on broader issues of 'community', 'identity' or 'presence'. There has been a particularly fruitful set of studies concerned with the social shaping of VR technologies (see Schroeder, 1996, 1997; Green, 1999), however the real-time organisation of interaction in VR is often overlooked or remains epiphenomenal. For example, there is a large edited collection entitled *Communication in the age of virtual reality* (Biocca and Levy, 1995), which makes many valuable contributions to our understanding of VR as a communication technology, but does not include a single empirical study of social interaction in VR.

The relative paucity of empirical sociological studies of social interaction in VR may derive in part from the lack of VR technologies 'in the wild'. Those that are available are mainly game-based systems that either feature on-line, in video arcades or maybe in science museums. These tend to have relatively limited communicative resources for participants – they are either designed for individuals to play games in isolation or only support text chat rather than full audio communication. Thus the interaction is focused on text chat and the like (for recent examples see Schroeder, 1997; Ducheneaut and Moore, 2004; Brown and Bell, 2004). There are however a small number of studies that have been able to analyse the use of systems that support full visual and live audio – the 'full' version of VR. For example, Bowers *et al*. have undertaken interactional analyses of meetings involving real-time interaction between CVE designers located at five sites and in three countries (UK, Sweden and Germany) (Bowers, 1996; Bowers, 1996). Bowers *et al*. only had access to quasi-naturalistic data drawn from meetings held by the research team themselves, and thus featuring the analysts as participants. Also, Büscher *et al*. examined a range of collaborative VR exhibits in the ZKM centre for art and media technology (in Germany), and were able to collect naturalistic materials of museum visitors encountering and experiencing, and communicating through, VR 'worlds' (Büscher *et al*., 2001).

In this paper, we extend these endeavours. Rather than explore the potential for VR and how it might form a distinctive site for sociality, we consider the ways in which individuals currently communicate in and through one version of VR. In particular, we are keen to explore the relationships between the 'virtual' setting for social encounters and interactions in relation to it; that is we are interested in the ways in which the virtual world features within sequences of interaction. Furthermore in order to pursue one aspect of the discussions of virtual environments, we aim to consider how the 'bodies' of participants – both their physical body and their virtual embodiment – 'mediate' interaction in virtual environments.

Research approach

For this study we had a relatively rare opportunity to collect data on social interaction in VR as part of a research project concerned with the development of one particular Collaborative Virtual Environment called MASSIVE-2[3] (see Hindmarsh *et al*., 2000 for the design output of the project). The research project formed a collaboration between social and computer scientists and focused around providing enhanced support for discussions of features of the virtual environment (virtual objects, scenes, etc.). Whilst studies have shown that much collaborative work rests upon the sharing and discussion of a whole host of documents, tools and other artefacts (eg Goodwin, 2000; Hindmarsh and Heath, 2000a), supporting such interaction is a general problem for the development of advanced shared workspaces (see for example Heath and Luff, 2000). Although for a range of video conferencing systems and media spaces[4] the problems associated with establishing what another can see or is looking at are well recognised, it is often argued that problems of recognising what views and scenes are available to the other are 'naturally' overcome in 3-D worlds (eg Smith *et al*., 1998). We wanted to explore the extent of this support in VR.

The sorts of technological interfaces most readily associated with Virtual Reality consist of head-mounted displays (HMDs) and the like, where the user is 'immersed' in a digital world. Sensors are taped to different parts of the user's body (most notably the head and hands) to calculate and reproduce physical movements such as turning and hand gestures on the user's embodiment, or avatar, in the virtual environment. However the act of 'walking' around the virtual world presents a technical problem. The user must remain relative stationary because they are wired up to the computing infrastructure and need to stay within the field of tracking sensors. Therefore users cannot physically walk around, but rather need an additional device, often a joystick, to control large scale movements such as 'walking'. Thus, users of this kind of immersive virtual environment often look something like the person depicted in Figure 1. Some of the early HMDs were so heavy and the position trackers so rudimentary that there was a heavy metal bar that attached the HMD to the ceiling – a reason why the first HMD developed by Ivan Sutherland at the University of Utah was known as 'The Sword of Damacles'. Now, as can be seen in Figure 1, the interface is much more lightweight, and is increasingly so.

Whilst Figure 1 presents a standard image of the interface to virtual worlds, the most commonly used interfaces do not adopt HMDs. Indeed VR is most commonly experienced using standard desktop computers (see Figure 2). The desktop interface allows people to view a world in pseudo-3D (so called 2.5-D). They can move their avatars around the world and undertake additional actions by operating the mouse.

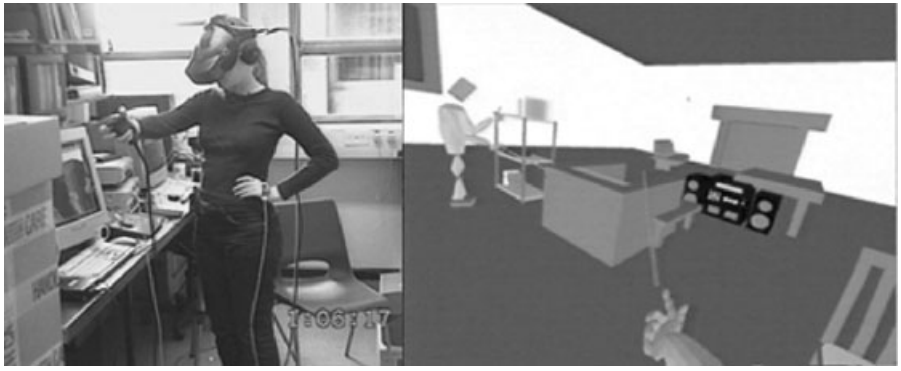


Figure 1 *A user wearing a Head-Mounted Display (HMD), position sensors and a joystick and their view on the virtual environment*

Figure 2 *The Desktop Interface and a Virtual View*

This paper will focus on examples of interaction between people using the desktop system, although interestingly similar issues emerged in both systems that we looked at.

In order to explore social interaction in the virtual world, we undertook a series of short trials with novice users. Thus, our research materials have been drawn from artificial or pseudo-experimental situations. Previously CVEs (that support *audio*, not simply text, for users to communicate) merely provided support for object-focused discussions through avatar position or orientation, talk and the ability to manipulate objects. However, we wanted to provide additional support by allowing avatars to gesture and so forth. Thus, we did not have access to any system that provides enhanced (or dedicated) support for object-focused interaction *and* that is currently used in a real-world context. So, we had to develop our own system and organised some trials with individuals interacting through that system. We enhanced MASSIVE-2 to

support simple pointing gestures – basically a virtual arm on a user's avatar would point towards selected items in areas of the virtual world (see Figure 2).

We set participants a very simple task, which involved two (or in some cases three) remote individuals collaboratively organising the layout of furniture in a virtual room. We did not design the task to assess 'performance' or 'efficiency', but simply as a means to encourage discussion and debate around visible features of the world. Asking them to complete this kind of task ensured that they would reference articles of furniture and spaces in the room. We were not interested in individual characteristics or behaviours, but rather the practices that participants used to establish mutual orientation to features of the world.

The analysis of the materials was informed by ethnomethodology and conversation analysis and in particular an emerging body of work concerned with the embodied organisation of social interaction (see for example, Goodwin, 2000; Heath and Luff, 2000; Hindmarsh and Heath, 2000a; and for a description of the approach, Heath and Hindmarsh 2002). Although studies drawing on conversation analysis are traditionally associated with the examination of naturalistic interaction, the prototypical nature of collaborative VR raises pragmatic barriers to the collection of naturally occurring materials, in that there were no systems featuring full audio communication in everyday use.

The data collection was somewhat more complex than in most studies of standard face-to-face interaction as there is no single scene of action for the analyst to focus on. Rather it was necessary to collect multiple audio-visual records: views of each user in the real world and view of each user's view on the digital world. As the trials featured two and occasionally three users, there were four to six video recordings being made simultaneously. In addition, and in order to ascertain the relative timing of different recordings, it was necessary to collect a further video recording that displayed all video and audio feeds simultaneously. This last recording also required audio mixing to produce an audio track that could meaningfully be associated with the collated display.

The complexity of the data collection was matched by the complexity of data analysis. Whilst identifying examples of reference to features of the virtual world was straightforward, the detailed description of these examples involved continual comparison between the different video records to establish the exact ordering of events from the viewpoint of different participants and the relationship between conduct in the physical and the digital domains. Our approach aimed to explore and uncover the kinds of interactional phenomena, practices and problems that may be of particular relevance to participants. Given the exploratory and fundamentally qualitative analytic approach our interest was to develop rich descriptions of specific data extracts. Moreover the examples that we present here are illustrative of a wider range of extracts that exhibit the underlying issues and phenomena discussed.

The material foundations to 'immaterial' worlds

It was noted earlier that the discourses of the virtual tend to revolve around notions that virtual domains provide an immaterial world distinct from the real; a world in which the physical body can be 'parked' and the virtual self can flourish. The materials collected in the course of our trials contradict this notion, but contradict it in interesting ways. The data that are presented in the following sections aim to illustrate the ways in which the materiality of using VR impacts upon the experience of, and interaction in, Collaborative Virtual Environments. Others have also discussed how activities and events in the physical environment can impact on the ways in which actions are perceived in the virtual world. For example Bowers *et al*. (1996a) discuss the phenomenon of '*corpsing*', which refers to those occasions on which users step away from their computer, but their avatar is still physically present. Thus, for other participants in the virtual world, it is unclear that the user is not available for interaction and only discovered through attempts at engaging interaction. In these sections we hope to reveal some further, more subtle, ways in which the design of the virtual world and the way in which it is 'entered' relates to the use and experience of the system, and especially how people interact 'in' or rather 'through' VR.

The bulk of this paper considers social interaction through desktop VR. Our studies of more immersive environments raise similar, if not heightened, difficulties for users. In particular the material realities of 'entering' the digital domain using immersive interfaces are inordinately constraining. The standard 'uniform' for entering immersive VR currently involves some combination of a head-mounted display, body sensors and a joystick or other input device. These devices are not comfortable to wear after a relatively short period of time. Even though they are getting lighter, the HMDs are a strain on the head and neck after extended 'immersion'. Furthermore in these kinds of dynamic and collaborative digital environments, there does tend to be some 'jitter' in the image and this can lead users to experience some motion sickness.[5]

The very physical character of the interface became an occasioned topic of conversation. Immersion in the virtual world rendered the participants blindfolded from their physical surroundings. The leads, cables and sensors that are required to capture and computationally process the users actions also tend to wrap around the user's legs and body (even though helpers were on hand to try to keep them as free as possible). To move around the virtual world, participants had to stretch both of their arms out. This produced the vision of Frankenstein both in the physical domain and in the virtual world because the avatar replicated the sense data captured on the movement of the participants' bodies. However this proved markedly counter-intuitive and users would recurrently attempt to step physically forward only to be halted by a stray cable or, more painfully, the sharp edge of a nearby desk. Whilst participants

avoided any serious accidents (and researchers were on hand to move potential obstacles out of the way), they would routinely say to their remote co-participant things like 'uh-oh, tangled up', 'I keep punching something – I can't tell what it is', 'ooh, I've just knocked something over'. Thus, the embodied, physical work of engaging in cyberspace became a topic of conversation amongst 'virtual' co-participants. In addition, even turning around to look in different parts of the virtual world led to tentative movements because of the fear or the feeling of wires around legs, cables on the floor or other nearby objects. They became very wary of moving too much or too fast. As a result movements in the virtual world were even slower than the computational speed demanded and reduced participants abilities to attend to the actions of others.

Whilst the desktop interface avoided many of these physical problems there were some difficulties that arose in sequences of interaction due to aspects of the physical interface that were common to both types of interface. These form the basis for the following sections.

Fragmenting interaction

One of the physical features of using desktop VR is that the digital environment is presented on a standard computer screen. The dimensions of this screen may be 15' or a little more. However, most 'desktop' virtual environments can only provide something like a 55 degree horizontal field of view on the digital world.[6] It is this narrow in order to minimise distortion on the graphical view of the world.

The implications of this narrow view on the world and on the action therein are critical, especially with regard to the ways in which participants are able to indicate and discuss features of the virtual scene. In everyday face-to-face interaction, when an individual asks a co-participant to look at an object at which they are pointing, that co-participant can usually see them in relation to their surroundings. They can simply look away from the body of the other to find the referenced object. Alternatively they may be able to see the other looking at an object out of the corner of their eye and be able to turn towards it. In CVEs, however the field of view restricts participants' opportunities to see avatars in relation to objects. When an utterance is produced, individuals are rarely in a position to see *at a glance* the avatar alongside the referenced object. It turns out that it is critical that they do see them in relation to one another. So, participants routinely turn to find the other's avatar before then looking for the object to which the avatar is visibly pointed. To illustrate, consider Fragment 1, in which Sarah asks Karen about the 'desk-thing' in the room. Before they can discuss where they might put the desk, they need some 25 seconds to achieve a common orientation towards it (square brackets indicate overlapping utterances; a dot in brackets indicates a short pause in talk).

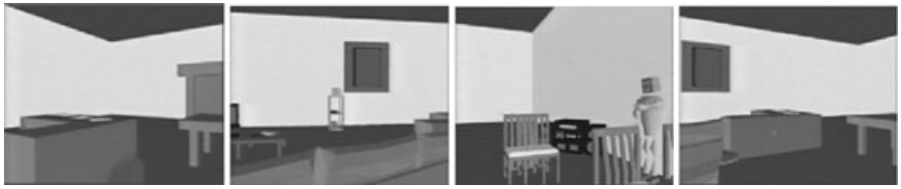


Figure 3 *Karen's view as she turns 360° to find Sarah's avatar and then the 'desk-thing'. The first view and the last both show the 'desk-thing'*

Fragment 1
S: You know this desk-thing?
K: Yeah?
S: Can you see what I'm pointing at now?
((K Turns to Find S))
K: Er I can't see you, but [I think
S: [Its like a desk-thing.
K: Er-where've you gone? [heh heh heh
S: [Erm, where are you?
K: I've. the. I can see
S: Turn (.) oh, oh yeah. you're near the lamp, yeah?
K: Yeah.
S: And then, yeah turn around right. (.) and then
its like (.) I'm pointing at it now,
but I don't know if you can see what [I'm pointing at?
K: [Right yeah I can see.

When Sarah asks if Karen can see what she is pointing at, Karen starts to look for Sarah's avatar and her pointing gesture. She is actually facing the desk very early on in the sequence, but ends up turning 360°, via Sarah's gesture, to return to the very same desk. See the sequence of images in Figure 3.[7]

In this CVE, due to the narrow field of view available, participants often do not have the other's avatar in view during an utterance and the sluggish shifts in orientation that are possible make it a relatively slow process to find the other's avatar. They might turn out to have initially had the referent in view, but without seeing the object in relation to the pointing gesture, they have little information with which to assess if they are looking at the appropriate object. In other words, they may see a 'desk-thing', but is it the relevant 'desk-thing'? In some cases, then, they cannot be certain that they are looking at and discussing the same object without seeing that object in relation to the other's avatar.

Thus participants often turn in the virtual world to be able to see the other avatar and then they use this as a resource to find the referent. Even in short

and straightforward instances, participants can be seen to turn their avatar from an object to find the other's avatar gesturing, only to subsequently return to face that object. Participants may, however, need to engage in an extensive search for their co-participant's avatar before being able to find the relevant object, especially because there are few clues to where the other's avatar might be. The sense of another's physical presence that we might rely on in everyday contexts (recognisable through sounds and the like) is abjectly unavailable in this 'immaterial' world.

In everyday environments, the local constellation of objects and artefacts provides a resource for making sense of the actions of others. The production of action is situated in an environment and its intelligibility rests upon the availability of that environment (Heath and Hindmarsh, 2000). However, participants in CVEs only encounter a fragment of the visible world. Fragmenting the visibility of an avatar from the objects or scenes to which they refer creates difficulties for participants, as their overall sense of the action is impoverished. They are rarely in a position to see both object and avatar simultaneously and have problems in relating the two. Critically, the sense of talk or action is based upon the mutual availability of that relationship. In co-present interaction, a question about an object is routinely conflated with the resources to find that object in the local milieu (see Hindmarsh and Heath, 2000b). In contrast to co-present interaction, therefore, referencing objects and spaces becomes *a topic in and of itself* within CVEs.

Even during the course of these relatively brief technology trials of half an hour or so, participants became familiar with some of the problems that these environments pose for interaction and had ways of managing certain difficulties. For example, with the participants' limited perspective and the inflexibility of avatars, participants use talk and vocalisations, such as 'hang on hang on', 'I am looking', 'er:::::::', to inform their co-participant(s) that the relevant action is being undertaken and that it is not yet complete. These utterances 'voice the avatar' and provide resources through which individuals display the ongoing action and project its completion, without having the co-participant expect or demand an immediate response. They are promissory notes, shaping conduct to render audible the activity which is being undertaken in the very course of its production.

Disrupting a world in common

Many instances show that participants face problems when the other's avatar is not visible in their window on the world. However, even when the other's avatar is visible, troubles often emerge. In particular, certain idiosyncrasies of the technology 'hide' how avatars are viewing, and acting in, the digital domain. Participants often presuppose mutual accessibility in shaping their conduct, only to find that how they see the world is incongruous with their co-participant's perspective at any particular moment in time. It was difficult

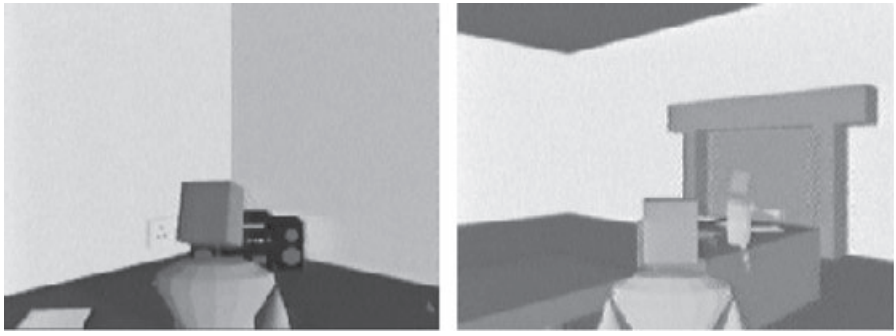
for example for participants to get a sense of how they appeared to the other(s).

In relation to the last fragment (Fragment 1), not only does the field of view cause problems for the person trying to find an object being referred to, but it also raises difficulties for the person doing the referring, namely Sarah. Note how she asks 'Can you see- what I'm pointing at now?' and then part way through the fragment, she says 'Erm, where are you?' before turning to look for her colleague. In everyday environments, referrers will routinely shape, time and organise their talk and bodily conduct *for the recipient*, even to the extent of delaying the production of a pointing gesture for a few moments while a colleague gets in a position to see more clearly (Hindmarsh and Heath, 2000b). In CVEs, referrers are usually not in a position to be able to see both the other and the object, and therefore attempts to produce actions that are subtly sensitive to the conduct of the other are rendered highly problematic. In Fragment 1, at the start of the sequence, Sarah can just see Karen's avatar on the edge of her screen. However, within moments Karen's avatar slips out of view as she is turning to find Sarah. At this point, Sarah explicitly asks whether Karen can see where she is pointing and then later turns to find where she is in order to provide her with directions to the 'desk-thing'.

Problems associated with how others are visually engaged in the world can lead to further difficulties in coordinating multi-party action. For example, the lack of access to the emerging visual orientation of others to an object under discussion leads to particular problems and difficulties in three-party interaction. In the case at hand, it takes just one participant to say that they can see the referent, for the speaker to proceed.

Fragment 2
A: You see this sofa here? That's ermm::
B: Yeah.
A: That's about in the right place isn't it?
((C turns to find the sofa))
C: Where?
B: [°The sofa? This one?°
A: [For the TV and everything like that. An: and the tables aren't too bad cos look this little coffee table here's alright I think.
B: But this sofa's in totally the wrong place.

After Alan's question about whether the sofa can be seen, he momentarily waits for confirmation. However once Brenda alone confirms that she can see the sofa, Alan immediately continues with a discussion of the merits of its current location. Unfortunately, this leaves the third party (Carrie) struggling to find the sofa. This problem is exacerbated because Alan not only moves the discussion on but drops his pointing gesture which provides one major resource for finding the location of the sofa. Carrie starts to turn around to find



Pete's View Rick's View

Figure 4 *A Comparison of Views*

the sofa and indeed her participation in the ongoing discussion becomes restricted as she tries to interrupt the others to 'catch-up' with the conversation. Unfortunately her interruption ('where?') does not manage to elicit further direction to the sofa and it is not for a further few moments before she engages in the conversation (after this transcript ends).

In a more extreme example consider how Rick and Pete fail to share a joke due to the limitations of the vertical field of view.

Fragment 3
R: Pete, what are you actually doing hah hah.
R: I'm just pointing at things.
R: heheheh you're standing in the middle of the desk.
P: Am I? Ooh sorry.

Rick asks Pete what he is doing and then chuckles: 'Pete, what are you actually doing hah hah'. Rick replies quite simply with 'I'm just pointing at things'. He is unaware that his avatar, from the co-participant's point of view, is standing in the middle of the desk (one of the objects in furniture world) (see Figure 4). If we look at Pete's view, the perspective from just to the rear of his own avatar, it is clear that he is unable to see where he is standing and, more importantly, unable to determine how he appears from the standpoint of his co-participant. Interestingly, the very design of Rick's question assumes in part that Pete is able to see how he is being viewed and what might be noticeable or humorous about his actions. We find that the participants presuppose that what they see, and the ways in which to see it, is available to the other, just as they assume that the other's perspective is available to them. Thus we find an intriguing 'incongruity of perspective'; an incongruity which is often discovered by chance in the course of particular actions and activities.

Once again, due to the physical characteristics of the computer screen, there is a restricted *vertical* field of view, which leads to difficulties for individuals interacting because the idea of what might be a remark-able or comical aspect

of the scene is unavailable to one of the participants. Indeed not only is the phenomenon unavailable to him at that particular moment, but he is unable to shift his viewpoint to be able to see it at all. The common approach to the design of avatars assumes that having legs (and a general pseudo-humanoid form) is important, but in this instance it turns out to lead to confusion rather than making social interaction more 'realistic'.

Corrupting trajectories of action

Practical problems also arise for participants with regard to the collaborative manipulation of objects; that is, when one participant moves an object and the other directs that movement. When participants move an object in the CVE, their avatar does not turn with it. Therefore, if they need to move an object out of the range of their field of view, they have to move it in stages – to the edge of their on-screen viewpoint, drop it, turn and then move it again. In Fragment 4, Mark and Ted are discussing where the standard lamp should be placed. Mark manipulates the lamp and Ted comments on his actions.

Fragment 4
M: Do you want the -er lamp moving?
T: Could do. see what it looks like.
M: Where d'ya want it put?
T: Erm- you decide
((pause as M moves the lamp))
T: That's right in front of the window, that's probably not so clever actually. over the other side, over here.

Mark drops the lamp and Ted suggests that this is an inappropriate position for it ('that's right in front of the window, that's probably not so clever'). The video data suggest that Mark has not completed the movement and that this is a first 'step' in the positioning of the lamp. Notice in Figure 5 that Mark places

'T: That's right in front of the window, that's probably not so clever actually.'

Figure 5 *Mark's View as he Moves the Lamp*

the lamp to the very edge of his screen before immediately beginning to turn to his left. However, the technology 'hides' this movement from Ted and at this point he rejects the positioning.

Ted cannot see how Mark is viewing and engaging with this activity. All he can see is the object being moved into an inappropriate position. So, he suggests an alternative. Ted is prevented from getting a sense of how Mark views the lamp. He cannot see that the object is crossing the edge of Mark's window on the world. He is unaware that Mark cannot move the object further without first temporarily releasing it and turning around. In such a way, the 'boundary' of Mark's actions and his physical engagement with a desktop interface are hidden from Ted.

In everyday situations, if an individual finds an object too heavy or too awkward to carry, they can display the temporary nature of its setting-down through their visual conduct and demeanour. Bare virtual avatars, on the other hand, conceal reasons for the placement of an object. Thus, the technology 'disrupts' access to the trajectory of action, the trajectory of moving the lamp. Objects that seem to have been placed once and for all, are often still in the process of being moved somewhere. This renders problematic an individual's ability to assess what the other is doing and how they are engaging with the objects in the digital domain. The technology thereby conceals critical aspects of the ongoing activity.

So, although an avatar and the relevant object may be visible on screen, the relations of one to the other may not be visible or available. Moreover, the action is visibly produced in a different light to which it is seen and understood. This is important, because it makes it difficult for an individual to imagine 'being in the other's position'. As a result it is difficult to assess the nature or trajectory of the other's actions. Thus, 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. The physical character of using a desktop computer is hidden in the virtual display and this leads to confusion about what they are doing. The technology distorts access to the common resources for working with others and making sense of their actions. Thus an individual's understanding of the activity at hand can be disorientated by the technology.

Misleading avatars

It was mentioned earlier how the common approach to avatar design in VR prioritises pseudo-humanoid form. In part this is due to the concern to attain graphical realism and yet due to the material foundations to action in virtual environments, can exacerbate problems in collaborating with colleagues. In the following instance, for example, Pete is explaining to Rick where the fireplace is located. As he does so, both Rick's avatar and the fireplace are visible on his screen.

Pete's View Rick's View

Figure 6 *A Comparison of Views*

Fragment 5
P: Do you reckon it might be better if we moved the T.V. over by the fireplace?
(.)
R: By the fireplace?
P: Yeah [in the cor-
R: [Is there a fireplace in here?
((R Rotates))
P: In the cor- yeah you're facing it now.

Although Rick has the fireplace in his view, he does not recognise it as a fireplace. So when he says 'is there a fireplace in here?', he simultaneously begins to rotate his avatar to the right to look for it. Pete treats Rick's avatar as still facing or at least able to see the fireplace ('you're facing it now'). Unfortunately, at the moment he says this, Rick's viewpoint is focused on the door to the right of the fireplace.[8] Compare their viewpoints in Figure 6.

This reveals a problem for participants in assessing what the other can see. Even though they may have the other 'on-screen', it is hard for them to ascertain what is visible on that other's screen because their view on the world is not available. Indeed the avatar used to somehow 'represent' or reveal it seems to confuse the issue. In other cases, for example, participants assume the availability of their gestures, when the other is visible to them. Unfortunately, it turns out that the other cannot see them.

It may be that seeing a pseudo-humanoid form for an avatar is confusing. This kind of 'embodiment' may give participants a sense that it possesses a roughly 'human-like' field of view, ie around 180°. However, the users' field of view in this CVE is only 55°. Moreover, the CVE does not facilitate stereoscopic vision and the avatars are often large virtual distances from their interlocutor(s), which exacerbates the problem, further concealing the other's perspective. So, it is very difficult for participants to assess what the other

might be able to see. It is not simply that they need to get the other 'on-screen', because even then their sense of what the other is seeing is confused. Thus, the strived-for realism works to conceal the view available to the other user. In the attempt to replicate everyday face-to-face interaction and the utility of the body as a resource for interaction, the avatar design in fact renders problematic our abilities to read the action through embodied conduct. The material interface can therefore render the pursuit of immaterial realism both ineffective and counter-productive.

Discussion

The theoretical concern with the potential impact of instantiations of the 'virtual' on contemporary society continues to grow. However these accounts are often based on design visions for VR and they routinely fail to consider the existing status and emerging character of VR. Thus, this paper has attempted to balance emerging theoretical work with empirical materials concerning the organisation of action and interaction within virtual environments. Indeed, the data provide quite a caution to those interested in the impact that virtual reality could have on society. For example, the new era thinking evident in the sociology of the virtual might raise the question of whether the participants in our data fragments (Karen and Sarah, Pete and Rick, Trevor and Mark, etc.) are 'cyborgs'? Certainly there are claims that the development of VR brings about the advent of the cyborg (eg Dery quoted in Hillis, 1999). We could indeed talk about these people as cyborgs or the avatars as 'virtual prostheses'. Hopefully, however, the fragments discussed in this paper reveal a cyberspace more akin to *Thunderbirds* than *The Matrix*. The avatars (or puppets) are animated not simply by their physical manipulation but more strikingly by the voice that breathes life into them. Intriguingly we find that far from the unconstrained, immaterial worlds that are envisaged in many sociological accounts, current VR is a clumsy place. Furthermore these issues are not simply due to current instantiations of VR but point to significant difficulties facing designers over the coming years; difficulties that relate to support for interaction in virtual environments as opposed to issues of graphical realism (which has improved markedly since these data were collected). So, an uncritical adoption of these design visions for virtual reality is highly problematic, especially when faced with data like these.

Much of the interest in the possibilities of VR to enable participants to transcend the limits of the flesh relates to issues of concealing the impact of visible features of identity inevitable in face-to-face interaction. This aspect of virtual technologies has been undermined by recent empirical work. For example, Hardey's work on internet dating reveals how people pursue 'real world' relationships and face-to-face encounters rather than attempt to create and sustain otherworldly or fantasy selves (Hardey, 2002). Our study extends these concerns by revealing further limitations to the visions of the immaterial

and disembodied character of VR. The paper highlights some of the different ways in which material features of technology use impact on the organisation of interaction in these virtual or 'immaterial' worlds. These 'immaterial' environments are used, experienced or indeed 'entered' in very material ways. Descriptions of action in virtual environments, whether by scholars or participants, often tend to highlight action in the digital domain whilst adopting the language of face-to-face embodied action. This may involve discussions of 'talking' when participants are actually sending a text message, or 'walking' when they are actually clicking a button on a joystick. Such descriptions conceal the achieved character of action and interaction in virtual worlds and the essentially material foundations that underpin the digital domain. On the one hand, we have seen how physical factors, such as screen size, impact on the interaction between people – how they assess others' actions and orientations; how they make sense of objects in the virtual world; how they orient to or constitute a (virtual) world in common. On the other hand seen how the production of action in the virtual world relies on the operation of material technologies – mouse, screens, wires, helmets – and these demand 'working the avatar' by grappling with technologies. However, participants are not allowed to see the virtual action in the context of its embodied production. Rather they see only the immaterial pseudo-realism of the CVE. They are denied access to the context of action production. The physical engagement in the virtual world is concealed and indeed ruptured from the movements of the bare avatar.

Thus, the empirical materials presented here reveal the critical interdependence of action in the physical and the digital realms and the ways in which embodied conduct animates the immaterial domain. The various problems associated with working with computer screens, computer mice and the like lead to profound difficulties in reconciling views on the virtual world, assessing trajectories of action and even identifying what might be noticeable or remarkable in the shared world. It is useful to contrast this with Dorothy Smith's claims that practices of referring '. . . organise among participants in a social act a shared universe of objects' (Smith, 1996: 185). However, when our basis for these practices is fractured, it can lead to the kinds of problems noted in our studies of VR. That is to say, assuming the mutual availability and visibility of 'common' objects and indeed our own 'actions' can lead to difficulties in coordinating and organising social interaction.

Unfortunately many sociological accounts of cyberspace focus purely on the immaterial domain and neglect the import of the physical sites of production of action in those virtual worlds. It would seem critical to focus instead on the relationship between the physical and the digital.

Our studies seem to suggest various methodological recommendations for sociologists of the virtual. At the very least it would be worth founding claims on empirical evidence of current technologies and trajectories of technological design. Furthermore in analysing those technologies in use, the data presented here encourage a concern for both action in the virtual and the physical realms as they have been shown to be fundamentally interrelated. Indeed, this rec-

ommendation can be broadened out, for whilst we have focused on virtual reality, often claims about virtual society relate to more mundane technologies – mobile phones, email, the Web and the like – and the analytic issues surrounding communication are likely to be just as relevant there. Thus sociological accounts might be richer through serious consideration of the practical, embodied production of 'virtual' action.

A further methodological contribution of the paper is to suggest that there is significant value in the study of prototype or inchoate technologies, what Suchman *et al*. (2002) call 'technologies-in-the-making'. This study demonstrates that the ways prototype technologies are encountered and managed by 'novice' users can reveal much about the nature of everyday interaction and can throw into stark relief the taken-for-granted practices and resources that we draw on in sustaining sense-able social encounters. So, rather than being overwhelmed by the complexities of the new technologies, an all too common feature of sociological writings, such studies can remind us of the complexities of mundane social interaction; the artful practices that animate everyday life. These technologies routinely corrupt, or to borrow from Garfinkel (1967), 'breach', conventional resources for coordinating conduct in real-time. Thus the paper demonstrates the value in undertaking qualitative studies of inchoate technologies such as Collaborative Virtual Environments.

Whilst it may seem that these issues and problems might be tied simply to the prototype technologies we used, it seems that they are of more widespread significance. There is evidence to suggest that the problems permeate more popular, but in some ways more basic, instantiations of VR. Massively Multiplayer Online Games (MMOGs) are an increasingly prevalent in today's virtual society and the issues described here seem equally relevant for them (see for example Brown and Bell, 2004; Ducheneaut and Moore, 2004). Indeed whilst we point towards some of the potential problems of multi-party interaction in these domains in Fragment 2, these difficulties will be heightened in massively multi-player environments. Furthermore, whilst VR is proving a valuable domain for supporting game playing, the issues raised by our data may prove more difficult to overcome when considering more complex object-focused discussions, such as with planning surgical procedures. This will become increasingly significant as it is argued that there are heightened opportunities for associated collaborative applications to emerge (Biocca and Levy, 1995) following the growing adoption of 3D visualisations and single-user VR technologies within numerous industrial and entertainment domains (Schroeder, 1996).

King's College London

Received 19 December 2003
Finally accepted 2 February 2006

Acknowledgements

This work was initially supported by the ESRC (Grant R000237136). We are grateful to Steve Benford, Tony Glover and Chris Greenhlagh for their contributions to the research and we would

like to thank Barry Brown, Tim Dant, Jason Rutter, the reviewers and members of the WIT research group for their comments on earlier versions of the paper.

Notes

1 Interestingly the character of these design visions is not exclusive to the emergence of Virtual Reality or even to the 20th Century. Indeed similar statements of intent regarding the pursuit of graphical realism foreshadowed the proliferation of television, film and photography. The search for an 'essential copy' is even found in the history of art, as during various epochs '. . . painters expressed the desire . . . for a canvas that would become a magic window or mirror on the virtual world created with brush stroke.' (Biocca *et al*., 1995: 8).
2 'Collaborative Virtual Environments' is a term commonly used in technical disciplines associated with VR research. It originally referred to systems providing 3D graphical, and often audio, media to multiple users; or 'multi-user virtual reality'. In recent times, the designation 'CVE' has been applied to an increasingly broad range of systems which are not discussed here, such as MUDs and 2D graphical spaces.
3 MASSIVE stands for 'Model, Architecture and System for Spatial Interaction in Virtual Environments'.
4 Media spaces are infrastructures that incorporate combinations of audio, video and computing links to interconnect remote participants in common activities (see Heath and Luff 2000).
5 As part of our procedures for securing informed consent, all users were warned of potential problems related to the use of VR. Following University ethics procedures, we also insured that no participants suffered from any of the following: 'migraine, recurring headaches, pregnancy, back pain or back problems, neck or shoulder strain, heart condition, asthma or respiratory disorder, epilepsy (photosensitive or other), problems with depth perception or other serious injury or illness'. However during our short trials, no participant experienced any physical difficulties.
6 Head-Mounted Displays provide similar horizontal and vertical fields of view to desktop displays.
7 These images, as with many of the other illustrative images in the paper, are taken from actual video data, and thus are subject to the resolution constraints of this medium. Note also, that for this one trial we did not provide users with a view of their own avatar.
8 In this case the delivery lag is negligible. In other instances, however, the lag disrupts the notion that this is a stable, common environment.

References

Barnatt, C., (1995), *CyberBusiness: Mindsets For a Wired Age*, Chichester: John Wiley.
Biocca, F., Kim, T. and Levy, M.R., (1995), The Vision of Virtual Reality, 3–14 in Biocca, F. and Levy, M.R. (eds) *Communication in the Age of Virtual Reality*, Hillsdale, NJ.: Lawrence Erlbaum.
Biocca, F. and Levy, M.R., (eds) (1995), *Communication in the Age of Virtual Reality*, Hillsdale, NJ.: Lawrence Erlbaum.
Bowers, J., O'Brien, J. and Pycock, J., (1996a), Practically Accomplishing Immersion: Co-operation in and for Virtual Environments, 380–389 in *Proceedings of CSCW'96*, New York: ACM Press.
Bowers, J., Pycock, J. and O'Brien, J., (1996b), Talk and Embodiment in Collaborative Virtual Environments, in *Proceedings of CHI'96*, New York: ACM Press.
Brown, B. and Bell, M., (2004), CSCW at play: 'There' as a collaborative virtual environment, 350–359 in *Proceedings of CSCW 2004*, New York: ACM Press.

Büscher, M., O'Brien, J., Rodden, T. and Trevor, J., (2001), 'He's Behind You': The experience of presence in shared virtual environments, 77–98 in Churchill, E., Snowdon, D. and Munro, A. (eds), *Collaborative Virtual Environments*, London: Springer Verlag.

Dale, K. and Burrell, G., (2000), What Shape Are We In? Organization theory and the organized body 30 in Hassard, J., Holliday, R. and Willmott, H. (eds), *Body and Organization*, London: Sage.

Ducheneaut, N. and Moore, R.J., (2004), The social side of gaming: a study of interaction patterns in a massively multiplayer online game, 360–369 in *Proceedings of CSCW 2004*, New York: ACM Press.

Featherstone, M. and Burrows, R., (1995), Cultures of Technological Embodiment: An introduction, 1–20 in Featherstone, M. and Burrows, R. (eds), *Cyberspace/Cyberbodies/Cyberpunk*, London: Sage.

Garfinkel, H., (1967), *Studies in Ethnomethodology*, Cambridge: Polity Press.

Goodwin, C., (2000), 'Action and Embodiment within Situated Human Interaction', *Journal of Pragmatics*, 32: 1489–1522.

Green, N., (1999), Strange Yet Stylish Headgear: VR consumption and the construction of gender, *Information, Communication & Society*, 2 (4): 454–475.

Hardey, M., (2002), Life beyond the screen: embodiment and identity through the internet, *The Sociological Review*, 50 (4): 570–585.

Heath, C. and Hindmarsh, J., (2000), Configuring Action in Objects: From mutual space to media space, Mind, *Culture & Activity*, 7 (1/2): 81–104.

Heath, C. and Hindmarsh, J., (2002), Analysing Interaction: Video, ethnography and situated conduct, in May, T. (ed.), *Qualitative Research in Action*, London: Sage: 99–121.

Heath, C. and Luff, P., (2000), *Technology in Action*, Cambridge: Cambridge University Press.

Heim, M., (1995), The Design of Virtual Reality, 65–78 in Featherstone, M. and Burrows, R. (eds), *Cyberspace/Cyberbodies/Cyberpunk*, London: Sage.

Heim, M., (1998), *Virtual Realism*, New York: Oxford University Press.

Hillis, K., (1999), *Digital Sensations: Space, identity and embodiment in Virtual Reality*, Minneapolis: University of Minnesota Press.

Hindmarsh, J. and Heath. C., (2000a), Sharing the Tools of the Trade: The interactional constitution of workplace objects, *Journal of Contemporary Ethnography*, 29 (5): 523–562.

Hindmarsh, J. and Heath, C., (2000b) Embodied Reference: A study of deixis in workplace interaction, *Journal of Pragmatics*, 32 (12): 1855–1878.

Hindmarsh, J., Fraser, M., Heath, C., Benford, S. and Greenhalgh, C., (2000), Object-Focused Interaction in Collaborative Virtual Environments, *ACM Transactions on Computer-Human Interaction (ToCHI)*, 7 (4): 477–509.

Hughes, J.A., O'Brien, J., Randall, D., Rouncefield, M. and Tolmie, P., (2001), Some 'Real' Problems of 'Virtual' Organisation, *New Technology, Work and Employment*, 16 (1): 49–64.

Jordan, T., (2001), Language and libertarianism: the politics of cyberculture and the culture of cyberpolitics, *The Sociological Review*, 49 (1): 1–17.

Lanier, J., (2000), Virtual Reality is the Telephone of the Future, in Rucker, R. (ed.), *Mondo 2000: A user's guide to the new edge*, New York: Harper Collins.

MacKenzie, D. and Wacjman, J., (1985), *The Social Shaping of Technology: How the Refrigerator Got its Hum*, (first edition) Milton Keynes: Open University Press.

McCormick, N. and Leonard, J., (1996), Gender and Sexuality in the Cyberspace Frontier, *Women & Therapy*, 19 (4): 109–119.

Schroeder, R., (1996), *Possible Worlds: The Social Dynamic of Virtual Reality Technology*, Boulder: Westview Press.

Schroeder, R., (1997), Networked Worlds: Social aspects of multi-user Virtual Reality technology, *Sociological Research Online*, 2 (4), <http://www.socresonline.org.uk/socreonline/2/4/5.html>

Smith, D.E., (1996), Telling the Truth after Postmodernism, *Symbolic Interaction*, 19 (3): 171–202.

Smith, R.B., Hixon, R. and Horan, B., (1998), Supporting Flexible Roles in a Shared Space, in: the *Proceedings of Computer-Supported Cooperative Work 1998*. New York: ACM Press: 197–206.

Suchman, L., Trigg, R. and Blomberg, J., (2002), Working Artefacts: Ethnomethods of the prototype, *British Journal of Sociology*, 53 (2): 163–180.

Sutherland, I., (1965), The Ultimate Display, in *Proceedings of the IFIP Congress*, 2.

Thrift, N., (1996), New Urban Eras and Old Technological Fears: Reconfiguring the goodwill of electronic things, *Urban Studies*, 33 (8): 1463–1493.

