The 'out-of-avatar experience': object-focused collaboration in Second Life

Greg Wadley [1] and Nicolas Ducheneaut [2]
[1] The University of Melbourne, Australia [2] Palo Alto Research Center, U.S.A.
[1] *greg.wadley@unimelb.edu.au* [2] *nicolas@parc.com*

Abstract. Much of our current understanding of collaboration around objects in collaborative virtual environments comes from studies conducted with experimental immersive systems. Now Internet-based desktop virtual worlds (VWs) have become a popular form of 3d environment, and have been proposed for a variety of workplace scenarios. One popular VW, Second Life (SL), allows its users to create and manipulate objects. This provides an opportunity to examine the problems and practices of object-focused collaboration in a current system and compare them to prior results. We studied small groups as they assembled objects in SL under varying conditions. In this paper we discuss the problems they encountered and the techniques they used to overcome them. We present measures of camera movement and verbal reference to objects, and discuss the impact of the UI upon these behaviors. We argue that while well-documented old problems remain very much alive, their manifestation in SL suggests new possibilities for supporting collaboration in 3d spaces. In particular, directly representing users' focus of attention may be more efficient than indirectly representing it via avatar gaze or gestures.

Introduction

Much of our current understanding of collaboration around objects in 3d environments comes from laboratory studies conducted five to ten years ago, with the state of the art defined in books edited by Churchill et al. (2001), Schroeder (2002), and Schroeder and Axelsson (2006). Since then, Internet-based 'desktop' virtual worlds (VWs) have become popular environments in the form of massively-multiplayer games and social worlds. These systems use standard PC

hardware rather than specialized I/O devices: limited hardware being both a boon, allowing VWs to achieve mass popularity, and a detriment to the realism of the user experience. There is renewed interest in using online 3d environments in non-recreational contexts such as online meetings and education, with several workplace-oriented VWs in use or under development.

Of the recreational VWs that are currently popular, one of them, Second Life (SL) is exceptional in that it allows users to build and manipulate the contents of the world using 3d editing tools built into the client (Ondrejka, 2005). The emergence of SL as a popular platform provides an opportunity to update our understanding of collaboration around virtual objects by comparing practice in this system with that observed in prior research.

There is current academic interest in VWs at CSCW and elsewhere (eg Brown and Bell, 2004; Moore et al., 2006; Yee et al., 2006), but most of this work has focused on social interaction and communication rather than the mechanics of collaboration in three dimensions. To bridge this gap, we report in this paper on a study of collaborative building in Second Life. We gathered data in several ways. To facilitate comparison with earlier research we conducted a laboratory study inspired by experiments such as that of Hindmarsh et al. (1998). We asked groups of two or three participants to collaborate on building tasks, and recorded their screen video and conversation for later analysis. With each group we discussed the problems they faced, how they solved them, and their thoughts on the user interface. We conducted quantitative measures of participants' deictic verbal references, and their use of SL's detachable camera, an interesting UI feature that potentially impacts collaboration. Finally we discussed themes that arose with other expert users discovered in-world and in the SL forum. We did not undertake traditional online ethnography, because it would be rare to chance upon instances of collaborative building and impossible to see the users' view of the virtual scene portrayed on their screens. However one author was an intensive user of SL during the study and was able to discuss and observe in-world practice in order to ground the laboratory observations.

Collaborative virtual environments

The primary communicative affordance of virtual worlds is a simulated 3-dimensional space in which users are represented to each other as avatars. This allows the simulation of some aspects of offline interaction such as how people position and orient their avatars and how they refer to objects. Research has investigated the mechanics of 'simulated face-to-face' for several years. One thread of inquiry focuses on the similarity and differences between simulated and physical spaces. Yee at al. (2007), for example, demonstrated that Second Life avatars obey real-life proxemic rules. Moore et al. (2006) emphasized that in

current avatar systems, most user actions are not publicly accountable and therefore hinder the micro-coordination of activities.

VWs should be suited to supporting collaborative work in problem domains that are inherently three-dimensional, such as design, repair, and medicine. For example, a technician fixing a machine might converse with remote experts while referring to a 3d representation of the machine, with all parties able to manipulate the model. A significant body of CVE research has focused on collaboration around objects. A central issue has been how one user can deduce another user's point-of-view in order to reference objects deictically: a problem compounded by the limited gestural abilities of current avatars. Hindmarsh et al. (2001) found that groups struggled to achieve common reference to objects even when able to 'point', since users could not always see both the pointing arm and the referent due to the narrow horizontal field of view of desktop systems. Gestures also forced users to spend too much time 'driving the avatar'. Pinho et al. (2002) studied methods of allocating degrees of freedom so that one user moved an object along a ray while another rotated it. Thus collaborators at different positions could combine their viewpoints to place an object efficiently.

Research in shared-video systems has shed light on collaboration, though avatars are not used and the choice of vistas is usually limited to 'scene' or 'head-mounted' cameras. Kraut et al. (2002 and related work) had helper-worker pairs complete a screen-based jigsaw-puzzle. In this arrangement the worker manipulates objects while the helper can offer only verbal assistance. Sharing the scene view, but not the head-mounted views, improved performance, especially when the task was complex and the objects difficult to describe verbally. Goebbels et al. (2003) had pairs collaboratively manipulate a virtual object with the assistance of haptic control and video-conferencing, finding that users spent more time looking at the object than each other except while resolving misunderstandings, and that voice quality was more critical than video.

While many experimental CVEs allowed subjects to communicate by voice, desktop VWs have until recently offered only typed text for linguistic communication. The mechanics of textual turns-at-talk in a VW was studied by Brown and Bell (2004). Text communication during object-focused collaboration was examined in the VW 'ActiveWorlds' by Herring et al. (2003), who found that novices tried to refer deictically to objects, but resorted increasingly to describing them by name. The recent addition of voice-over-IP to systems such as Second Life contributes to making them more 'lifelike' and indeed, recent research shows that it has considerable, though situation-dependent, benefits for coordination of groups in MMORPGs (Williams at al., 2007; Wadley et al., 2007). While VW users engaged in identity-play often prefer to communicate by text, we took it as a given that voice would be used by collaborating workers, and allowed our participants to speak.

Second Life

Since its inception in 2003, Second Life has grown into one of the most popular commercial VWs, with 1.5 million registered users. The SL client uses a standard PC screen for output and keyboard and mouse for input. Since 2007 a voice channel has been included for user communication. SL user accounts may be paid or free: the approximately 80,000 users who pay can own virtual land and build on it. The right to build permanent structures (which are collections of simple shapes called 'prims') is the main advantage conferred by a paid account, so this is a reasonable estimate of how many users are building content. A number of authors have commented on SL's potential as a tool for CSCW: Van Nederveen (2007) proposed it be used for collaborative architectural design, while Rosenman et al. (2006) tested a design system in which SL was supplemented with tools such as a 2D sketch-pad.

The SL client allows users to choose between first-person (through the avatar's eyes) and third-person (from behind the avatar) views. Unusually, it allows users to move their camera independently of the position and orientation of their avatar, by anchoring it to an object in the local scene. This technique allows a user to gain multiple perspectives more quickly than is possible by walking an avatar around an object: thus it is commonly used for object-related activity such as building and looking at other users' creations. Unlike most VWs which support limited camera movement near the avatar, the SL camera can be moved over a wide area, oriented in any direction including up and down, zoomed a long way in and out, and unlike avatars, moved through objects. While users' avatars are publicly visible, their camera positions are not (Irani et al., 2008). The ability to decouple one's camera from one's avatar is similar to techniques suggested by Hindmarsh et al. (2001) and Bailensen et al. (2006).

At any given moment an SL user is viewing the virtual landscape from either their avatar location or the location to which they have moved their camera: thus an SL user's presence is divided between two different locations. We call these modes 'in-avatar' and 'in-camera' to emphasize that while avatar locations are visible, camera locations are private. These are illustrated in figure 1. Since the objects that a user can interact with are those that are currently visible to them, rather than those that are in the vicinity of their avatar, it is their private camera-location rather than their public avatar-location that defines their focus of attention.

The detachable camera feature means that there is no reliable relationship between what an SL user can see and what their avatar appears to be looking at. While the feature is useful while editing, prior research (e.g. Hindmarsh et al., 1998) suggests that when users cannot deduce each others' vistas, their ability to collaborate is lessened.

Second Life provides no specific support for collaborative building beyond the ability to visualize one's collaborators in the shared workspace and to communicate using text or voice. Users can display a map of the local area showing avatar locations and the outlines of buildings; however the map is not sufficiently detailed to assist with object manipulation and tends to be used only for coarse navigation. There is no analogy to peripheral vision in the SL display. Limited pointing is possible: when a user is editing an object, their avatar's arm reaches toward the object and a dotted line (the 'selection bar') connects arm and object. This is similar to the line provided by Hindmarsh et al., the rationale for which was that moving an object at a distance represents projection beyond the avatar. In SL this line provides a rough indication of which object a user is editing; however if object and avatar are sufficiently far apart it is difficult for others to follow the selection bar between them. The editing user sees a highlight on the object, but this is not visible to others.

Given the current popularity of virtual worlds and interest in them as platforms for CSCW, we perceived an opportunity to update the understanding of collaboration in 3d by observing it in a current VW. We chose Second Life because it has a significant user base, a focus on social interaction, and offers object manipulation via in-built tools.

Methods

We used methods based on Hindmarsh et al. (2001) and Kraut et al. (2002) to conduct a 'quasi-experiment' in the sense of Hindmarsh. Participants logged into Second Life in groups of two or three and collaborated on building tasks. Groups were co-present in our lab and arranged so they could hear but not see each other. We observed the groups and recorded their screen output and voice conversation for later analysis. Building sessions were followed by focus-group discussions.

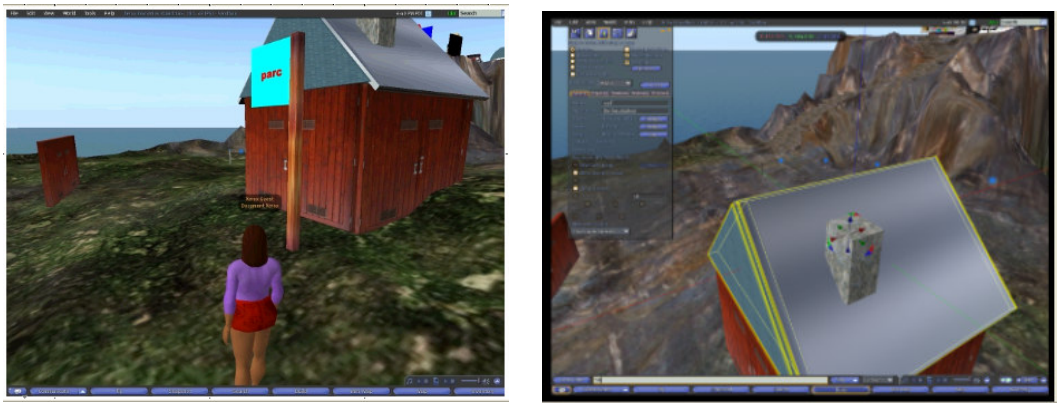

Figure 1: Two users' vistas, captured simultaneously, just before completing the House task. The user on the left is 'in-avatar'. The user on the right is 'in-camera' and editing the roof.

Each complete session was designed to last an hour and a half, although some enthusiastic groups worked over time, which we allowed. We observed ten groups; a total of 22 participants. Their ages ranged from 20 to 50. Half were male. Some knew each other before undertaking the task, while other groups were meeting for the first time. We chose participants with a broad range of SL experience, reasoning that while expert users would demonstrate cutting-edge practice, it was important also to understand novice use. Two participants' recordings were excluded due to technical faults: for statistical tests this left a cohort of four experienced builders (recruited via Craigslist and the SL forum), ten users with experience of other VWs or modeling tools, and six novices.

After spending a few minutes familiarizing themselves with each other and the lab setup, group members undertook two tasks. The first adapted the 'furniture world' task of Hindmarsh and the 'jigsaw puzzle' task of Kraut (though in 3d). The researchers provided a set of virtual objects including four walls, two roof parts, two gables, a flag, flagpole and chimney. The group had to assemble these into a house. Some of the objects, such as the flagpole, were unique, meaning they could be referred to by name. Others, such as the walls, were indistinguishable from each other, requiring spatial reference to identify them. Each group member received a screen-shot of how the house should look when complete. The 'House task' was unstructured, allowing groups to discover their preferred method of collaboration, naturalistic in that dwellings are a popular building project in SL, and relevant to remote repair in being a complex object made of smaller parts.

In pilot studies we found that some users chose to collaborate with minimal communication (more on this later), and so we provided a second task, the 'Garden task', using the helper-worker arrangement of Kraut and designed to force closer collaboration. One group member was designated the helper, while the other(s) were workers. Helpers could not use the building tools themselves. They received a screen shot showing the house now surrounded by extra objects such as garden furniture and a fence, and had to direct their worker(s) to build this scene.

After the two tasks were complete we conducted a focus-group in which we asked participants about their experience, the problems they encountered, how they solved them, and whether the UI could be enhanced to better support collaboration. Focus-groups were semi-structured to allow exploration of themes.

We used the screen recordings to conduct two quantitative analyses of group performances, counting 'salient events' (Schroeder et al., 2006). We measured the proportion of time users spent with their camera decoupled from their avatar, and we categorized verbal references to objects by their linguistic form. We compared these measures across experience levels and tasks to discover patterns of use.

As a final check we discussed emergent themes with a range of experienced users discovered on the SL discussion forum and at building classes and competitions.

Results

Our participants displayed a broad range of building and communication styles, illustrating the variation that general-purpose systems need to support. They encountered communication problems of the kind discovered in earlier research, but worked around them by experimenting with different referential practices (as in Hindmarsh et al., 1998) and rarely became mired in problems of reference.

The following exchange illustrates the kinds of problems that arose and the variety of reference techniques attempted. This group of three is deciding how to position four walls to form the base of their house. The walls look identical and are currently positioned randomly around the vicinity.

A. So are you putting the walls together?

B. I'm moving one wall … a third wall, towards the other two … the one that's tilting. *[B marks a wall by changing its orientation in a fashion visible to her team-mates.]*

C. Oh, that was you! *[A and C now know which wall B is editing.]*

A. Why don't you turn that over, and I'll move the other wall? *['That' refers to the tilting wall, and 'other' to the fourth wall not discussed yet.]*

C. Are you moving the one on the lower leftmost of the walls?

A. Well .. your left? *[A and B laugh, because C attempted a spatial reference which A and B cannot decode.]*

A. I'm going to move the one that I'm standing right next to. *[A doesn't attempt to correct C's attempt at deixis, but instead moves her avatar beside a wall to mark it to the others.]*

A. The one that's I guess kind of closest to [C] ... *[A uses the position of C's avatar as a reference point]* … why don't we leave that one still, and then we can put the other three around it?

C. Mine's above the ground. *[C refers to the wall closest to his avatar as 'his' wall.]*

A. That's fine I think. Why don't we just leave that one and put the other three around it?

The group tried different referential methods until one worked and they could proceed with the task. Participants often used their avatars to mark positions, such as this exchange from a helper-worker pair performing the Garden task:

H: You see this table I'm standing next to? Don't move this one – this one stays in place.

W: Yep. So why don't you just move where you want the others.

H: Yeah. *[walks to a different spot]* The other one is going to go here. In front of me.

W: Right, hang on … *[W moves the table]*

H: And the last one is on the other side ... *[walks around the house]*… Just about here.

Another pair's exchange illustrated several techniques: pointing with the edit bar, marking places with avatars, and verbal description of an object:

X: Is there a way to point? What's the thing you thought was the flagpole?

Y: Hang on, let me just walk into it. *[walks to the flagpole]* See this thing that's right near my hand? *[He is currently editing a wall – a different object - so his arm is in the air.]*

X: Which hand?

Y: Right in front of me. Can I point at it? *[He places the flagpole in edit mode so as to point directly to it.]* There we go. Why don't I move it. If you're watching it, I'm moving it back and forwards now. Can you see an object that keeps moving left and right?

X: Yes that's the flagpole isn't it? […] You just walked past a cement block. Are we supposed to do something with that?

Y: I think that's the chimney.

Experience made a clear difference to participants' ability to collaborate. Yet all groups were able to complete their tasks, albeit at different speeds and with varying quality.

Use of the detachable camera

The ability to rapidly gain multiple perspectives of an object by detaching one's camera from one's avatar is not available in most VWs. While this feature supports efficient building by individuals, it also breaks the relationship between an avatar's orientation and what the avatar's owner can see. Since prior research indicated that deducing collaborators' viewpoints was a significant problem in CVEs, we were interested in how often users detached their cameras while working in SL and whether this affected either their ability to communicate or their experience of virtual embodiment.

Referential problems caused by the mismatch between avatar and camera viewpoints are illustrated by the following exchange. This pair performed their tasks well, but were plagued by an on-going misunderstanding over viewpoints, because one (here called 'A') stayed mostly in-avatar while the other ('C') had his camera zoomed out and seemed to ignore the avatars. Here they have assembled four walls and are about to place the gables. Their avatars stand at opposite ends of the house, however C's camera is near A's avatar, so that unbeknownst to A they are viewing from the same side of the house.

C: Let's place those triangle things. *[the two gables]*

A: Where are those? Oh, the triangle things are around the front aren't they? *[It is not clear which end of the house is the front.]* I'll place the one on my side if you place the one on the other side.

A: I don't know whether I've selected the same one as you. I'm selecting the one that's further from the house.

C: Ok, do you see one moving? I selected one that I just raised up.

A: Yes I see that one, ok good. I'll pick a different one then. Oh you're putting it on that edge?

C: I put it on the nearest spot I could find. *[They both intended the end nearest their views.]*

A: Where's your character? *[he means 'avatar']* Oh ok, I see where your character is. I tell you what, can you put the gable on the house section closest to you, and I'll move the one that's closest to me? Unless you want to finish placing the one that you had. *[A is still using avatar-relative reference while C assumes he means relative to camera.]*

C: Does it matter? I'm maneuvering the one that I had.

A: Ok, I can move the other one I think. I'll just walk around so I can see it better. *[He walks his avatar to the other end of the house, where C's avatar, though not his camera, is placed. While walking he apologizes for bumping into C's avatar, though C was unaware of it.]*

At one point during the Garden task this pair tried the house's frame of reference, and then spatial deixis, before being successful with avatar marking:

C: If you're facing the front of the house, you need one table in the front of the house with two chairs, one to the left of the house with four chairs, and one behind the house with two chairs.

A: So we're going to treat me as facing the house right now? Do you want to see where I am?

C: Um, I see where you're facing.

A: I tell you what, can you walk your avatar to what you're calling the front of the house?

[C goes back into avatar and walks to the front]

A: Ok. So you're currently at the front of the house?

C: Yes I'm facing the front. *[A proceeds to place the furniture.]*

Some experienced MMORPG users rarely detached camera from avatar, suggesting that extensive gaming experience may make disembodied viewing in 3d systems feel unnatural. Experienced SL users said that maintaining both an avatar and a camera location did not bother them. On being questioned about "being in two places at once", most said that this had never occurred to them. When asked, "How would you describe your location right now?", experts usually chose their avatar rather than camera location. One felt that the detached camera was simply a tool, and that while using it he continued to equate his avatar with himself. Conversely, others felt that, while they were building, their avatar was irrelevant and even got in the way. One expert said that as a beginner she had identified with her avatar, but that over time she had begun to experience SL more as a building tool than a virtual reality. But her equally experienced building partner felt that his avatar mattered because avatar locations are how people find each other, and because: "something can happen to your avatar. You can get pushed or shot. Nothing can happen to the camera: it's just a view of a picture."

We calculated how much time each participant spent 'in-avatar' and 'in-camera'. Averaged over all participants, about half of task time was spent in each mode. Suspecting that this correlated with experience, we classified participants and compared their camera use. Group A (n=4) had significant SL experience, while group B (n=10) were competent users and group C (n=6) were novices.

ANOVA showed expertise to have a significant effect on camera use ($F(2,33)=8.93$, $p<.001$), with expert SL users, as expected, more inclined to detach their camera (figure 2).

Task (House vs. Garden) was not a significant factor in camera use, and there was no interaction between expertise and task, however this may represent two effects canceling each other out. In the Garden task, helpers often used their avatars to mark locations, and were in-camera less often. But some workers spent more time in-camera during this task, perhaps due to increased familiarity with the UI.

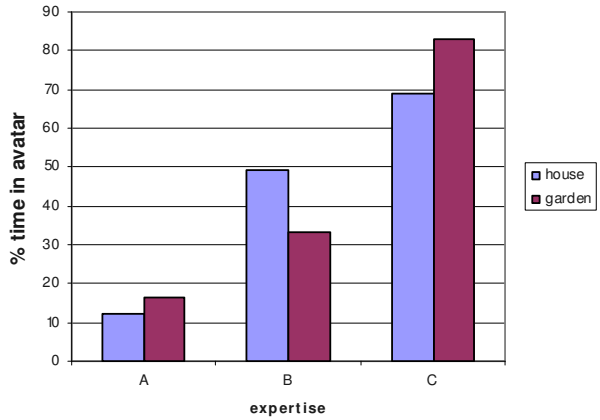

Figure 2: Time spent 'in avatar'.

Verbal reference to objects and places

Natural language offers several ways to refer to objects and places. CVEs afford spatial deixis because users are embodied at particular locations and orientations which are visible to other users. If a user is looking through the eyes of his avatar, then a deictic reference relative to the avatar is also relative to the user's vista, and should be understandable by him. Therefore a detachable camera should make deixis less reliable.

We examined our participants' use of verbal reference, counting references to objects and locations and categorizing them according to the frame of reference used. A number of categorizations are available: we used that of Levinson (1996), who recognizes relative, intrinsic, and absolute frames. A 'relative' reference involves deixis from the speaker's or receiver's point of view and is the key form of interest here. Some objects have their own 'intrinsic' frames of reference, for example houses may have an obvious front and rear to which other locations can be compared. Finally locations can be relative to an 'absolute' frame of reference such as compass points or a prominent object in the distance. We added a fourth category, 'reference by name or property' to count references such as "the brown

rectangle". Other researchers have used slightly different schemes: for example Herring et al. (2003) categorized references as "deictic", "fixed unique" and "fixed non-unique". Their 'deictic' category corresponds to our 'relative' category, while their 'fixed unique' corresponds to our 'name or property' category. We did not count how many of the instances of deixis were successfully interpreted, as this is not always clear to an observer.

Although both Cartesian (x y z) and cardinal (north south east west) frames are available in Second Life, they were rarely used in our study. On only two occasions participants made use of Cartesian coordinates to describe locations. Only one group used the cardinal frame. Although landmarks were visible, on only one occasion did a participant use one for spatial reference, describing a wall as "the side closest to the sea". The 'absolute' category is excluded from the graph below.

Figure 3 illustrates the relative frequency of these forms of reference. These were consistent across groups ($F(2, 27) = 6.37$, $p<0.01$). Neither task nor expertise level were significant factors, though expertise affected the overall number of references, with experienced participants making more. While this appears to contradict a finding of Kraut et al. (2002), it suggests that those who were better able to handle SL's particular style of representation were more comfortable communicating about location. Some novices seemed to be so focused on grappling with the building tools that they neglected to communicate with team-mates.

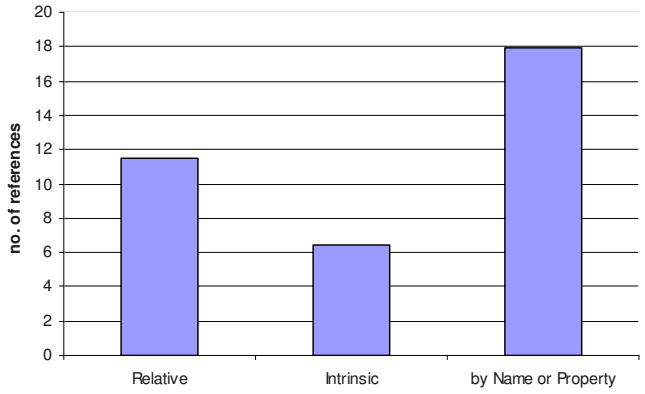

Figure 3: Frames of reference used to refer to objects in speech

It is noteworthy that even experts used deixis relative to a collaborator's avatar despite knowing the collaborator was probably 'in-camera'. This was not a problem for experienced users, who seemed able to use their avatar's frame of reference even while their visual focus was elsewhere. Experts interpreted a reference such as "to your left" in the only reasonable way, meaning "to the left of your avatar". (Sometimes reference was explicitly relative to "where your avatar

is standing".) If their avatar was out of view, they moved their camera to bring it into view, pressed Escape to return their camera to their avatar, or asked for more explanation. Expert informants confirmed that they were able to interpret avatar-relative deixis, the only problem being that, if they had to return to their avatar, they lost their camera position and would have to find it again. In Second Life, avatars are permanent cursors that mark a location that is easy to return to, while camera locations are invisible and impermanent.

Occasionally, as illustrated in the exchanges above, novice participants forgot that their avatar but not their camera position was visible. But in most cases, deducing collaborators' vistas did not seem to be a major concern. No participant ever asked a team-mate, before using deixis, whether they had detached their camera. We asked participants whether they would like an extra screen displaying their collaborator's vista, but none thought this would be useful, though some novice users proposed a 'see what I see' feature that could transmit their vista to a colleague when required.

Non-verbal reference

Referencing an object by pointing to it (selecting it for editing) was only rarely observed in our study, though one expert claimed that this technique was used "all the time" in SL. Some participants remarked that although they could see a colleague's avatar point while they were editing, the highlight on the object was only visible to the editor, which reduced the usefulness of pointing as a means of collaborative reference. Avatar gestures other than pointing were never used by participants. Moving one's avatar to stand beside an object however was frequently used. Some participants jiggled objects back and forth or moved them to mark them for collaborators, as illustrated above and in the following exchange:

A: Are you rotating one of your walls?

B: Yes, is it rotating on your screen?

A: I'm rotating another wall. Yep, I just saw yours rotate.

One expert changed the colour of a wall to mark it: however she said this was not a common technique because it is usually hard to restore the original texture.

Some experts said that it was common practice in SL to create a prim to mark a location, to be deleted later after use. However no-one did this in our trials. It is possible that the technique did not occur to novices, and that experts did not find our tasks sufficiently difficult to require it.

Problems editing in 3d

Although not all participants were regular users of 3d environments, none of them had difficulty navigating their avatar around the space. However those unused to SL found it difficult to manipulate objects, which are constrained to move along orthogonal axes. Novices found translating and rotating objects frustrating, some stating they would prefer to simply drag an object from one location to another rather than execute separate moves along each axis. One participant typed destination coordinates rather than use the mouse. An environment for supporting remote repair might require more natural movement of objects.

Prior experience with modeling packages substituted for experience with SL's building tools. One such participant felt that SL was "Blender grafted onto a game". Experienced users sometimes performed an entire build 'in-camera' and on completion walked around it 'in-avatar', suggesting that they experienced building as a distinct activity within the VW.

Novices seemed to assume that objects were subject to collision detection and gravity, though they clearly were not. Experts made use of the fact that objects could hang in the air and pass through other objects, and had a natural orientation along the world's axes. Experts did not express interest in improving the building UI, though several complained that the permissions system was not conducive to collaboration.

Group organization and division of labor

Groups performing the House task were free to organize themselves any way they wished. Different groups divided their labor at different levels of abstraction. A striking proportion chose one method, which was to decompose the house into 'base' and 'roof' sub-assemblies, to be completed separately by individuals and joined in the final stage. This appeared to be a strategy for reducing the need for close collaboration. Most research on collaboration around virtual objects has focused on closely-coupled collaboration, yet we rarely observed participants choosing to work simultaneously on the same prim.

Expert informants reported that although team building is common in SL, especially on large, complex projects, closely-coupled collaboration at the level of individual prims is rare. Experts find it more efficient to decompose a project into sub-tasks, allowing specialization and schedule independence. For example, one user might create a building's skin, while another creates its furniture, another builds walkways and a fourth applies textures. Often these components are not even built at the same site, but are created on the individuals' own land and moved into place in the final phase of building.

Apart from reducing the need for coordination, another explanation for SL users' disinclination to collaborate closely might be that is only necessary in systems which tie vistas to avatars. In those, a user cannot easily obtain multiple

viewpoints of an object and might benefit from feedback from colleagues located at different viewing angles, in the fashion described by Roberts et al. (2006). SL's camera reduces the need for this, and we observed only two occurrences.

Close collaboration in SL is also made less effective by the way visual feedback on object movement is shared. While a user is dragging an object they receive visual feedback at all positions along the object's path. However collaborators only see the end point of the movement. The is probably designed to reduce the number of scene-update messages sent over the network, however it makes it harder for one user to guide another user's placement of an object.

Discussion

Comparing old and new

Our purpose was to observe how the problems and practices of collaboration around virtual objects have evolved since CVEs emerged from research labs to become mainstream technologies. We chose Second Life as our study system because it is a popular VW that allows users to manipulate objects. We conducted a semi-naturalistic exploratory study rather than a formal experiment, but exploited the lab setting to conduct two quantitative analyses.

Many of our findings can be directly compared to previous work. Hindmarsh et al. experimented with extended pointing, peripheral vision and a plan view. SL provides such an extended pointing via the 'selection beam', but we heard mixed views on its usefulness. Instead, some users said it would be more useful if the object highlight seen by an editor was also visible to others, so that knowing who is working on what was more transparent. In other words, there seems to be little need to tie selection feedback to the avatar: changing the visual appearance of the selected object is enough. This seems to contradict (for object-focused work) Moore et al.'s argument that avatar systems should be made richer to enable tighter coordination. In fact it could be argued that representing user's bodies is superfluous during this form of collaboration. Certainly some participants forgot about their avatars, or used them as object-marking cursors.

Related to this, Hindmarsh et al. reported that stylized gestures were not useful for collaboration around objects, and this is supported by our study, which observed no use of SL's pre-programmed gestures. With regard to peripheral vision, it is not provided in SL, and most participants felt that extra screens would be a burden. SL offers a plan view (the map) but none of our participants used it, probably because it displays insufficient detail about objects to be useful.

Hindmarsh et al. found that avatars often gave the wrong impression of what their users could see. SL's detachable camera would seem to make this problem intractable. Yet our participants were able to communicate successfully, if at

times slowly, about objects and locations. Moreover they stated that an extra window showing their collaborator's camera view would be superfluous. This accords with the finding by Fussell et al. (2003) that collaborators preferred seeing the shared workspace to looking through the head-mounted camera of a collaborator (see also Hindmarsh et al., 2001: p134-5). Herring et al. (2003) found that novice users of ActiveWorlds ceased attempting spatial deixis when they found their colleagues could not dereference it. Our participants were more successful with deixis, which may be due to better graphical representation in SL, or the availability of voice, which is better for quickly resolving ambiguity (Löber et al., 2006). Alternatively it may reflect an increased familiarity over time with the representational style of VWs.

The 'out-of-avatar experience'

Second Life is unusual among avatar-based systems in allowing users to detach their camera from their avatar. This feature trades the benefit of rapid acquisition of multiple viewpoints against the drawback of making some deictic references irresolvable and the foci of activity invisible. SL users effectively have two locations, their (public) avatar position and their (private) camera position, a situation which can foster deception (cf. Irani et al. 2008) and conceivably, a reduced sense of embodiment in the avatar.

Experts seem to maintain a sense of where their avatar is facing, and use this to dereference their collaborators' deictic references. When this is impossible they simply return their camera to their avatar to translate the reference. This costs only the time taken to subsequently return to the camera position, and does not seem to significantly impact users, except for a temporary loss of 'state' – they have to remember where their camera was and manually re-establish their view. This difficulty could be easily addressed by adding a 'toggle' to switch back and forth between the two perspectives. Supporting such an ability to smoothly transition between various states and viewpoints might be a fruitful avenue to explore by future VW designers, especially if they intend to support collaborative activities.

It is possible, using scripting, to provide the location and gaze direction of one's camera to other users. This can enable 'you see what I see'. This feature is not available in SL's standard UI, but has been implemented by one entrepreneur as a software add-on which is available for purchase. We obtained this but found it had limited utility. Only one user could send their camera position, and their collaborators could only receive. Switching between one's own camera and the sender's was slow and awkward. Only the camera position was transmitted, not other screen visuals such as editing highlights, thus masking much of what the sender was doing (cf Irani et al., 2008; Moore et al., 2006). One of our participants, an experienced SL builder, was familiar with this add-on but did not use it in his work.

One might expect a detachable camera to diminish the relevance of avatar location. Despite this, researchers have found that SL avatars obey physical-world proxemic norms of inter-personal distance and eye-gaze (Yee et al., 2007). For proxemics to work, users must perceive each other as having a definite location and orientation. In our study, participants often 'parked' their avatar while building, moving it to a socially appropriate position only when interacting with other users. In one session, a participant stayed in-camera except when a new user appeared nearby, whereupon he went in-avatar and walked over to them. It seems that SL users regard their avatar as a mediator of social interaction which can be ignored while editing objects. However in collaborative building, which is simultaneously object-focused and social, these two attitudes contradict.

People communicating in SL often place their avatars face-to-face, even if they also detach their camera. Arguably it is a form of 'perception management' to maintain proxemic norms with one's public embodiment while one's private focus is elsewhere. A user who maintains a conversational orientation while moving their visual focus must be aware that other users may be doing the same thing. A solution to this ambiguity might be to display camera positions on screen. An option in the SL client's 'Advanced' menu allows one to see the locations of nearby cameras; however these are not labeled with avatar names - thus users can know that they are being examined, but not by whom. Cameras are often moved so quickly that to keep track of them is cognitively difficult and would require increased network traffic.

It is sometimes argued that generations who have grown up with 3d videogames will readily adapt to collaboration in virtual environments. But participants with game but not building experience reported difficulties using the building UI, which was perceived as being unaligned with the avatar UI.

Articulating collaboration

Using the terminology of Schmidt and Simone (1996) we can analyze SL as a CSCW system for creating the virtual world's contents. During collaborative building the common field of work is objects and the virtual space within which they reside. SL provides no specific mechanism for articulating work beyond its regular communication tools.

We were surprised to find that the style of collaboration we have referred to as 'closely-coupled', in which two users work on the same primitive object at the same time, was rarely performed. On the contrary, the first impulse of many groups was to modularize their task. Users seem to have devised organizational processes that preclude the need for fine-grained collaboration, and there may be several reasons for this. One is that the articulation work required for close collaboration in a 3d environment represents too high a load. Another is that it might be easier to gain multiple viewpoints by moving one's camera than

receiving verbal feedback from a collaborator. A third is that SL's permission system forces one to explicitly change a default setting in order to allow collaborators to edit objects one has created. A division of labor that involves individual construction of separate modules seems to better leverage the benefits of having more than one person involved in the task.

We implemented a helper-worker task in order to encourage more communication about objects and location. It is noteworthy that in other research where participants worked closely around individual objects, close collaboration was also 'forced'. For example, Pinho et al. (2002) required one user to move an object which was distal to their avatar, while another user closer to the object guided its placement. Roberts et al. (2006) implemented gravity so that two users were required to lift objects while a third joined them together. By contrast SL allows users to rapidly acquire a variety of viewpoints and does not implement gravity by default, so that objects can be lifted by a single user and will stay in place while the user works on other objects.

It may be that VW users will only collaborate closely around objects if physical-world constraints such as gravity and strict embodiment of camera within avatar are reintroduced. But these constraints are not necessary in a virtual environment. Rather than insisting on mimicking physical reality to encourage tightly-coupled interaction, it seems more productive to embrace a VWs' 'unrealistic' properties. As an example of this dichotomy, we would cite again the suggestion by Hindmarsh et al. that users should be made aware of their collaborator's viewpoint (mimicking the accountability of actions from the physical world, cf. Moore et al., 2006) and compare it to a possibility suggested by our study, namely, that users should be able to switch at will between several viewpoints. The 'unrealistic' ubiquity we propose might turn out to be more productive than insisting on reproducing the more familiar, but ultimately more limiting, 'one body – one view' paradigm.

Conclusion

The appearance of Second Life, a popular Internet-based virtual world that allows users to edit its contents, provides an opportunity to update our understanding of collaboration around virtual objects. It is interesting to note that problems identified more than ten years ago in experimental CVEs are still prevalent in a 'mass market' environment like SL. In particular, difficulties with the UI (especially the lack of transparency and feedback about a collaborator's actions) can lead to a tendency to partition collaborative building into isolated, individual sub-tasks that can be completed in parallel and assembled only at the very end. But our users did not react positively to suggestions from past research that could have made tightly coupled collaboration easier. Shared viewpoints, for instance,

were considered to be cumbersome and unnecessary; avatar gestures for pointing were rarely used; etc.

Instead, our data suggest another avenue for supporting collaboration in VWs: 'decoupling' them from physical reality to leverage their unique properties. For instance, while users frequently did not use their avatar like we would use our bodies in physical-world collaboration (by pointing, orienting to the object, etc.), they asked instead for the object itself to be more accountable: for instance, making visible the fact that it is selected by someone else, relaying movement as it happens rather than only at the end of a sequence of modifications, etc. It is technically easy to make the state of an object visible in VWs and yet they have remained for now silent partners in collaborative tasks. We argue that a lot could be gained by thinking about how to make objects, rather than avatars, richer and more interactive.

In a related fashion, the separation between 'in-avatar' and 'in-camera' modes did not introduce as many coordination problems as one might have expected – and in fact, experts used the two modes to conveniently handle relative positioning and object manipulation synchronously. Rather than trying to reconstruct a collaborator's ever-changing viewpoint, it might be more productive to accept that users can literally be in several places at once and instead make the transition between various modes more straightforward. As we saw, users will then switch to whatever viewpoint is necessary for the task at hand without much misunderstanding of what their collaborator says.

The fact that, while building, many users 'parked' their avatar and concentrated on the objects instead suggests that while it is necessary to represent users' focus of attention for collaboration around objects, it is less necessary to represent their bodies. Indicating attention directly via a shared cursor or by highlighting objects could be more efficient than indirectly representing it via an avatar's eye-gaze or gestures.

As collaborative VWs become more mainstream it will be interesting to see whether the practices we observed are truly widespread. In the meantime, we hope this study will inspire VW designers to explore more 'unrealistic' interfaces to better support collaboration in 3d spaces.

Acknowledgements

We would like to thank our study participants, and especially Don Wen, Diane Schiano and Mike Roberts at PARC, Jonas Karlsson at Xerox, and Martin Gibbs at the University of Melbourne for their help with this study.

References

Bailenson, J., Yee, N. and Merget, D. (2006) 'The Effect of Behavioral Realism and Form Realism of Real-Time Avatar Faces on Verbal Disclosure, Nonverbal Disclosure, Emotion Recognition, and Copresence in Dyadic Interaction', *Presence: Teleoperators and Virtual Environments*, vol. 15, no. 4, August 2006, pp. 359-372.

Brown, B. and Bell, M. (2004) 'CSCW at play: 'There' as a collaborative virtual environment', in *Proceedings of the 2004 ACM conference on Computer supported cooperative work*. Chicago, pp. 350-359.

Boellstorff, T. (2008) *Coming of Age in Second Life*, Princeton University Press, Princeton.

Churchill, E. F., Snowdon, D. N. and Munro, A. J. (eds.). (2001) *Collaborative virtual environments: digital places and spaces for interaction*, Springer, London.

Fussell, S. R., Setlock, L. D. and Kraut, R. E. (2003) 'Effects of Head-Mounted and Scene-Oriented Video Systems on Remote Collaboration on Physical Tasks', in *Proceedings of the SIGCHI conference on Human factors in computing systems*, Ft Lauderdale, pp. 513-520..

Goebbels, G., Lalioti, V. and Göbel, M. (2003) 'Design and evaluation of team work in distributed collaborative virtual environments', in *Proceedings of the ACM symposium on Virtual reality software and technology,* Osaka, pp. 231-238.

Herring, S. C., Borner, K. and Swan, M. B. (2003), 'When rich media are opaque: Spatial reference in a 3-D virtual world.' Invited talk, Microsoft Research, Redmond.

Hindmarsh, J., Fraser, M., Heath, C., Benford, S. and Greenhalgh, C. (1998) 'Fragmented Interaction: Establishing Mutual Orientation in Virtual Environments', In *Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work*, New York, pp. 217–226.

Hindmarsh, J., Fraser, M., Heath, C. and Benford, S. (2001) 'Virtually missing the point: configuring CVEs for object-focused interaction', in Churchill, E. F., Snowdon, D. N. and Munro, A. J. (eds.): *Collaborative Virtual Environments: Digital places and spaces for interaction*, Springer, London, pp. 115-139.

Irani, L. C., Hayes, G. R. and Dourish, P. (2008) 'Situated practices of looking: visual practice in an online world', in *Proceedings of the ACM 2008 conference on Computer Supported Cooperative Work*, San Diego, pp. 187-196.

Kraut, R. K., Gergle, D. and Fussell, S. R. (2002) 'The Use of Visual Information in Shared Visual Spaces: Informing the Development of Virtual Co-Presence', in *Proceedings of the 2002 ACM conference on Computer supported cooperative work (CSCW 02)*, New Orleans, pp. 31-40.

Levinson, S. C. (1996) 'Frames of reference and Molyneux's question: cross-linguistic evidence', in Bloom, P., Peterson, M.A., Nadel, L. and Garrett, M.F. (eds.) *Language and Space*, MIT Press, Cambridge.

Löber, A., Grimm, S. and Schwabe, G. (2006) 'Audio vs chat: Can media speed explain the differences in productivity?', in *Proceedings of the 14th European Conference on Information Systems*, Goteborg, pp. 2172-2183.

Moore, R., Ducheneaut, N. and Nickell, E. (2006) 'Doing Virtually Nothing: Awareness and Accountability in Massively Multiplayer Online Worlds', *Computer Supported Cooperative Work*, vol. 16, no. 3 pp. 265-305.

Ondrejka, C. R. (2005) 'Escaping the Gilded Cage: User Created Content and Building the Metaverse', *New York Law School Law Review*, vol. 49, no. 1, pp. 81-101.

Pinho, M. S., Bowman, D. A. and Freitas, C. M. D. S. (2002) 'Cooperative Object Manipulation in Immersive Virtual Environments: Framework and Techniques', in *Proceedings of the ACM symposium on Virtual reality software and technology (VRST 02)*, Hong Kong, pp. 171-178.

Roberts, D., Wolff, R. and Otto, O. (2006) 'The Impact of Display System and Embodiment on Closely Coupled Collaboration Between Remote Users', In Schroeder, R. and Axelsson, A.-S. (eds.) *Avatars at Work and Play: Collaboration and Interaction in Shared Virtual Environments*, Springer: London.

Rosenman, M., Merrick, K., Maher, M. and Marchant, D. (2006) 'Designworld: A Multidisciplinary Collaborative Design Environment Using Agents In A Virtual World', *Automation in Construction* vol. 16, pp. 37-44.

Schmidt, K. and Simone, C. (1996) 'Coordination mechanisms: Towards a conceptual foundation of CSCW systems design' *Computer Supported Cooperative Work* vol. 5, no 2-3, pp. 155-200.

Schroeder, R. (2002) *The social life of avatars: presence and interaction in shared virtual environments*. Springer, London.

Schroeder, R. and Axelsson, A.-S. (2006) *Avatars at work and play: collaboration and interaction in shared virtual environments*, Springer, London.

Schroeder, R., Heldal, I. and Tromp, J. (2006) 'The Usability of Collaborative Virtual Environments and Methods for the Analysis of Interaction' *Presence: Teleoperators & Virtual Environments*, vol. 15, no. 6, pp. 655-667.

van Nederveen, S. (2007) 'Collaborative Design In Second Life', in *Proceedings of the Second International Conference World of Construction Project Management*, Netherlands, 2007.

Wadley, G., Gibbs, M. and Benda, P. (2007) 'Speaking in character: using voice-over-IP to communicate within MMORPGs', in *Proceedings of the Fourth Australasian Conference on Interactive Entertainment*, Melbourne, 2007.

Williams, D., Caplan, S. and Xiong, L. (2007) 'Can you hear me now? The impact of voice in an online gaming community'. *Human Communication Research*, vol. 33, no. 4 pp. 397-535.

Yee, N., Bailenson, J. N., Urbanek, M., Chang, F. and Merget, D. (2007) 'The unbearable likeness of being digital: the persistence of nonverbal social norms in online virtual environments.' *Cyberpsychology and Behaviour*, vol. 10, no. 1, pp. 115-121.