Chapter 4. Collaborating with Objects in Furniture World

This chapter will discuss the initial developments and experiments through which object-centred interaction was explored. Chapter 2 has already described that, whilst many distributed CSCW systems provide some technological support for working with objects, very few investigations have explored how these technologies enable (or disable) collaboration with and around artefacts. One significant exception to this has been the work on the MTV mediaspaces developed at Rank Xerox EuroPARC (Gaver et al., 1993, Heath et al., 1995). The previous chapter mentioned the relevance of methodological implications of the MTV studies for studying interaction between technologically distributed participants. However, the concerns and findings of these experiments also provide direction for an exploratory investigation into the role of object-focused work in CVEs.

This chapter will follow the course of just such an exploratory investigation. From initial construction, through to analysis and implications for design, this chapter charts work that motivates the technological and analytic developments pursued within this thesis. Firstly, it will discuss the adaptation of the concern and task of the MTV experiments for CVEs, and mention some reasons why this course might prove beneficial. Secondly, it will describe the functionality added to an existing CVE system, MASSIVE-2, to support object-focused participation in general, and the relevant task in particular. Thus, thirdly, it will present the resulting CVE application and task called ‘Furniture World’. Fourthly, this chapter will discuss the experimental data generated from these trials. The resulting exploratory evaluation of the Furniture World data will then be outlined. This analysis was produced as part of collaboration between the author and the Work, Interaction and Technology group at King’s College London, and published in (Hindmarsh et al., 1998). The evaluation itself was conducted by Jon Hindmarsh, and is reported in this chapter as such. Finally, the
implications of this analysis for designing CVEs to support collaborative work with virtual objects will be raised.

4.1 Background

Chapter 2 has already mentioned the ways in which CSCW systems support collaborative work with objects. In particular, it is felt that there has been little consideration in CSCW systems development of the ways in which distributed workers might accomplish interaction with and through objects. Additionally, it has been noted in chapter 3 that an important exception has been work on the MTV mediaspaces (Gaver et al., 1993, Hindmarsh, 1997). In this section, the reasons for and findings of (rather than method of) this work will be mentioned, in order to outline the design of the exploratory virtual environment and task, Furniture World.

MTV I and II were mediaspace systems developed at Xerox Europarc in Cambridge, England. MTV studies have been published in a number of research papers and articles (Gaver et al., 1993, Heath et al., 1995, Hindmarsh, 1997). The mediaspace design consisted of two participants in separate offices, aware of each other through a number of cameras capturing views in each office and a monitor presenting the analogous views of the other. The camera views used for the systems changed with each iteration, but included a wide-angle view of the office, an over-the-shoulder view of the participants, a face-on view of the participant, and a view of the desktop in the office. MTV I allowed participants to switch between the camera views on a single monitor, using a round-robin switching mechanism. The important difference between the two systems was that MTV II included multiple monitors to display each of the video views simultaneously, thereby avoiding the use of a switching mechanism. This change was instigated because it was found that participants had trouble getting an overall and coherent sense of the other’s workspace through the former system. Understandably, it was felt that presenting all views at all times might alleviate this problem.
However, the findings of these experiments highlighted that object-focused mediaspace discussions were still highly problematic in MTV II, namely a ‘fragmentation’ of different elements of the collaborative space in which the work occurred. Hindmarsh (1997) describes the following example, which gives a good indication of such problems in the MTV II system. Doris (D) and Colin (C) are engaged in arranging furniture in an architectural modelling design task. The aim of the task is to accomplish the design of a room using a physical doll’s house model located in D’s office (and shown in C’s view from the desktop camera).

Fragment 9.1 (taken from Hindmarsh, 1997: 209)\(^{17}\)

C: So ah- the speakers need to go: (.s) some somewhere like that:

D: [Yeah, I would suggest that the one on the fireplace, actually comes down to

thi:ss ((D points to own ‘doll’s house view’ monitor)) (0.8) the corner

(.) actually you can’t see where I’m pointing, can you

Doris’s utterance ‘thi:ss (0.8) the corner’ is accompanied by a pointing gesture to her video screen displaying the remote model. The actual model is within Colin’s physical environment. However, as Colin only has access to video views of Doris’s physical environment (and not her video views of his physical environment), he can only see her gesture, and not what it points to. Of course, he has potential access to the corner Doris has in mind – the corner is part of the model located in his physical environment. Without being able to see how Doris sees her gesture in relation to how she sees the model, however, her gesture is rendered senseless to him.

Within the various video views which were provided in MTV II, the participant trying to understand the reference might be able to see the speaker’s face, gesture and the relevant object. However, the relevant aspects of the action would often be presented

\(^{17}\) See Appendix B for the notation used to transcribe conversation.
in different parts of the world, a combination of disjointed video views and local, physical artefacts (Hindmarsh, 1997). Participants could not re-assemble the relations between body, gesture and object and thus object-focused discussions became highly problematic.

4.2 Initial development of Furniture World

When we turn to an initial consideration of working with objects in virtual environments, it might seem that virtual worlds offer a solution to some of these problems. Virtual environments are spatially consistent, in the sense that spatial relationships between objects in the environment, and between those users represented in that environment, are analogous to the three-dimensional world in which we routinely interact. It would seem that the use of virtual environments for object-focused work might prove a worthwhile venture.

The task chosen to be adapted from MTV was the one involving the collaborative design of a doll’s house. This particular setting was picked for the simple reason that an adaptation of it was practically achievable using current CVE technology. Additionally, and as shown in the MTV II data, designing a living room in this way aptly encourages participants to collaboratively engage in object-focused work. The design task of moving furniture around in a CVE seemed interesting for a number of reasons. Firstly, the task invites participants to engage in certain actions that have been shown to be significant in workplace interaction, such as organising gesture, gaze, grasping and moving and so on to work with objects. Secondly, the (virtual) objects involved are not just momentarily invoked for the task, but rather their placement is critical to its accomplishment. The achievement of layout for a room of furniture means that participants may use the location in which objects reside to constitute the completeness of the task. The task in itself therefore appeared suitable for exploring the practices through which participants undertook object-focused work in CVEs (and thereby apt to raise issues for the design of those systems). Thus it was decided that this experiment, derived from MTV II, would prove an interesting task with which to investigate working with objects in CVEs.
It should be noted that the use of this task from the MTV experiments was not intended to assess whether CVEs provide more adequate or better support than mediaspaces for distributed working with objects. Indeed, a direct comparison would be rather deceptive as the technological differences make the task quite different – in mediaspaces, participants have asymmetrical access to a real, physical model. In CVEs, participants have equal access to virtual objects. The use of a task previously analysed in mediaspace research does allow some more general comparison of issues involved in designing systems for distributed interaction. In the main, however, this design task simply provides an opportunity to encourage participants to discuss objects within the virtual environment. It is interesting to note as a counterpoint that the differing geographical and spatial resources provided by these types of systems meant that quite different practices in object-focused work were expected to arise. Indeed CVEs provide coherent spatial relationships between (represented) participants and objects. Therefore, following the emergence of difficulties with mediaspaces for interacting with and through objects, it was anticipated that CVEs might offer some potential solutions to these difficulties.

Furniture World was realised using a particular CVE system, MASSIVE-2. This system is a general purpose CVE platform that has been developed at the University of Nottingham (Greenhalgh, 1999). It allows multiple participants in a shared virtual world to communicate using a combination of 3D graphics and real-time audio, to grasp and move virtual objects, and to populate virtual environments using ‘avatars’. In addition, during the work on this thesis, this system was continually updated and revised based upon different aspects of CVE design and analysis. Therefore, the use of MASSIVE-2 became beneficial in two additional ways. Firstly, the author was allowed the freedom to discuss development directly with those who were creating aspects of the overall system. Secondly, the work presented in this thesis was allowed to propagate within a short time frame, because both relevant program code and explanations were available to others engaged in other aspects of CVE research.18

18 The dissemination of this work into other areas of CVE design is discussed in Appendix A.
Having been developed as a generic CVE platform, and prior to the development process of this thesis, MASSIVE-2 had some existing technical support for manipulating, orienting to, and referring to objects within the virtual world. The remainder of this section describes the support for these activities already provided by MASSIVE-2. This support was a typically *ad hoc* collection of resources with which current desktop CVE systems support object-focused interaction, as described in chapter 2. The provision of techniques to work with objects in MASSIVE-2 had been added here and there, and received no empirical analysis.

In order to manipulate objects in the MASSIVE-2 virtual world, users could pick up and translate or rotate them using a mouse. However, only one method of grasping was provided. Users pointed at artefacts with the mouse cursor, and then selected them with a mouse button. They could then be moved to positions on a constant radius from the user’s avatar. Thus, there was no possibility of moving an artefact away from or towards oneself, without first relocating the avatar to an orthogonal position. Additionally, there was no representation of this action available on the avatar; it was solely shown through the movement of the object itself. In a multi-user environment, there was no designation of who might be making a certain object move.

MASSIVE-2 already supported spatialised audio communication between users, through headphones and a microphone. Thus it was possible to refer to an object by discussing its features or location with others connected to the virtual environment. In addition, the default MASSIVE-2 avatar had a recognisable face and body, and thus denoted the general orientation of the user’s display. One could be seen within the environment to be facing a certain direction, and thus facing a certain object. However, there was no support for explicit referential gestures, such as pointing.

It seemed clear, therefore, that various aspects of the MASSIVE-2 system would need to be enhanced in order to produce an environment that could be used for experimentation. The following sections describe the revisions made to both avatar
and interface in order to support the particular ‘Furniture World’ experiment to be undertaken.

4.2.1 Revising the MASSIVE-2 Avatar

It was clear that the default MASSIVE-2 avatar would need to be enhanced, in order to more explicitly support grasping and referencing objects. It was decided that the new avatar design should convey the impression of a ‘human-like’ appearance. Although photo-realism was not possible for system performance reasons, adopting this approach seemed the most obvious choice, given the ‘human-like’ activities such as gesturing, picking up and moving objects and so on in which users would engage. These activities could be represented in avatar forms that were easily identifiable with those of their human counterparts. Indeed, the reproduction of humanoid figures has dominated the representation of users by CVE designers for these very reasons. This meant that the avatar should be have a recognisable head, body and limbs.

In addition to referencing objects in talk, or by facing them, as have already been described, the avatar design supported pointing gestures as a way of referencing objects. A participant could choose to point at a target (an object or a general area of space) by selecting it with the right mouse button. The avatar would then raise its nearest arm to point at the target and would also incline its head towards it, as shown in Figure 4.1.
Users were able to grasp objects by selecting them with the left or middle mouse buttons and then position them by moving the mouse. The left button corresponded to a translation of the virtual object across the floor of the virtual room. Selecting with the middle button rotated it around its vertical axis. In order to simplify the task in which users would participate, the ability to lift objects and rotate them around other axes was removed (although these abilities would clearly be required in other tasks). These activities were also shown on the avatar by raising an arm and tilting the head towards the object, in the manner of a pointing gesture. In addition, a connecting line was drawn between the avatar and the object being moved. This connecting line was the only extension beyond a normal humanoid avatar. It was included for two reasons. Firstly, the connecting line would differentiate, for an observer, the action of pointing at an object as opposed to grasping it (and also presumably for users, if they became confused as to which button performed which activity). Secondly, the connecting line was provided to reflect the ability to move objects at a distance in the CVE. This is not a familiar experience in the physical world, yet the virtual environment will enable one to move artefacts without being within arm’s length of
them. An example of the user’s view of the world whilst grasping a sofa is shown in Figure 4.2.

![Figure 4.2: A User Grasps a Sofa](image)

**4.2.2 Revising the MASSIVE-2 Interface**

The interface to MASSIVE-2 was modified with two aims in mind. Firstly, the additional functionality described in the previous section was made visible to the user. This involved adding certain features to feed information back as to which objects were being pointed at or grasped, as described in greater detail below. Secondly, MASSIVE-2 supported an array of interface techniques and tools, which were felt unnecessary and over-complex for this particular experiment. Thus, the interface was simplified in order that participants could only carry out a limited set of actions. The actions available through the revised interface design were *looking* (i.e., navigating so as to adjust one’s view of the virtual world), *speaking*, *pointing*, and *grasping* (to translate or rotate objects).
Specific aspects of the interface included for the task were:

- Restricting movement to the ground plane only. In other words, neither avatars nor objects could be moved in the vertical plane. Users could not ‘fly’ up and down, and objects could not be moved up and down (for example, in was not possible to place one on top of another).

- The use of an existing 2D map facility to allow a ‘birds-eye’ view of the environment. This was provided to allow users to inspect intricate relationships between objects (e.g. placing one exactly adjacent to another), if these perspectives were difficult to discern in the 3D view.

- A reporting facility to give feedback on object selection. This tool displayed the actions of pointing at an object, rotating an object and moving an object. It was included for two reasons. Firstly to give the task objects their own nomenclature: to let users know that a certain artefact was, for example, ‘the comfy chair’ and thus allow them to discuss artefacts without the confusion of deciding which was which. Secondly to remind users that mouse buttons performed particular actions...
– the three reporting texts, which described the relevant actions, were ordered in the same way as the operation of the mouse buttons.

- Use of an ‘out-of-body’ camera view that showed one’s own body in the foreground of the scene. This technique was initially introduced in the MASSIVE-1 system to extend one’s field-of-view and to provide feedback as to the state of one’s avatar (Greenhalgh and Benford, 1995). When testing the implementation of pointing with a viewpoint situated through the avatar’s ‘eyes’, it was not possible to see the ‘arm’ or its movement. Thus the only feedback that pointing was in progress was through the reporting facility on the interface. A ‘behind the body’ view allows participants to see their own avatar pointing and grasping. The field-of-view provided in our application was 55 degrees horizontally and 45 degrees vertically, in order to minimise the distorting effects of a wide view on the virtual world.

4.2.3 The Furniture World Environment

The graphical environment was mostly constructed by creating 3D models using the AC3D modelling package (Colebourne, 2000). The simple design involved creating walls, floor, ceiling, a fireplace, two plug sockets, a door and two windows. These artefacts were background items intended to simply give the feel of a living room, and to constrain or inform the design to some extent. Various artefacts intended for use in the design task were then scattered to populate the room. The ‘sofa’ object was obtained from a free repository of 3D objects on the Internet. However all other artefacts were produced by the author, hence the observably limited artistic flair in the room design (Figure 4.4).
4.2.4 Data Collection

Six trials of two participants and two trials of three participants were performed. Most participants were students, twelve male and six female, with a broad mixture of previous acquaintance. None of them had a background in CVE technology. Each trial took approximately one hour and consisted of the following stages:

- ten to fifteen minutes for participants to get used to the system.
- half an hour to perform the given task.
- a short interview concerning the participants’ experience of using the system.

The participants in the experiments used Silicon Graphics O2 workstations connected via an Ethernet, with speech being supported by headphones and a microphone. A VCR was used to record each participant’s on-screen activities and audio from their ‘auditory perspective’ – their own voice in real time, plus the other participant’s voice(s) after a short delay over the network. Depending upon network traffic, the delay varied from being almost negligible up to imposing a one-second difference on
transmission of audio and graphics. Video cameras simultaneously recorded the participants in the real world and contained audio from the participant in shot (Figure 4.5).

![Image](image_url)

**Figure 4.5: A User Being Filmed**

The interview with participants was videotaped, and used as a guide to the different experiments – to ‘get a feel’ for the individual session as the report of the participants had it. The interview data was not directly used in analysis, but rather to give a sense of the situation as recalled by participants prior to inspection of the video data.

Additionally, it is interesting to note that the ‘real-world’ view of participants sitting at their computer terminals was rarely presented in the final analysis. The isolation in which participants were asked to perform the task meant the distractions from the real, physical, environment were few. Participants seemed to focus on the task within the virtual environment, and thus the data showing the surroundings of participants was only necessary when actions produced in the virtual environment had seemingly little relevance to the virtual context in which they were produced\(^{19}\). Nonetheless, the

\(^{19}\) See Bowers et al. (1996b) for a discussion of the relevance of real-world interaction to the sense of that achieved in the virtual world.
exploratory analysis was initially approached (by parties other than the author; see below) using a four picture-in-picture shot containing both real and virtual camera shots for each participant to get a more complete sense of their interaction. Where three participants undertook the task, the picture-in-picture recording simply showed the three views of the participants’ on-screen (and thus in-world) activity.

4.3 Furniture World task

Participants were asked to collaboratively arrange the layout of furniture in the room and agree upon a single design. They were told that the virtual environment was being used to plan the layout of a real showroom in the Ideal Home Exhibition (an annual event for homeowners held in the U.K.). Additionally, they were informed that their company and another (the representative of which was the other participant) had been accidentally given the same space in which to create a single living room for a house. One participant worked for ‘The Tasteful Design Company’ and one for ‘Cheap and Basic Designs’. Where a third participant was involved, they worked for the ‘Abstract Design Company’. These conflicting roles were provided in order to encourage the participants to debate and discuss their ideas, and to give the opportunity for both collaboration and conflict. It was not intended (or indeed pursued) that any kind of consideration would be given to how ably or accurately the remit of the task itself was fulfilled. Therefore not completing the task was also not considered relevant or problematic. The finalised and, where relevant, incomplete living room environments were ignored, deleted upon the departure of the participants, and re-set for the next trial.

4.4 Exploratory analysis

This section will discuss the analysis of the data generated in the Furniture World trials. It is hoped that the orientation and methodology described in the previous chapter will become clearer in its application to empirical materials. As previously mentioned, the author was not primarily involved in the production of this evaluation,
yet the findings and issues raised are intrinsic to the process of this thesis. Thus, whilst involved in the exploratory Furniture World trials in other ways described, the reader should treat the following analysis as a discussion of the work of Hindmarsh, which informs the subsequent technical and evaluatory chapters of this thesis.

Inspection of the collected video data was used to provide intricate and detailed results concerning the modes and methods through which participants had accomplished the task asked of them. In particular, the problematic ways in which participants collaborated around and with objects in the virtual world were highlighted. These problems are discussed in the following sections.

The first section will consider the practical use of a CVE in working with objects. It has already been outlined that the MTV mediaspaces fragmented the environment through which objects could be invoked and referred to. Additionally, it has been mentioned that the use of virtual environments might be thought to provide a solution to some of these problems, allowing a common spatial frame of reference with which to assemble and understand the on-going actions therein. This section discusses empirical evidence that shows how this assumption may or may not hold. Ultimately, however, it describes how the affordances of the space itself, as enabling this kind of coherent work, cannot be partitioned from the abilities of participants to engage in the environment and undertake object-focused activities.

**4.4.1 Achieved coherence in the workspace**

In Furniture World, participants are represented as avatars in a shared and coherent virtual space. This enables them to indicate features for each other within the common environment. The ability to produce a pointing gesture towards an object is seen in relation to the object in question. Thus the spatial coherence of the virtual environment can sometimes provide that the sense of the gesture can be re-assembled and understood. Hindmarsh et al. (1998) provide the following example to show how pointing gestures can be successfully used to encourage co-participants to look at an
object with them. Sarah and Karen are repairing a confusion over which table should be moved.

Example 4.1 (Perspective: Sarah)

K: It’s this table I’m talking about. this one yeh?

((K Points))

S: Yeh.

K: Can you see me point?

S: Yeah, it’s the one in front of you isn’t it.

Figure 4.6 – Sarah’s view as Karen points

Sarah is able to see Karen’s gesture in relation to the object-at-hand. The fact that Sarah and Karen are both represented within a coherent, three-dimensional space

---

20 Talk is transcribed from the video recording of one participant’s perspective.
21 In the examples of talk in this section, square brackets indicate overlapping utterances, a dot in brackets indicates a short pause in talk, and double-brackets indicate a comment by the transcriber of the talk; see Appendix B for more details of transcription notations used.
22 For this single experiment, participants were not provided with a view of their own avatar.
23 This picture, as with the other illustrative pictures in this thesis, is taken from the video data collected from the screen of the users. Thus, they are subject to the combined resolution constraints of the computer display and the video medium.
means that Sarah is able to see the point in relation to the table. Thus she can
distinguish this object (as opposed to, for example, another table in the vicinity) as the
one relevant for the on-going activity. The ability to gesture in a common
environment is used as a successful resource with which to indicate an object in the
virtual world. The representation of Sarah and her virtual gesture are located in the
same domain as the object being referred to. This enables Karen to relate or ‘connect’
the gesture to the object referred to in the talk, and it is this connection that gives
sense to Karen’s talk, to her gesture, and to the object-as-relevant.

In Furniture World, the gesturing avatar and object can potentially be seen alongside
one another. The coherent and recoverable relationship between objects and
(representations of) participants is maintained. We no longer encounter the disruption
of body and object which was indicative of troubles in interaction throughout the
MTV experiments.

It should be noted that this coherence of virtual participation appears critical to the
ability of participants to accomplish the task. The hour provided within which they
were asked to get to know the system, perform the task, and respond to a discussion
regarding their performance, proved entirely sufficient. Indeed, some participants
even claimed that they had enjoyed using the system, or requested to continue as
‘subjects’ in future trials.

However, this is not to say that working with artefacts in Furniture World could be
undertaken without any hindrance. The analysis reported in (Hindmarsh et al., 1998)
describes a number of disruptions to working practice within the scope of this
seemingly simple design task. The following sections discuss the ways in which
participants found object-focused interaction through Furniture World problematic, as
well as considering some of the coping strategies that were formed.

4.4.2 Fragmenting the Workspace

The Furniture World CVE provides a 55-degree horizontal field-of-view of the virtual
world. Desktop displays provide limited horizontal space in which to render a user’s
view of the virtual world. Rendering 3D graphics on such screens uses a perspective projection (Carlomb and Paciorek, 1978) onto the display. Thus the rendering software projecting desktop views must accept a trade-off between causing perspective distortions, and allowing a wide field-of-view on the virtual world. Yet when ‘simulating’ real worlds or tasks as in the case of Furniture World, the realistic (undistorted) appearance of the environment seems paramount. Highly distorted views of such virtual worlds may render realism unclear. Thus, whilst the field-of-view of the desktop CVE can be varied, it is usually limited to a small fraction of a ‘human-like’ field-of-view.

### 4.4.2.1 Objects in view

The limited field-of-view provided in Furniture World restricts participants’ opportunities to see avatars in relation to objects. As opposed to the case of Example 4.1, individuals are often unable to see an avatar alongside the relevant artefact at the time an utterance indicating that object is produced. Yet it is critical to the accomplishment of working with those objects that they are seen in relation to one another. In order to compensate for this inability, participants would firstly turn to find the other’s avatar and then look for the object.

To illustrate this phenomenon, consider Example 4.2, representative of the kinds of such ‘fragmentation’ encountered throughout the Furniture World data. The same participants as discussed in Example 4.1, Sarah and Karen, are talking about the ‘desk-thing’. Before they can decide where they might put the desk, they need approximately 25 seconds to achieve a common orientation towards it.
Example 4.2 (Perspective: Karen)

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- th- I can see
S: Tur- (.). 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.
Figure 4.7 – Karen’s 3D view as she turns from the ‘desk-thing’ to find Sarah’s avatar and then returns to view the same ‘desk-thing’

When Sarah asks if Karen can see what she is pointing at, Karen starts to look for Sarah’s avatar and pointing gesture. Karen is actually facing the object in question, the desk, very early on in the example. However, because she cannot see the gesture, and thus relate it to the desk, she ends up turning away from it to find Sarah’s avatar and gesture. She subsequently turns all the way around to view the very same desk.

Studies of working with objects in co-present interaction reveal that 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 simply turn from the body of the other to find the referenced object (Hindmarsh, 1997). In Furniture World, participants often do not have the other’s avatar in view during an utterance. They might well have initially had the referent in view. Yet 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 this example, Karen can see a ‘desk-thing’; yet without access to, and attending to, the demands of the gesture she is unable to determine if it is the relevant ‘desk-thing’. Participants cannot be certain that they are looking at and discussing the same object without seeing that object in relation to the other’s representation.

Fragmenting the world in this way, separating views of avatars from the objects on which they are acting, creates difficulties for participants. As they are rarely in a position to see both object and avatar simultaneously, participants have problems in relating the two. Critically, the sense of their talk and action is based upon the mutual availability of that relationship. When participants have a referenced object in their view of the world, they need to see it in relation to the other’s avatar. When participants have an avatar in view, they cannot understand another’s action without seeing it in relation to the object. Without seeing both relevant features of the virtual world together, they cannot simultaneously and simply constitute body and object as relevant features of the on-going action. In co-present interaction these resources are often concurrently available. However, in Furniture World, participants often find the relevant object by following a particular sequence. First they turn to find the gesture and then they use this as a resource to find the referent. Even in short and straightforward instances, users can be seen turning away from an object to find the other’s gesture, only to subsequently return to face that object. As has been shown, however, engaging in an extensive search for their co-participant’s avatar before being able to see the relevant object is often necessary.
4.4.2.2 Avatars in view

These problems often arise because the other’s avatar is not visible at the onset of a new activity. However, misunderstandings can also arise when the other’s avatar is visible, but is again separated from the objects made relevant in interaction through the virtual environment. This is particularly relevant in incidents where the intelligibility of the kind of gesture is dependent on seeing both body and object concurrently. Hindmarsh et al. (1998) describe how seeing an object being grasped can be confused with seeing an object being pointed at when that object is not available in the other’s view. This situation arises because the differentiating visual aspects between these actions are the movement of the object and a thin connecting line between the two. The fact that the object will often be out of view when observing the avatar’s representations means that its movement will not be seen. When combined with the very narrow connecting line, which blends in colour against the background features of the room, these two representations may be very difficult to distinguish. This problem points to a more general issue of representing actions when working with objects. The fragmentation of views means that the way in which actions are represented across different parts of the environment is critical to participants’ comprehension of those actions. Seeing the same actions in different ‘fragments’ of the world may change the presentation, and thus understanding, of those actions.

Yet more troubles emerge when the other’s avatar is still visible. Certain idiosyncrasies of the interface ‘hide’ how avatars are viewing, and acting in, the world. Hindmarsh et al. (1998) describe two examples of the ways in which the interface conceals a user’s local experience of the virtual environment.

Firstly, an example shows how it is difficult to recover what another has visible at any particular moment. When one participant can see another’s avatar, there is no indication of the extent of their rendered view, other than the orientation of the avatar. This leads to statements such as “you’re facing it now”, when in fact the limited field-of-view means that the object in the other’s view is actually peripheral or entirely unavailable. Unfortunately this means that it is difficult to take as given the visibility
of actions in, and features of, the virtual world. It is suggested that seeing a human-like avatar gives rise to this confusion. Pseudo-humanoid forms may give participants a sense that the represented user possesses a ‘human-like’ view of the virtual world (180°) rather than the limited desktop view provided (55°). Additional issues exist which compound this problem, further concealing the other’s perspective. Firstly, this CVE does not allow rendering of stereoscopic vision, so there are problems in getting a sense of depth of an avatar, or object, compared to that of another. Secondly, avatars are often large virtual distances from each other during talk, as distance is less of a factor in conversing than it is in co-present interaction (although this CVE supports attenuation of audio volume over distance, this facility was not used in Furniture World). Therefore, talk may be simply accomplished whilst views of the common environment are highly disparate. Thirdly, the view of the user is not actually positioned at the location implied by their avatar. The use of a camera view means there is a discrepancy between the actual and represented positions of the user’s ‘eyes’ in the virtual world.

A second example – that of problems in ‘seeing what the other can see’ – shows how the orientations and intentions of a user moving an object are hidden from the common environment, and thus unavailable to others. Users are only able to grasp and move an object within their 3D view. If they wish to move that object to a point outside of their visual range, they must move the object to the edge of their screen, drop it, and then turn and repeat the process. However, because the extent of the grasping user’s view is not visible within the environment, participants are unable to determine whether an object has been dropped because it is in its intended position, or whether it is at a temporary point on the stepwise transition to its final location. This leads a participant, for example, to disagree with the placement of a lamp even though, as the other sees it, it is residing at what is clearly a temporary location, while they turn around to continue moving it.

These examples show that it is very difficult for participants to assess what the other might be able to see. This multiplies the problems raised in previous sections. It makes it far harder for participants to attempt to design actions for others and to co-
ordinate actions with them. It is not simply that they need to get the other ‘on-screen’, because even then their sense is confused of what the other is seeing. The orientations of the other in the world are hidden from view, which can also lead 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. An individual’s understanding of the activity at hand can be disordered by the CVE interface.

4.4.3 Coping strategies

The previous section highlights a problem related to the narrow field-of-view provided by the CVE. We have seen how participants cope with their inability to concurrently and coherently discover the on-going activities of others. They rely on coping strategies to ensure legible interaction by first finding the gesture and then ‘manually’ connecting to the artefact. However, the gesturing participant is also sensitive to the possibility that the other is not in a position to see their avatar or the objects on which they are acting. They employ certain practices in order to aid co-participants in co-ordinating their actions.

Hindmarsh et al. (1998) note that one clear strategy used by these participants is to use their talk to make visual conduct more explicit, given the limited range of perceptual resources. In co-present interaction an individual may simply say, “what do you know about this?” alongside a gesture. Their co-participant can often turn quite easily to see what ‘this’ is and attend to the query. In such a way, the gesture and the projected activity can be conflated (Hindmarsh, 1997); that is, the presentation of the object and the initiation of the activity (e.g. asking a question) can be one and the same. In Furniture World, participants tend to engage in prefatory sequences of talk and reference in which the identity of the relevant object is secured – for all practical purposes – before the main activity continues. Typical utterances include “erm (.) you see the fireplace which is like (.) there?” or “See this sofa here?”. The activity only progresses when the other has found the referenced object. The participant wishing to make a certain object relevant attends to the difficulties that the
other may have in finding their gesture, by waiting for affirmation that the object has
been recognised before continuing. Example 4.2 illustrates this phenomenon, in
which a 25 second search for the ‘desk-thing’ precedes a discussion about where it
could be moved. In this case, Sarah begins by asking “Can you see- what I’m
pointing at now?”, and only continues with the task after the ‘desk-thing’ has been
identified. The Furniture World interface does not permit a user to point and move
their view at the same time (due to the limited dimensionality of the input device
used). Unlike co-present interaction, where the other can be monitored in finding and
following a gesture, this means that it is difficult to design the on-going point with
regard to the movements of the other in finding the object.

Such inhibiting ways of acting in the virtual world are compounded by the slow speed
of movement in the CVE, preventing quick glances to find the other and the object.
Given that movement in the world is relatively slow, participants often display that
they are trying to look for the gesture and that the other’s actions are not being
ignored. Examples provided in the analysis of Furniture World data include phrases
such as “hang on, hang on”, “I am looking” or “errr” noises to fill the gap in talk.
These actions would normally be available visually, through the sight of the other’s
body movement. Often these referential sequences can last longer than the very
activities that establishing mutual orientation to the object require – for example, the
length of time it takes for both to see the same object or location can be much longer
than the simple task-related query that follows.

Unfortunately, the additional time involved in establishing mutual orientation is not
the critical concern. This accommodation impairs common practices in working with
objects – an added sequence is inserted into the work activity. Drawing out elements
of the sequence of on-going activities disrupts their flow and organisation.
Participants are unable to engage in the activity at hand, designing the furniture,
whilst assuming that the other can see or quickly find specific objects. Within
Furniture World, participants become explicitly engaged in, and distracted by, the
problem of establishing mutual orientation; it becomes an explicit topic of
collaborative work.
When encouraging another to look at a particular feature of the local environment, members attempt to design their referential actions for each other. In co-present interaction, they are even able to transform the course of a pointing gesture, taking into account the emerging orientations and movements of their co-participants. Individuals routinely and continually configure a pointing gesture by monitoring a co-participant’s ‘response’ (Hindmarsh, 1997). In this CVE, one problem for participants is that when they point to something, they often cannot see their co-participant(s). The Furniture World interface does not allow participants to point at something and simultaneously look elsewhere. Therefore, it is much harder for them to be sensitive to the movements and visual conduct of the others’ avatar. Their ability to design gestures or utterances to indicate an artefact is constrained.

This highlights a more general concern for participants engaged in collaborative work in CVEs. The organisation and co-ordination of much co-present work is facilitated by the ability to ‘monitor’ the activities of others. The narrow field-of-view, however, cuts out the visible features of many of those activities. So, participants’ peripheral awareness of the other is severely constrained. The talk of participants does reveal features of their conduct to others. However, there is a much greater reliance on the talk than in everyday collaborative work. Normally, individuals can rely upon the availability of the others’ visual conduct and see that visual conduct with respect to objects they are working with. Here participants cannot. This leads to coarser practices for organising object-focused interaction. Whereas they would normally be able to talk and simultaneously reveal other ‘information’ via visual conduct, almost all their actions must be revealed through talk.

4.4.4 Principal Issues

This analysis of interaction accomplished through and with this virtual environment, highlights some ways in which CVEs can undermine resources commonly used to co-ordinate collaborative work. Although the Furniture World system does not prevent the task from being completed, it does set up a range of obstacles, outlined below.
• Features of the world are revealed in ‘fragments’, due to the narrow field-of-view. This often separates views of the other’s avatar from relevant objects. As a result, participants are provided with a deficient sense of action. They cannot make sense of talk and activity without seeing (and thus clumsily seeking) the avatar and its relationship to relevant features of the environment;

• Participants are forced to compensate for the problems of interacting in a limited environment by explicitly describing actions and phenomena that are unproblematically available in co-present interaction. In particular, referencing objects and features of the virtual world becomes a topic in and of itself, rather than being absorbed within on-going activities;

• Opportunities for participants to design and co-ordinate their actions for others are reduced. Peripheral awareness of the action and orientation of others is significantly undermined. Even when another’s avatar is visible on-screen, the technology disrupts the resources used to make sense of an individual’s activity.

When interacting in day-to-day life, we assume our perspectives on the world and on an object are irrelevant for the activity at hand (Schutz, 1970). An understanding of what each other can see within the local environment is relied upon. It is presumed that such differences are incidental to the practical tasks in which members engage. Indeed, others’ orientations are exploited in order to initiate new activities or to collaborate on particular tasks (Hindmarsh, 1997, Heath et al., 1995). One of the most notable findings of the MTV mediaspace research was that the technology that mediates co-participants’ interaction disrupts the mundane resources and practices upon which such presuppositions in working with objects are based. Just as with MTV (although in very different ways), the technical constraints imposed by the Furniture World interface render such activities more problematic. This is likely to lead to more intrusive means of monitoring others’ actions and interleaving activities with them.

An implicit thread running through the preceding discussion of problematic interaction has been that that these issues are important for the design of CVE
systems to support synchronous remote working. The following section, then, will take up these issues as regards their specific manifestations and causes in Furniture World, in order to address how CVE system design might take these problems into account. For this purpose, it must be noted that certain technical aspects of the Furniture World system relate to the problems observed in participating in object-focused interaction.

4.5 Implications

These observations lead to the conclusion that certain aspects of the initial Furniture World interface design (a design that resonates with many current desktop CVE interface technologies) should be addressed if CVEs are to provide a basis for distributed collaboration with and around features of virtual environments. Four key limitations seem to have contributed to the noted phenomena. These are: a limited field-of-view; limited or confusing information about the actions of others; clumsy and slow movement of views; and an inability to perform actions concurrently. These four issues form the bases of technical proposals outlined in this thesis, and are thus further underscored in the following sections.

4.5.1 Limited horizontal field-of-view

It is difficult to simultaneously view the source (i.e. avatar) and target (i.e. objects) of object-focused actions such as pointing and looking at things. The ability to engage and act within the virtual environment is displaced by difficulties arising from the limited perceptual capabilities provided by the interface. Confusion about these capabilities arise due to the difference between actual field-of-view for a participant and that assumed by co-participants.

4.5.2 Lack of information about others’ actions

Not all available actions are explicitly represented on avatars and target objects. Thus the ability to discern an action is limited by ‘perceptual’ problems. A strategy of
representing actions without taking these factors into account means a lack of coherence in representation. Where actions are represented on all relevant features of work with objects, it may not be easy to distinguish between them.

### 4.5.3 Clumsy movement

Movement in CVEs may be slow and clumsy. A number of factors influence these problems. Firstly, any movement to attend to relevant features and objects within the world is difficult due to the lack of peripheral vision and awareness stemming from a reduced field-of-view. Secondly, controlling one’s viewpoint through the interface is limited by a lack of intricacy and dimensionality in interface controls provided by desktop systems. Thirdly, the complexity of the required graphics, audio and other media resources in the CVE mean that the performance of system and network updates is relatively poor, and thus movement around the environment is cumbersome.

### 4.5.4 Lack of parallelism for actions

The interface prohibits some combinations of actions from being performed concurrently. Some examples, arising in analysis, which cause problems for participants are: the inability to move one’s view whilst grasping objects; and being unable to design and co-ordinate gestures with the movements of others by simultaneously pointing and looking around. This second issue is compounded by the limited field-of-view of the environment that users have when stationary, and thus their inability to monitor the other’s actions in attending to their reference.

### 4.6 Conclusion

This chapter has described an exploration into object-focused interaction in virtual environments. It has detailed the work in which a set of quasi-experiments were designed and undertaken, and data collected and analysed. The task presented was adapted from studies of mediaspaces in which working with objects was found to be
problematic. The analysis of these data conducted by Hindmarsh (Hindmarsh et al., 1998, Hindmarsh et al., 2000) enabled the author to derive four key elements of system design which produced problems for participants. In general, it was found that it is not possible to separate the affordances of the space as enabling collaborative work from the interface through which users can engage in that common world. In particular, the problems users faced in working with objects were directly related to four aspects of system interface design as reported. These phenomena will be investigated in the next chapter, which follows a course of re-design through a consideration of each interface phenomenon in turn. In this way, it is intended that the CVE interface, and thus the ability of participants in working with virtual objects, will be altered. This change will be based upon the empirical evidence presented. Thus, it will be left to the following chapter to describe an evaluation of the success or failure of these changes; i.e. whether they constitute improvements in the ability of users to undertake distributed collaborative work with virtual objects.