The Distributed Work of Local Action: Interaction amongst virtually collocated research teams

Dylan Tutt[1], Jon Hindmarsh[1], Muneeb Shaukat[2] and Mike Fraser[2]

[1]Work, Interaction & Technology Research Centre, Dept. of Management, King's College London, London, SE1 9NH. U.K. *{dylan.tutt; jon.hindmarsh}@kcl.ac.uk*
[2]Department of Computer Science, University of Bristol, Merchant Venturers Building, Bristol, BS8 1UB. U.K. *{muneeb; fraser}@cs.bris.ac.uk*

Abstract. Existing research on synchronous remote working in CSCW has highlighted the troubles that can arise because actions at one site are (partially) unavailable to remote colleagues. Such 'local action' is routinely characterised as a nuisance, a distraction, subordinate and the like. This paper explores interconnections between 'local action' and 'distributed work' in the case of a research team virtually collocated through 'MiMeG'. MiMeG is an e-Social Science tool that facilitates 'distributed data sessions' in which social scientists are able to remotely collaborate on the real-time analysis of video data. The data are visible and controllable in a shared workspace and participants are additionally connected via audio conferencing. The findings reveal that whilst the (partial) unavailability of local action is at times problematic, it is also used as a resource for coordinating work. The paper considers how local action is interactionally managed in distributed data sessions and concludes by outlining implications of the analysis for the design and study of technologies to support group-to-group collaboration.

Introduction

Over recent years we have witnessed the emergence of what have been termed "collaboratories"; formal collaborations between distributed research laboratories or groups that are connected via communications technologies. There is significant encouragement (through funding and other means) for inter-institutional research teams to be distributed nationally and internationally. However, time,

monetary and scheduling constraints restrict opportunities for research teams to congregate, to engage in research meetings and to collaborate on the analysis of data. Therefore there is a strong demand for systems and technologies to support virtually collocated research meetings. A particular challenge in the development of the collaboratory, then, centres on an obdurate problem in the development of CSCW systems – namely designing effective support for *synchronous* collaboration over and around common documents, objects and datasets.

Whilst "collaboratories" are primarily considered in relation to research in the natural sciences they are equally relevant to social scientific research. This paper explores one case concerning the use of new tools to support the real-time analysis of video data amongst distributed social scientific research teams. The tool they use, MiMeG, is designed to support a common practice for social science research communities engaged in video analysis - the 'data session'. In standard data sessions multiple individuals meet to view, comment on and collaboratively analyse video data. Thus, MiMeG is attempting to support 'distributed data sessions', where groups of geographically remote researchers can view video data simultaneously and conduct meaningful analytic work with those data.

The technical development of the MiMeG software has been introduced in an earlier paper (see Fraser et al., 2006) but here we discuss how it is used in practice to support the collaborative analysis of video data. In doing so, the paper highlights an issue rarely given serious consideration in the existing CSCW literatures on synchronous remote working – the organisation of 'local action'. Often studies report *that* action occurs at one site which is hidden from remote colleagues. However such action is generally treated as incidental, peripheral, disruptive, problematic or otherwise a distraction to the main business of a virtually collocated meeting. Few studies have focused on its interactional organisation in any detail. What is particularly interesting in this case is that, in contrast to other studies, it is not straightforwardly problematic and indeed at times is used as a resource by members of the research team to coordinate the business of the distributed group as a whole.

Remote Working & Virtually-Collocated Teams

There is a great deal of research in CSCW and cognate disciplines that highlights how remote working is a poor cousin to face-to-face meetings. Indeed physical collocation is usually considered "the gold standard of work environments" (Hinds & Kiesler, 2002: 56) and numerous studies powerfully reveal our "compulsion for proximity" (Boden & Molotch, 1994) and how "distance matters (Olson & Olson, 2000). However there is significant and increasing demand for various forms of virtual collocation and within CSCW there is a long-standing tradition of developing and evaluating systems to support synchronous remote working, from groupware through to various forms of media space and collaborative

virtual environment. A common concern for many of these systems has been an attempt to support group meetings focused on and around documents, objects and other media.

Studies of these technologies in use often highlight the lack of interactional cues available to remote participants and the difficulties that arise as a result. In particular they note how the bodily and material contexts of actions are 'hidden' from remote participants leading to a range of troubles for participants to assess the sense and significance of those actions. Cursor movements, avatar actions or even displays of conduct on video are somehow, and in various ways, disembodied and disembedded. Thus action at one site ('local action') is unavailable to remote participants and is disconnected from the work of the group as a whole.

The focus for many of these studies has been on systems that support collaboration between two or more *individuals* distributed across workplaces (Mark et al., 2003). However there is a growing demand for groups to be virtually connected to other groups, especially (but by no means exclusively) within research communities.

Studies of group-to-group collaboration also discuss how the lack of access to local action at remote sites is problematic in various, although somewhat different, ways. For example Olson & Olson (2000: 147) suggest that in co-present team meetings "[p]articularly important is the spatiality of human interaction … If a team member wants to observe his manager's reaction to a point someone made he can just glance quickly in her direction", whereas in virtual meetings this is not possible. In studies of virtually collocated, interdisciplinary teams at Boeing, Mark et al. (1999) found that participants had difficulty identifying who was talking over the audio conferencing link and that they were distracted by parallel activities that they undertook whilst attending the meeting. Ruhleder (2000) has also argued that the interactional demands in local sites can distract from the mediated communication. Aoki et al. (2003) dismissed 'casual conversation' in local sites as incidental to the business of the meeting. Sonnenwald et al. (2002: 125) also consider local action to be peripheral, as they report that it has become common practice in videoconferences for participants to "cover the microphone closest to them" to mute or muffle talk and action that may be heard at other locations "including whispers or side comments, munching on chips, sneezes and page turning".

Furthermore it is argued that asymmetries in access to local action can lead to the formation of 'sides' in distributed groups. Bos et al. (2004, 2006) pay careful attention to the dynamics of 'partially distributed groups' in which some participants are collocated while other individuals join in remotely. They show how participants "experienced 'collocation blindness' and failed to pay enough attention to collaborators outside of the room" (Bos et al., 2006: 1313) and have argued that distributed teams tend to form subgroup identities based on their shared physical location, where people typically enjoy more interaction and share more

information with each other than they do with remote partners and even begin exhibiting in-group behaviours (Bos et al., 2004).

So, the widespread research on remote working and virtually collocated teams reveals how the unavailability or partial availability of 'local action' at remote sites is treated as incidental or straightforwardly problematic (a nuisance, a distraction, etc.) or leads to undesirable conduct (e.g. the development of in-group behaviours). However none of these studies takes the interactional organisation of local action as a topic of inquiry in its own right or explores the wider range of ways in which 'local action' bears upon 'distributed work'. Our studies of the use of MiMeG to support distributed research teams provides opportunities to begin to treat these issues seriously.

The System: MiMeG

Within the social sciences, many researchers working with video materials recognise the value of being able to share, show and discuss data with others. One dominant means for doing this is the 'data session', where colleagues and peers congregate to view and collaboratively analyse video data. This enables participants to explore tentative formulations and analyses and to receive immediate comment, contribution and feedback from colleagues in relation to their data. Participants can range in number from a minimum of two to a quite sizeable small group, of possibly up to twenty or so – although beyond that the dynamics transform significantly. The data session can be relatively formally structured with an introduction to the data, the viewing of the data, time for participants to make notes and subsequently opportunities for each participant to make comments. Alternatively someone can just start the video and anyone can ask questions or raise issues. Whilst data sessions may vary in form, they are common activities for many in the video analytic research communities of sociology, psychology, education, anthropology, linguistics, geography, CSCW, HCI and more.

As mentioned earlier, there is increasing support for inter-institutional national and international research projects, consortia and networks. As a result there is growing demand for technologies to support 'distributed data sessions', where remote participants can see, discuss and collaboratively analyse video data in real-time. MiMeG is a preliminary attempt to do this and it enables members of a research team located at two or more sites to simultaneously watch and discuss fragments of video data. Note that it is not intended to replace face-to-face meetings – indeed there is evidence to suggest that tools for remote collaboration tend to work better if participants do meet up regularly aside from their virtual meetings (Olson & Olson, 2000). However there is demand for such systems to *supplement* existing face-to-face meetings.

The design of the system (for more detail, see Fraser et al. 2006) was founded on an understanding that the visibility and control of video data is central to the

data session. There are of course many systems designed to support distributed co-working on different media, however the support for work on video materials is rather primitive. Therefore the system ensures that high quality video is played simultaneously at all sites. There is no built-in video view of the other group(s) as we wanted to begin with a system that distributed the data coherently. Indeed recent studies have shown that, especially for visually complex tasks in which the focus of attention changes frequently (such as identifying and orientating to features in video data), a shared view of the 'task space' is essential, and can be more useful than a limited view of the 'person space' afforded by traditional videoconferences (see Kraut et al., 2002). The research team is however connected via Skype (free audio conferencing software) and whilst Skype can provide a basic video conferencing link, early trials indicated that it provides too basic an image of other sites for it to add value.
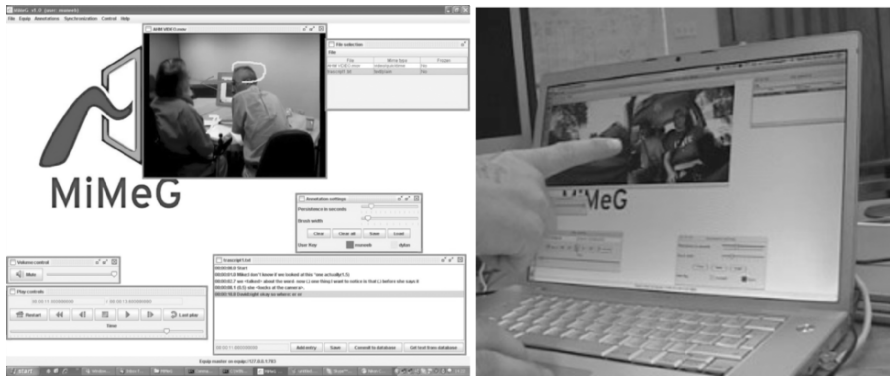

Figure 1: The Interface includes video windows, playback controls, annotation controls and windows for other media (transcripts, images, etc.). The system can be used with computer screen and mouse or projection screen and pen-based input.

The repeated playback of a video fragment is a routine practice in data sessions, whether to help participants become familiar with the sequence of actions in the data, or for more focused, finer grained video analysis. Therefore a range of controls for playback is available. Playback controls rest at a single site, although participants can choose to formally pass control to other sites.

Another key concern in design was to support participants in indicating features in the video. One of the major activities in data sessions generally is that participants encourage colleagues to notice phenomena on screen. This then forms the basis for analytic discussions regarding the significance of those phenomena. As video is not static (unless on freeze frame), the phenomena to be discussed are fleeting – they may only appear on screen for a second or less. A glance, a gesture, a nod, a movement of a pen, a stroke on a keyboard or whatever are difficult enough for an individual to *spot* on video, let alone to reveal to oth-

ers. Thus there can be substantial coordination work involved in getting a group of others with different viewpoints in the room to see some action on screen. Therefore MiMeG enables all sites to *annotate* the video stream in real time by drawing on it using mouse or pen input – these annotations appear on all screens simultaneously. Annotations can be added whilst the video rests on a single frame or indeed during playback.

Data sessions also routinely involve additional media, such as transcripts of talk, images, photographs, etc. Transcripts in particular tend to form a fundamental resource for participants to locate interesting features in the data. As a result, MiMeG enables all sites to display common additional media. Furthermore it allows transcripts to be time-stamped so that participants can navigate the video clips using the transcript that they have produced.

The development of this system will potentially be of value not only to the range of social scientists engaged in the analysis of video materials, but also to the broader range of practitioners keen to undertake collaborative video analysis at distance – for example, performance analysts, film and video editors, video forensics specialists and the like.

Real-World Trials

MiMeG has been widely distributed to social scientists. However a small number of groups have been provided with additional technical support and assistance. These groups have agreed to be studied during their use of the system and here we report on early experiences with MiMeG by one of these teams. The team comprises of four members of three different departments within two, geographically remote UK universities. In the past, members of the team regularly met for informal data sessions. However in the present context they are collaborating on a funded research project concerned with car sharing. They have collected a large corpus of audio-visual recordings of action and interaction in cars and regularly hold data sessions in which they explore issues concerning way-finding, instruction-giving, domestic routines and the like. They are interested in using MiMeG for two key reasons: firstly it would enable them to continue to hold data sessions when one or more of the team is visiting a more remote third institution for a significant period (which is often the case); and secondly it would provide opportunities for more ad-hoc, unplanned, less time consuming data sessions to discuss specific issues that arise during data analysis by an individual team member.

We recorded two one-hour distributed data sessions. As the participants already know each other well they can be seen as a "gelled social group" (Aoki et al., 2003). This situation can be placed in contrast to experimental groups brought together for a trial, or established groups facing a work task in which the system might be viewed to be of dubious benefit to them. This case provided a highly relevant test of the technology. The research team was engaged in a meaningful

data session, on real project issues, and where the team were keen to adopt the system more generally. For comparative purposes we recorded a regular, co-present data session involving this team as well as a number of co-present data sessions with other similar research groups.

We video-recorded the distributed data sessions by using two camcorders at either site – one to depict the participants clearly and the other to display a close-up image of the on-screen activity. We obtained written consent from all participants for the presentation of the data in this paper, although we decided to give them pseudonyms. In working with the data we adopted an analytic approach common to CSCW, namely video-based field studies informed by ethnomethodology and conversation analysis (e.g. Heath and Luff, 2000). In analysing the use of the technology the concern should not be seen as a traditional evaluation of the system properties or functionality. Rather we are attempting to explicate the interactional practices that emerge in managing the technology in use. Our understanding of these practices is intended to stimulate further issues for research and to inform the design of future technologies.

Local Action as 'Side Work'

Numerous studies of distributed collaborative work have described the nature of action at one site that is (at least partially) unavailable to remote colleagues. However such actions are usually dismissed as being off topic or incidental (Aoki et al., 2003) or a distraction from the core meeting activities (Ruhleder, 2000). Within the distributed data sessions under consideration, there were numerous instances of such 'local action', action that fell beneath the remote site's horizon of notice. However on close inspection such action can *at times* be seen to be a central resource for the coordination of the distributed team's work.

Consider the following example from a distributed data session involving the research team. Eddie and Ivor are based in one UK city (Site 1) and their colleagues Ben and Henry are in another UK city (Site 2). In **Fragment 1**, the two sites are tackling a common problem that emerged in our early trials with MiMeG – that is establishing whether or not the video playback is running in unison at each site. This is in order to ensure that they are discussing the same part of the scene. As the fragment begins Eddie asks about the positioning of the traffic in the paused video as a means of assessing video alignment (L.1-3).

After asking Ben about the relative positioning of the car and the van on their screen in Site 2, Eddie (to the right of Image 1.1, partly obscured) moves on to describe what can be seen at Site 1 (L. 9-13). As he does this, Ivor leans in towards the screen and alongside Eddie and slowly extends his right arm, holding his pen as a pointer (during the word "overlapping", L.12). The gesture arrives at the playback control window on the screen during Eddie's stretching of the word "by:::" (L.12), but does not intrude on Eddie's view of the video window. Eddie

is therefore made aware of Ivor's upcoming contribution, but is able to describe the scene in the video window, without having his view obscured or his account otherwise disrupted by Ivor's gesture.

Fragment 1 (Bold in the transcript marks the line where the image occurs)

```
1  Eddie: and in yours is- are the
2         van and the car right
3         beside each other?
4  Ben:   er (.) no:: the van's behind
5         the car
6  Eddie: by a ↑lot?
7         (0.9)
8  Henry: er- fifteen ↑feet
9  Eddie: we've got them (.) pretty
10        much parallel (.) in
11        fact the van's (0.4)
12        overlapping by::: one or
13        two foot probably
14 Ivor:  °the code reads°
15 Eddie: erm (.) our time code
16        ends 06↑4
```

Image 1.1 (L.14)

At the end of Eddie's utterance, Ivor moves his pen to the time code in the playback window and, as Eddie looks down, Ivor slides the pen across the scale and suggests that Eddie use the time code to specify the paused video frame. Spoken softly with his head tilted close to Eddie, Ivor's utterance "the time code reads" is designed to be heard locally and not remotely. Eddie subsequently reads out the code for their remote colleagues in order to assess the extent of the video misalignment by comparing frame numbers.

There are three points to raise here. Firstly, Ivor's actions are designed explicitly for his local colleague – the gesture is invisible remotely and his talk is very softly spoken so that he cannot be heard by the others. Secondly, his action is delicately coordinated such that it does not disrupt Eddie's talk to the remote site. Thirdly, it is designed to *contribute*, to support and not to distract from the general activities of the data session. It is not peripheral to the business at hand, but rather very much on topic. Ivor not only encourages Eddie to notice the time code, but to elaborate on his description and announce the accompanying time code to the remote site.

This sort of 'side work' at local sites is a fairly common feature in our data. Consider **Fragment 2,** which is a further example of how side work contributes to the meeting as a whole. Here, with the same arrangements of participants, Henry faces the problem of getting others (at both the local and remote sites) to see something that he has noticed in the video data – in particular evidence of 'no entry' to a particular side street. The research team is studying a car journey through a city and at this time they are interested in the organisation of directions given by one passenger. Eddie (at Site 1) questions how Henry (at Site 2) knows that there

is 'no entry' to a road in the route under examination. Henry and Ben (both at Site 2) look for evidence in the video.

Fragment 2

```
1 → Henry:  that was it there I think
2   Ben:    °oh was it
3   Henry:  yep
4           (3.1)
5   Henry:  we're just going back a bit
6           (3.8)
7   Henry:  °mark it on the screen°
8           (0.8)
9   Ben:    where ↑is it?
10  Eddie:  yeah so there's right hand turn markings on the
11          road there
12          (0.4)
13→ Henry:  °back a bit°
14  Ben:    Henry thinks he can see it °but I'm no(t)°
15  Eddie:  yeah there's a (.) you- there's markings ↑on the
16          road (1.9) so you can do a right turn
17          (1.1)
18  Henry:  there
19  Ben:    have we ↑missed it have we h[ere?
20  Henry:                              [no no:: (.) that's
21          just coming up
```

In data sessions, participants are routinely called upon to ground their analytic claims in observable evidence in the video data. Therefore much of the work of data sessions involves getting others to see such evidence. This often involves the rewinding, pausing and playing of the fragment at moments relevant to the observation being made. This is complicated considerably when the phenomena being pointed out are in the moving video rather than a paused image and thus may only be on screen for a moment or two.

In this case, the matter is further complicated as the data are from a moving vehicle and they are trying to spot something at the side of the road – thus it is at a distance from the camcorder (therefore small) and only visible momentarily (as the car passes by). When Ben plays the clip, Henry leans in to the screen to prepare to spot and point out evidence of 'no entry'. Moments later he reaches out and points to the laptop screen – an action only available to Ben in the local site. As he points, Henry leans towards Ben and quietly says to him: "that was it there I think" (L.1). This is only available locally and the participants at the other site display no orientation to the utterance to indicate that it was audible to them. However at the local site Ben's hand immediately moves to the playback controls to rewind the fragment, thereby displaying his understanding that the relevant moment in the video has passed.

Once that section of the clip begins to play again (L.6), Henry promptly points towards the screen (Image 2.1) and softly suggests that Ben "mark it on the screen". However this time Ben keeps the video playing. Indeed he asks quietly

"where ↑is it?". At this moment, having drawn back his pointing gesture a little (Image 2.2), Henry transforms the poised finger into a more substantial and cruder backwards-thumbing gesture to request rewinding (Image 2.3). This 'hitch-hiking' hand gesture suggests a sizeable rewinding required from Ben, and that the phenomenon is well past. Ben immediately pauses the clip before rewinding it. Much of this co-occurs with Eddie's comments in lines 10-11. Thus they are able to re-position the video without interrupting him.

Image 2.1 (L.7)   **Image 2.2** (L.8)   **Image 2.3** (L.11)

When the clip begins to play for a third time, Henry again points firmly towards the screen. Of particular note here is that as Ben continues to let the clip play, Henry's pointing finger starts to slide to the right of the screen as if to mark how the feature (the 'no entry' sign) is passing off screen. So, Henry transforms his referential practice into a representational gesture that 'marks up' a virtual route extending out of the screen. His finger indicates where the feature is going if the route could be seen trailing off screen. Indeed Ben treats it as if the phenomenon has gone off screen – he stops the video immediately and says "have we ↑missed it have we here?", but he is assured that it is still on screen.

This fragment reveals how side work rests upon the rich interactional resources available to co-present colleagues. The visibility of subtle and delicate gestures and movements facilitate the close coordination of conduct. These are of course unavailable to remote sites. However, crucially the side work is organised in parallel and in between contributions to the group as a whole. In Aoki et al's (2006: 398) discussion of 'aside' turn types, local talk in remote sessions is described as the production of a "turn in a soft voice (especially when produced in overlap) [which] targets the action towards people who are not attending to the main conversation as primary participants". However these fragments show that side work is not an "aside" to the meeting but rather conduct that contributes to the very core of the business at hand. Thus local action and distributed work in these distributed data sessions are deeply co-implicated.

Revealing Local Action

Often within these distributed data sessions, aspects of local action are explicitly announced or revealed to the remote site. That is to say, participants somehow describe or narrate what is unavailable remotely. The following fragments allow us to explore when and how these announcements are produced and how they are orientated to by participants at remote sites. Consider, for example, the previous fragment, **Fragment 2.** Both Henry and Ben produce accounts of their side work at certain moments. Their local action is only partially available to Eddie and Ivor. It is partially available as they are able to see that the video is jumping around and at times playing, but they cannot see the delicate embodied work of Henry's instructions and Ben's control of the playback.

At two moments Henry and then Ben reveal the nature of their side work. The first (L.5) follows a 3.1 second pause in talk, during which the video is stopped and jerks through a couple of still frames as Ben rewinds it. Henry accounts for the movements of the video when he explains that "we're just going back a bit". As with most turns in these data sessions the recipients are clear (indeed few problems regarding the intended recipient for a turn emerge in our data). The turn is designed for the remote audience. Henry leans towards the microphone and employs the pronoun "we" to mark out his local 'side', thereby further identifying the remote site as addressees. Its production at this moment accounts for the on-going pause in the discussion and the movements of the video playback.

The second instance (L.14) comes moments after a comment from Eddie. Neither Henry nor Ben attend to Eddie's turn immediately. However after a pause, and quiet local instruction from Henry, Ben starts to rewind the clip again. He says "Henry thinks he can see it °but I'm no(t) °", again revealing that they are engaged in ongoing side work to position the video clip. It accounts both for the visible rewinding of the clip and for the lack of response to Eddie's turn. However, in contrast to the previous instance, the design of the turn seems to mark this as Henry's perspective.

The next fragment shows how the activity of revealing local action may not necessarily take the form of a clear announcement in talk. Rather this one takes the form of a 'response cry' (Goffman, 1981). In **Fragment 3**, Ben and Henry are based at one site, while Eddie is the sole 'full' participant at his respective site. However, also in Eddie's room is Muneeb who is providing technical support on the use of MiMeG. He joins the action because Eddie experiences problems manipulating the video. The fragment begins with Ben introducing a new line of inquiry, which focuses on one particular utterance ("now we've got a problem") spoken by one of the people in their data. In response to this new topic (L.1-4), Eddie, who has the playback controls, starts to try to find that point in the video.

Fragment 3

```
1    Ben:    so erm (.) I >don't know if I've said this before<
2            but one thing I've been thinking about was the
3            (0.2) we've got a ↑problem >now we've got a
4            problem< line
5            (1.2)
6    Eddie:  yea↑h
7    Ben:    do you know where that ↑is?
8            (1.1)
9    Eddie:  yep hold on
10           (4.2)
11   Henry:  °you mean in the sense of [how does it become a=
12   Muneeb:                           [drag it
13   Henry:  =a problem?°
14   Eddie:  do I have to drag it
15   Ben:    no [I was thinking of just [what was the guy
16   Muneeb:    [ye↓ah
17→  Eddie:                             [↑tsh hhh
18   Ben:    doing (.) wi[th hh.
19→  Eddie:              [URGH↑hhh::
20   Ben:    wi[th ha::
21   Henry:    [ha hhh
22           (1.9)
23   Ben:    with that line and her reaction to it
24   Eddie:  er:: I think it's ↑here
```

Image 3.1 (L.3)   **Image 3.2** (L.5)   **Image 3.3** (L.19)
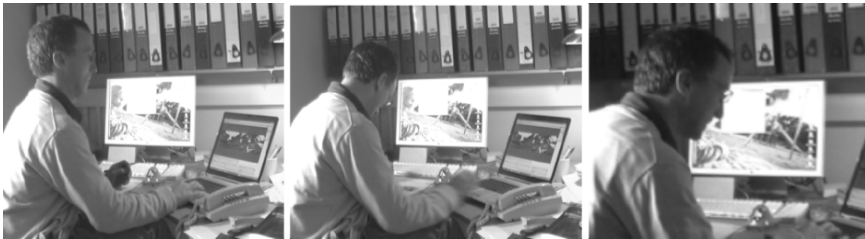
In data sessions, someone quoting a line in a transcript often initiates new topics. They may raise an analytic puzzle for others to explore or simply express interest in the utterance design. However it will routinely lead to that part of the video being played (and replayed) for analysis. In this case, after Ben quotes one line from their transcript, Eddie starts to search for the relevant part of the video. He first turns away from the screen to the paper transcript on the desk to his left in order to find the relevant line of transcribed talk (Images 3.1-3.2). This will be used to help locate the right position in the video fragment. This takes place during the 1.2 second pause (L.5) in which Eddie's activity is unavailable or observable to Ben and Henry. When Eddie says "yea↑h", the higher intonation at the end of the utterance works to request more from Ben, who then asks explicitly "do you know where that ↑is?" (L.7). As Ben cannot see that Eddie is looking for it, this maybe raises uncertainty about how he is participating during these moments.

Eddie's next utterance ("yep hold on", L.9) reveals to the remote site that he is engaged in the task and 'buys' him some time to find the relevant part of the video. Whilst Eddie tries to work the system, the video remains static at the remote site. However Muneeb, in the room with Eddie, notices his difficulties with the controls and advises him to "drag it" (L.13).

As Eddie repeatedly fails to 'drag' the bar on the playback controls, he produces two "response cries" (Goffman, 1981) which interrupt a comment by Ben (L.17: "↑tsh hhh"; L.19: "URGH↑hhh::"). Goffman defined response cries as "exclamatory interjections which are not fully-fledged words. *Oops!* is an example", which display "evidence of the alignment we take to events" (Goffman, 1981: 99-100). In this case, considering that the task involves the movement of a finger across a laptop touchpad, rather than one demanding physical exertion, the second cry (Image 3.3) is somewhat overstated. However this 'strain grunt' demonstrates his difficulty in operationalising Muneeb's prior suggestion to "drag it" and more generally reveals problems in working the system to the remote site. Thus it reveals problems to both the local audience (Muneeb) and the remote audience (Ben and Henry).

Ben and Henry's laughter acknowledges the trouble (L. 20-21) and Ben pauses for a while before completing his turn (L. 23) – indeed Goffman argues that these cries serve as "a warning that at the moment nothing else can claim our concern" (Goffman, 1981:105). Thus Eddie's 'strain grunt' both interrupts Ben's turn and simultaneously accounts for the interruption, by alerting the remote site to the extent of, and his preoccupation with, local problems.

Revealing local action to the remote site at opportune moments is significant to the success of the distributed data session. In particular, these moments when participants announce, describe or otherwise reveal key aspects of their local action often follow, or are produced, during pauses in conduct. Revealing the character of local action often indicates that a moment or two is needed before they can resume with the distributed work of the research team (cf. Hindmarsh et al., 2000). They are timed and designed to do this work. Another example we have is when a participant leaves the room. This is not announced immediately but rather timed and designed with regard to practical issues at hand. It only comes to light after the end of an ongoing turn and a pause where the absence of the other may be accountable. As much of the local action in these clips is partially available to remote colleagues through the video playback, participants render visible the hidden work that makes sense of the movement of video on display. However it is of course not only about revealing what can be seen at this moment, but informing remote colleagues of what can be expected in the moments to come. Thus local action is revealed to support the smooth coordination of the distributed work.

Tag Work: Re-working local action

In these distributed data sessions, there are a number of examples in which local actions are *re-worked* to be made available to the distributed team as a whole. There is not an explicit description or announcement of local action, as with the activities of revealing local action, but rather the utterances are built upon and re-designed in subsequent contributions to the data session. We term this 'tag work' as it usually takes the form of one person taking on or building up the local action of a co-present colleague.

As an illustration, consider **Fragment 4**, from a data session involving Ben and Henry at one site, with Eddie at the other. We join the action after a long 10.4 second pause. The team has been discussing the 'duties' of the driver to the passenger during a journey. Following the pause, Ben quietly quotes an utterance spoken in their video data – "I just need to get to the bottom of this" (L.1) – which is only hearable at the local site.

Fragment 4

```
1     Ben:   °I just need to get to the bottom of this:°
2            (3.0)
3→    Henry: I suppose there in that statement too about I just
4            need to get the- to the bottom of this there's
5            also a sense in which (.) it will be her: that
6            takes over the task once again once they've- when
7            they've arr↑ived (.) you know that (.) th[at
8     Eddie:                                          [that's
9            nice
```

Towards the end of the long pause Ben sits back, strokes his forehead and raises his eyebrows (Image 4.1). This embodied display of 'thinking' or 'ponder-ing' is accompanied by the quote from the data, which is only clearly audible to Henry. Eddie, at the remote site, cannot be seen to acknowledge or in any way display having heard this utterance, and from our recordings it seems that it only comes across faintly as muffled talk. It is clear that the turn is not designed to *de-mand* response from any party, but it does offer opportunities to take it up, espe-cially to Henry.

Of particular interest here is the way in which Henry *re-works* the turn mo-ments later ("I suppose there in that statement too", L.3). Essentially Ben's quote encourages Henry to pursue this line of analytic inquiry. As such Ben's local talk is 'picked up' and transformed by Henry. He uses the fact that the talk was clearly hearable to him to proffer an analysis of the utterance to which Ben is referring. However his comments are not just designed for Ben (in the form of side work), but are rather re-directed to the whole research team.

He does not look at Ben, but rather turns towards the screen and microphone, and makes his talk clearly audible remotely by raising his voice (Image 4.2). In doing so he orients to the team as a whole, rather than simply the colleague along-

side him. Talking more loudly and leaning towards the microphone ensures that the remote site can hear. This is a common trend in our data and interestingly 'general' talk seems to have the flavour of addressing the remote site as opposed to the local site. So, in this case, Henry transforms the audience for, and potential participants to, the comments.

Image 4.1 (L.1)   **Image 4.2** (L.7)   **Image 4.3** (Beyond transcript)

His vocal stress on the words "there" and "too" acts as key building blocks in the transformative activity. They position Henry's talk as not designed to display a *new* point, but rather to display that it is built on, and is adding to, a prior. Furthermore, by repeating Ben's quote "I just need to get to the bottom of this" Henry reveals uncertainty as to whether Eddie will have heard Ben's talk. By integrating the quote into his new contribution, it ensures that whether or not Eddie has heard Ben's turn, he will still understand his comments. So the turn *does not require* that the prior has been heard to be made sense of, and yet it very much builds on and attends to that prior.

 After he receives positive alignment from Eddie ("that's nice", L.8-9), Henry turns to face Ben (Image 4.3) while continuing with his tentative analysis of the utterance originally quoted by Ben. In doing this Henry then draws a series of nods from Ben. This local action, which is unavailable to the remote site, provides visible alignment to (parts of) Henry's analysis and encourages him to continue.

Local action can also be re-worked for the remote site through the shared MiMeG workspace. Consider how **Fragment 2** continues beyond the point that we previously discussed. Once Ben and Henry reach a frame of the video that features the relevant side road, Ben's draws on the frame, using the annotation tool, to ask Henry whether the sign on-screen is the no entry sign that he is looking for. Prior to this point, Henry had been gesturing over the screen – gestures that were not available to Eddie and Ivor. However Ben's annotation crucially transforms the audience by making it available to the remote site through the shared workspace (Image 2.4).

Fragment 2 con't

```
19   Ben:    have we ↑missed it have we h[ere?
20   Henry:                             [No no:: (.) that's
21           just coming up
22→ Ben:    about there?
```

```
23          (2.2)
24   Henry: think [so
25   Eddie:       [erm (.) you've got the guy on our monitor
```

      Image 2.4   **Image 2.5**

During "No no:: (.) that's just coming" (L.20-21), Henry holds a pointing gesture still while Ben scrolls the cursor arrow up to an on-screen location. He starts to draw a red circle at the no-entry road and Henry pulls back his pointing finger. Once the annotation is complete, Ben asks Henry "about there?". While the annotation tool features in the discussion between Ben and Henry to clarify a feature locally, it renders previously unavailable features of that discussion available to the remote site for their scrutiny. Thus Ben 'builds on' Henry's pointing gestures and adds to them by specifying them on screen.

Interestingly it is only when the annotation is committed to the video stream that a problem with video alignment between the two sites is revealed. Ben's data mark-up appears a sizeable distance from the relevant feature on the remote screen ("erm (.) you've got the guy on our monitor"). Unbeknown to the other, each site had set their video window to a different size during the course of the data session, which caused the annotation to mark up different parts of the scene across the sites (Image 2.5). The extent of the misalignment is revealed across sites as the teams set about circling the heads of the subjects in the video stream.

So 'tag work' involves the transformation of locally available conduct, whether purely vocal or additionally non-vocal, to make it available to the distributed team as a whole. As we have seen, this transformation can be undertaken purely through talk or through the additional affordances of the shared digital workspace. This sort of transformation marks something as relevant or significant enough to be shared more generally rather than remain as side work. Also the new comment or contribution builds on the prior local action. It is designed in such a way that it indicates that local action is being referenced, but it also captures and re-iterates that local action; the prior action is integrated into the new contribution. Thus the new turn ensures that it can be understood even if the prior has not been heard or seen by remote colleagues. This also reveals that participants can be uncertain whether local action has been heard remotely.

Discussion

This paper has begun to unpack the nature and organisation of local action within data sessions held by virtually collocated teams. In CSCW, often such local action has been treated as straightforwardly problematic, peripheral, irrelevant, a nuisance, or a distraction to distributed group work. However the data presented here reveals a more complex picture of local action. It should be noted that we did not search out 'constructive' local action. Rather we were interested in the interactional practices that underpinned the work of the distributed data sessions and in taking such an interest we found local action to be critical to that work. Let us also be clear that we are not suggesting that the unavailability of local action is *always* beneficial. Indeed we have a number of instances in the data that demonstrate troubles that arise due to the lack of information of activities at the other site (e.g. the start of Fragment 3). However we are keen to emphasise that local action should not be straightforwardly glossed as peripheral or problematic, so we have attempted to redress the balance by focusing on examples that do not routinely appear in the literature.

A key message here is that local action should not be disregarded in analysis, evaluation or design. The examples of local action in our data cannot be disconnected from the work of the distributed group, but rather feed into the central organisation of the group's work. Taking it seriously in the study of emerging group-to-group conferencing systems may involve challenges in terms of capturing such action, but it may also reveal critical practices. In terms of design, if anything the implications are really rather positive. The fragments certainly do not focus on a problem to be solved. Rather they show how participants exploit the technological asymmetries to coordinate work. So while there are good reasons for designers to consider technically complex solutions that reveal local action to overcome well-known problems (indeed we are pursuing such a line in parallel work), designers should not be put off more lightweight solutions that may be more immediately deployable (at least for the group sizes that we have discussed here). In doing so they might do well to consider the best configurations of technology to embrace local action.

There are two further implications that we would like to raise, which concern our understanding of 'schisms' in virtually-collocated meetings and our understanding of how 'sides' form in distributed work.

The phenomenon of 'side work' that we discuss relates to the concept of 'schisms' familiar to studies of co-present interaction. In studies of co-present meetings multiple parallel conversations, or 'schisms', often break out. The examples in the section on 'side work' could be seen as schisms. However unlike the literature on schisms, side work represents local action that is very much *in-topic* rather than off-topic in that it contributes to the immediate work of the distributed group as a whole. 'Side work' is also similar to the 'side sequences' dis-

cussed by Jefferson (1972). Again however there are notable differences. Side sequences refer to breaks in activity that clarify problems or issues before that activity resumes. While side sequences are inserted within the flow of conversation, due to the configuration of MiMeG, side work occurs *in parallel* to ongoing talk. Furthermore, the environment makes the side work potentially *invisible* to some participants. In a co-present meeting one can see someone whispering to or gesturing at another. Here such conduct is invisible. As we have seen, this provides opportunities for participants to do local work to make the distributed data session work. This makes it possible for Ivor to design his contribution in Fragment 1 so that it is not seen or heard remotely. It is positioned *alongside* the ongoing talk but such that it then re-shapes the work of the group as a whole. Thus standard concepts in the analysis of co-present interaction take on a new form, organisation and significance in these mediated encounters.

A number of studies have noted *that* co-present colleagues in distributed groups tend to develop 'sides' (e.g. Bos et al. 2004) and the reason for this is often linked to the additional social cues available to co-participants in local sites. The findings presented here contribute to this work by revealing some of the interactional practices that underpin *how* sides emerge. Take for example the case of 'tag work' presented in Fragment 4. In a co-present data session with four participants, all parties would be able to hear what Ben said. Thus all parties would be able to build on his comment to progress the analysis. However due to MiMeG, Henry is given *unique access* to the comment of his co-present colleague and is the only party able to build on from it. Sacks (1992) writes of the ways in which finishing another's sentence can give an impression of a team. Here we can extend that to suggest that building on the contribution of another also does so. Furthermore in revealing local action, participants routinely announce what 'we' are doing, again giving the flavour of a co-operative. Moreover as talk to the team as a whole seems to be directed to the remote site, it possibly gives the remote site primacy in claiming next turn, giving rise to a site-to-site (side-to-side) turn-taking system. Each of these practices fosters some sense of 'local team' and provides the turn-by-turn basis in and through which sides emerge.

These observations may be of interest to those studying or developing large group conferencing systems. Sometimes people at the same site do not represent the same interests and maybe more importantly people that represent the same interests are not necessarily at the same site. Therefore they are denied opportunities for side work; to coordinate contributions, check facts, help out, etc. Our study here helps us to encourage work that is exploring mechanisms that could facilitate *cross-location* side work. To this end, Access Grid and other conferencing systems might consider text chat (Mark et al. 1999) or 'space like systems' which can support sub-conferences or allow directed comments through "whispering" (Berc et al., 1995; Yankelovich et al., 2005). Whichever tools are selected it is critical to ensure that they do not distract from the meeting but pro-

vide resources for side work to be intertwined with ongoing team activities or even to facilitate opportunities for tag work.

That said, clearly this paper is presenting early findings from our programme of studies. In our concern to provide support for 'real-world' teams, we are directed in part by their group composition and the technologies that they use. This particular team had four members and ran data sessions using standard computers or laptops. In future work we are keen to explore larger group sizes and the use of different display technologies to see how practices translate or 'scale'. For example we are interested to see if the relationships between local action and distributed work become more complex when there are opportunities for parallel instances of side work in local sites. How then does *simultaneous* side work get drawn into the meeting as a whole? We are also eager to consider the impact of different technological ecologies, ranging from multiple screens, to tabletop displays, to wall projections and the like on the organisation of action at local sites. It is likely that these different ecologies impact on how local participants can manage relationships between local action and distributed work.

In conclusion, the data in this paper reveal practices that are quite distinct from co-present data sessions. Whilst the tasks are the same – finding and showing phenomena, making analytic claims, supporting claims with video evidence, etc. – the interactional asymmetries imposed by MiMeG lead to new forms of coordination. With the development of "collaboratories" there will inevitably be further study of group-to-group(s) systems and, as we have started to see here, the focus on groups interacting at distance and over and around data may well reveal intriguing practices; practices that allow us to refine our understanding of concepts such as 'awareness', 'involvement' and even 'activity'. Indeed these systems give rise to novel and complex participation frameworks that may contribute as much to our understating of the dynamics of social interaction as to the design of new technical solutions.

Acknowledgments

References

Aoki, P., Romaine, M., Szymanski, M., Thornton, J., Wilson, D. & Woodruff, A. (2003): 'The Mad Hatter's Cocktail Party: A social mobile audio space supporting multiple simultaneous conversations', *Proc. CHI 2003*, ACM Press, New York, pp. 425-432.

Aoki, P., Szymanski, M., Plurkowski, L., Thornton, J., Woodruff, A. & Weilie, Y. (2006): 'Where's the "party" in "multi-party"?: Analyzing the structure of small-group sociable talk', *Proc. CSCW 2006*, ACM Press, New York, pp. 393-402.

Berc, L., Gajewska, H. & Manasse, M. (1995): 'Pssst: Side Conversations in the Argo Telecollaboration System', *Proc. UIST '95*, ACM Press, New York, pp. 155-156.

Boden, D. & Molotch, H. (1994): 'The Compulsion of Proximity', in R. Friedland & D. Boden (Eds.) *NowHere: Space, Time and Modernity*, University of California, London. pp. 257-286.

Bos, N., Olson, J., Nan, N., Shami, N., Hoch, S. & Johnston, E. (2006): 'Collocation blindness in partially distributed groups: is there a downside to being collocated?', *Proc. CHI 2006*, ACM Press, New York, pp. 1313-1321.

Bos, N., Shami, N., Olson, J., Cheshin, A. & Nan, N. (2004): 'In-group/out-group effects in distributed teams', *Proc. CSCW 2004*, ACM Press, New York, pp. 429-436.

Fraser, M., Hindmarsh, J., Best, K., Heath, C., Biegel, G., Greenhalgh, C. & Reeves, S. (2006): 'Remote Collaboration over Video Data: Towards real-time e-Social Science', *Computer Supported Cooperative Work*, 15, 4, pp. 257-279.

Goffman, E. (1981): *Forms of Talk*, University of Pennsylvania, Philadelphia.

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

Hindmarsh, J., Fraser, M., Heath, C., Benford, S. & Greenhalgh, C. (2000): 'Object-Focused Interaction in Collaborative Virtual Environments', *ACM ToCHI*, 7, 4, pp. 477-509.

Hinds, P. & Kiesler, S. (Eds.) (2002): *Distributed Work*, MIT Press, Cambridge, MA.

Jefferson, G. (1972): 'Side sequences', in D. Sudnow (Ed.) *Studies in Social Interaction*, Free Press, New York, pp. 294-338.

Kraut, R., Gergle, D. & Fussel, S. (2002): 'The Use of Visual Information in Shared Visual Spaces', *Proc. CSCW 2002*, ACM Press, New York, pp. 31–40.

Mark, G., Abrams, S. & Nassif, N. (2003): 'Group-to-Group Distance Collaboration: Examining the "Space Between"', *Proc. ECSCW'03*, Kluwer, Dordrecht, pp. 99-118.

Mark, G., Grudin, J. & Poltrock, S. (1999): 'Meeting at the Desktop: An empirical study of virtually collocated teams', *Proc. ECSCW'99*, Kluwer, Dordrecht, pp. 159-178.

Olson, G. & Olson, J. (2000): 'Distance Matters', *Human Computer Interaction*, 15, 2/3, 139-178.

Ruhleder, K. (2000): 'The Virtual Ethnographer: Fieldwork in distributed electronic environments', *Field Methods*, 12, 1, pp. 3-17.

Sacks, H. (1992): *Lectures on Conversation. Vols. 1 & 2*. (Ed. G. Jefferson), Blackwell, Oxford.

Sonnenwald, D., Solomon, P., Hara, N., Bolliger, R. & Cox, T. (2002): 'Collaboration in the Large: Using video conferencing to facilitate large group interaction', in A. Gunasekaran, O. Khalil, & M. Syed (Eds.) *Knowledge and Information Technology Management in 21st Century Organizations*, Idea Group Publishing, Hershey, PA, pp. 115-136.

Yankelovich, N., McGinn, J., Wessler, M., Kaplan, J., Provino, J., & Fox, H. (2005): 'Private communications in public meetings', *Proc. CHI 2005*, ACM Press, New York, pp. 1873-1876.