

Video and the Analysis of Social Interaction. An interview with Christian Heath

Edited by Barbara Pentimalli and Andrea Spreafico

Christian Heath is Professor at King's College London and co-director of the Work, Interaction and Technology Research Centre. Drawing on Ethnomethodology and Conversation Analysis, he specialises in fine grained, video-based field studies of social interaction. He is currently undertaking research in settings that include auctions, control centres, operating theatres, and museums and galleries. His previous research involves a range of projects UK Research Councils and the European Commission in areas that include command and control, health care, the cultural industries, and advanced telecommunications. He has held positions at the Universities of Manchester, Surrey, and Nottingham and visiting positions at Universities and industrial research laboratories in the UK and abroad. He is an Fellow of the Academy of the Social Sciences (AcSS), a Freeman of the Worshipful Company of Art Scholars and in 2015 was given the EUSSET-IISI Lifetime Achievement Award presented to scholars for an outstanding contribution to the reorientation of the fields of computing and Informatics. His publications include: "The Dynamics of Auction: Social Interaction and the Sale of Fine Art and Antiques" (Cambridge 2013: awarded the Best Book Award in 2014 by the International Society for Conversation Analysis), "Video in Qualitative Research: Analysing Social Interaction in Everyday Life" (Sage with Hindmarsh, J. and P. Luff, 2010), "Technology in Action" (with P. Luff, Cambridge 2000), "Workplace Studies: Recovering Work Practice and Informing System Design" (Cambridge with Luff, P. and J. Hindmarsh 2000), "Body Movement and Speech in Medical Interaction" (Cambridge 1986) and numerous articles in journals and books. With Roy Pea and Lucy Suchman he is editor of the book series published by Cambridge University Press, Learning and Doing: Social, Cognitive and Computational Perspectives.



Professor Heath, the use of video recording is an established practice in various academic disciplines (e.g. anthropology, linguistics, sociology, psychology) and also in several interdisciplinary approaches. In your opinion, is there nowadays a proper sociological perspective in the use of video recordings or, instead, are we witnessing the formation of a disciplinary cross-bordering “community of scientific practices”? Ethnomethodology, for instance, is practiced in various forms by sociologists, linguistic anthropologists, linguists, social psychologists, and so on: in your experience as ethnomethodologist and “sociologist”, how do you describe, if there is one, a sociological ethnomethodological perspective?

Within sociology there is a wide diversity in the ways in which scholars use video recordings for research and fundamental differences in the methodological assumptions they bring to bear in using visual data. These differences encompass such matters as what constitutes data, the forms of analysis, and the character of explanation and theory. Even within for example Ethnomethodology and Conversation Analysis we find a variety of ways in which scholars use and analyse audio-visual recordings. These differences do not necessarily reflect the disciplinary background or commitment of the scholars in question and it is not unusual to find that particular ways of using and analysing audio-visual recordings breach disciplinary boundaries – for example particular sociologists, applied linguists and social psychologists may bring very similar methodological commitments to bear upon the ways in which they employ video recordings for analysis. With regard to our own research, much of which has focused on the workplace, we have had a long-standing concern to address the endogenous characteristics of activities as they are accomplished in and through social interaction within particular settings and environments.

Ethnomethodologists had studied different interaction contexts with the video recording, such as doctor-patient or user-copier interactions, control rooms of the metropolitan, rail and air traffic, museums, laboratories, operating theatres and several others. Do you believe that there are nowadays other contexts, little or nothing studied, that deserve to be considered? And why? Are there specific issues to which the contribution of visual ethnomethodology could bring significant results? Do you think that the exploration of the interactional organisation of human activity is offering contextual description and analysis that are limited to the particular social interaction we are studying or do you think that it's possible to deduce some wider teachings?

In the couple of decades, we have witnessed the emergence of a burgeoning corpus of Ethnomethodological/Conversation Analytic video based studies of social interaction. Research has addressed broad diversity of settings and activities including institutional environments in areas that include health care, markets, the cultural industries, and transport, as well as interaction in less

formal settings – for example in the home, the car, and museums and galleries. In one sense, we are at the beginnings of video-based studies of embodied or multimodal interaction even within research settings such as health care with which we have some familiarity. There remain numerous settings of which we know little, but more importantly the multimodal, interactional organisation of much human activity remains relatively unexplored.

Videorecordings of social interaction are always gathered within particular settings or environments. Their analysis provides an opportunity to explore two interrelated issues. On the one hand, what are ways in which the organisation of interaction enables the investigation of the characteristics of particular activities as they are accomplished with regard to the demands and constraints of particular task or setting. On the other hand, they also provide the resources with which to explore generic features of an activity's accomplishment and the ways in which particular forms of interactional organisation enable the concerted production of particular sequences and actions. Over some years, we have found that having a substantial data corpus of video-recordings of social interaction that includes a broad range of settings, some similar, some very different, provides an important resource to compare and contrast the organisation of particular actions or activities as they are accomplished on different occasions, within different circumstances (see for example Heath and Luff 2013).

In the study of social interactions and of the multimodal resources through which the participants make visible and accountable what they are doing during an ongoing action, what is the benefit for researchers to make a video recording instead of an exclusively “naked eye” observation (e.g. classical ethnography)? Within the trans-disciplinary community using video-based research is there actually an agreement concerning the necessity of doing fieldwork before audio-visual recording the social interactions?

It has long been recognised that audio-visual recordings of naturally occurring events provide unique opportunities for social science research and in particular research concerned with the organisation of social interaction. Unlike more conventional qualitative data, such as field observations, it enables scholars to capture versions of naturally occurring events and subject them to repeated scrutiny using slow motion facilities and the like. It provides an opportunity to share and discuss those materials with others and enables members of the research community to inspect for themselves the data on which observations and analyses are founded. It also allows researchers to build a corpus data on which various interests and concerns can be brought to bear, indeed interests and concerns that may not been envisaged when the original materials were collected. In the coming years, it is likely that we will find new

and distinctive ways of presenting video based studies that allows us to provide 'readers' and more broadly the research community with more powerful ways of experiencing analyses.

However, for many of the settings in which we undertake research, including more specialised setting in institutional environments, fieldwork is a prerequisite to the collection and analysis of audio-visual recordings. We have discussed this matter in more detail elsewhere (see Heath *et al.*, 2010). To very briefly summarise – field work provides the resources with which begin to become familiar with the characteristics of staffing and activities, with the tools, equipment, systems and so forth that feature in those activities, with the particularities of the ecology and its structure, to identify suitable and relevant locations for cameras and microphones, and not least, an opportunity to establish a working relationship with the participants themselves to secure their cooperation with audio-visual recording. The character of this fieldwork, its extent and focus, varies significantly between projects and settings and can depend in part on the analytic concerns and focus of the research. Moreover, it is not unusual, to undertake preliminary fieldwork and recording, and find that further fieldwork is necessary to clarify a range of matters or to gain access to particular activities or even domains that were previously unavailable.

Given the need to frame with the camera the various elements relevant and pertinent to the description and understanding of “what is going on” in a particular situation, what are the important choices about fixing or not the camera in one point, its possible position, its orientation and the number of cameras to be used in a specific context?

Gathering data, data that enables 'reasonable' access to the action and interaction of the participants always proves challenging and each and every setting poses distinctive demands and requires particular consideration. Indeed in our experience it is true to say that we find difficulties with almost all our initial recordings gathered within particular settings. Our own methodological commitments provide a broad framework of issues and concerns that inform the ways in which we collect data. We are primarily concerned with the collaborative, interactional production of activities and their sequential characteristics, securing reasonable access therefore to the visible and vocal features of the participants' actions, including their use of various tools and equipment, is critical. In some cases it proves necessary to use a single camera, in others we collect data simultaneously from multiple standpoints. So for example in a wide-ranging project, concerned with collaboration and control in the station operation rooms on rapid urban transport systems in London, Paris and Brussels, we found that, in most cases, a well-positioned single camera provided a rich and analytically fruitful data corpus (see Heath *et al.* 2002).

In a very different project, concerned with the interactional organisation of auctions of fine art and antiques, we soon realised that it was necessary to use three cameras – one camera primarily focused on the auctioneer and his/her assistant(s), the others focused on buyers and bidders including the sale assistants representing buyers on the phone (see for example Heath 2013). There are always certain limitations to recordings and particular angles and views constrain what sorts of actions, activities and phenomena will be accessible to analysis. Consider the following images from these different projects.

Figure 1: *A Station Operations Centre on London Underground*



The ecology and equipment of the control room, the operational characteristics of the transport network and the importance recording where possible the actions of the station supervisor sets certain demands on where, when and how to record. With some small variation the recording position we largely used is illustrated above. It was decided, at that time, that using more than one camera within severely constrained space would be overly obtrusive. The position and orientation of the one camera had to capture, as far as possible, the actions of the supervisor as well as the information sources on which he relied to produce activities, primarily interventions to regulate human traffic through the station. The video-recordings enable us to see and hear the actions of the supervisor including his use of various devices, such as the Public Address, the staff Radio and CCTV system, it also enabled us to see, the on screen material, including the CCTV images and traffic data, as well as, at least in part, the supervisor's view of the foyer (see for example Heath and Luff 2002).

Figure 2: *An Auction of Fine Art and Antiques*

Multiparty settings pose significant problems for data collection and it is often necessary to use multiple cameras. We normally find using a roaming camera, or multiple roaming cameras, proves highly problematic, obtrusive and results in much action in particular its initiation being inaccessible (see Heath *et al.* 2010). In recording auctions we had to vary the position and orientation of the cameras depending on the particular setting and number of arrangement of participants. Even then we found it was rarely possible to include everyone present within the scope of the cameras. The example we have included here shows a well-attended auction and one can see the challenges it poses for recording. Despite its limitations, the data proved very useful indeed, enabling us to examine the actions of the auctioneer and those of potential buyers as they prepare to bid, bid, and respond to the invitations of the auctioneer to issue further bids (see Heath 2013).

During the video sequences analysis process (data session), how the researchers could avoid to overlap their own interpretational categories of “what is happening there” to the sense that the participants in an interaction were mutually and cooperatively giving to what they were doing, seeing and making visible in that local context – through words, gestures, gazes, postures and the use of objects?

Sequence and the sequential character of action in interaction provide a critical resource for exploring the ways in which participants' themselves orient to and constitute the sense and significance of particular actions and activities. However, the embodied or multimodal production of action can pose important challenges to the ways in which we have traditionally warranted our analysis of particular actions. For example it can prove difficult to expose a demonstrable orientation to particular features of an action or activity. These challenges can be exacerbated when one considers for example the seemingly simultaneous co-production of particular actions. In the coming years, we suspect these challenges will become of increasing analytic significance to video based studies of social interaction.

How researchers made the choice to analyze certain micro-sequences rather than others? And what is the relevance and influence of this choice? In your experience of video analysis, have you ever find a new very important and surprising question completely unrelated to the initial research commitment (serendipity)?

It is sometimes suggested that the selection of particular fragments and the analytic focus will depend on the interests that one brings to bear in undertaking a particular project, previous research, commitments and the like. For my own part, in beginning to work on new data, I begin to look at the material and within a few minutes stumble across some action or activity that I then transcribe. Indeed, the problem is often looking beyond a collection of fragments drawn from the material and undertaking a thorough overview of the data corpus. This is not to suggest that having found an action, a sequence, an activity, a phenomenon, that looks potentially interesting, that the next step is not to review the materials and put together a candidate collection of whatever one thinks it is, or might be, at that stage of the analysis. However an extensive review and detailed cataloguing of all the data associated with a project can demand substantial resources and prematurely disregard potentially interesting phenomena within the materials.

Video-recordings, coupled with a methodological framework that drives analytic attention towards the fine-grained interactional accomplishment of social action, provides an important, if not unique opportunity, to discover phenomena that would otherwise pass unnoticed (see for example Heath and Luff 2000, Heath *et al.* 2010, Heath 2013). Consider for example a recent project led by vom Lehn (2013) that examined optometric consultations and in particular the interaction between optometrists and patients or customers. In the course of the analysis we discovered that optometrists employ a characteristic gesture when placing a trial lens in front of the eye of the patient, a sweeping movement in which the optometrist raises and immediately places the lens from above the patient's eye.

Figure 3: *Placing a Lens in Optometric Consultations*

We characterized the movement as a ‘professional gesture’ and found that it is designed to engender an immediate response to the effect of the lens rather than a consider reaction from the patient (see Webb *et al.* 2013). It proves critical to securing reliable results from the eye test. Surprisingly, while the gesture proved to be a recurrent way in which a lens is placed in front of the eye, across many different consultations, optometrists were unaware of the practice. When it was pointed out, they realized it is important feature of the way in which they generate reliable results.

Could you describe how your research group carries out the “data session” and how you “negotiate” the meaning of small gestural details (gazes, body gesture, and so on...) performed by the participants to social interactions that are evolving on the video screen?

The presentation of original extracts or fragments of videorecordings at seminars, colloquia and conferences, is a critical aspect of the ways in which we assess the quality and reliability of our observations and analyses. It provides the opportunity for members of the academic community to see and to comment upon the materials on which analysis is based and explores the cogency and persuasiveness of the underlying argument.

Within the Work, Interaction and Technology Research Centre, we hold regular data sessions in which extracts of video-recordings are presented and discussed in detail. These data analysis workshops are an important aspect of the ways in which we develop and assess our observations and subject our analyses to the scrutiny of others. We also hold frequent data analysis workshops with the members of research groups in both the UK and abroad undertaking related research. These workshops provide an important opportunity for faculty and students to present and discuss preliminary analyses as well as more developed studies. For example over the past year we have held regular workshops with colleagues and research groups from the Universities of Paris (Paristech), Trento, Fribourg, Basel and the Institut für Deutsche Sprache, Mannheim. They also provide the opportunity to present and discuss data and analyses with scholars with rather interests and in some cases

drawn from different disciplines. At the Centre at King's we also frequently host visits from academics from institutions in the UK and abroad and their participation proves invaluable during workshops and colloquia in which data is presented, discussed and analysed.

What difficulties involves to insert images in a scientific article so that readers are able to see the relevant details (gazes, gestures ...) that the participants in interaction and the sociologist have seen and have taken into account? Which is the role of images in theoretic argumentation?

There are significant challenges with the ways in which research based on the analysis of audio-visual materials is presented and illustrated in textual form. The various solutions that are commonly found in articles and monographs fail to provide a clear sense of the action, its sequential organization, or the significant characteristics of the participants' conduct (see for example Heath *et al.* 2010). Transcripts and images can appear self-serving and selective and frustrate the reader with detail. More frustrating still, is that in live presentations, an audience can enjoy and experience the complexities of particular fragments that when rendered in text can seem laboured and pedantic. Notwithstanding the practical, ethical and analytic challenges posed by the presentation of audio-visual fragments to a 'readership', in the coming years, if video based studies of social interaction are to gain traction within the social sciences that I believe they deserve, we need to address new and distinctive ways of providing our academic communities with a way of seeing and experiencing the data that is sensitive to the complexities of its analysis.

The analysis of video fragments implies the use of multimodal transcription. While conventions for the transcription of talk are well established, for the bodily behaviours researchers use a large variety of ways. Which kind of multimodal annotations do you use? How do you describe and categorise the different gestures, gaze orientations, bodily behaviours? Do you use different lines for the transcription of talk and for the description of bodily comportment? How do you insert images and video screenshots in the textual multimodal transcription? How do you put together different views of the same moment filmed by more than one camera?

Elsewhere we have discussed the transcription of multimodal action in detail and described the techniques that myself and other use to delineate aspects of bodily comportment and talk (see for example Heath *et al.* 2010; Luff and Heath 2015). We have also described the ways in which these techniques are modified to encompass the demands of particular activities and settings for example those that involve the use of complex technologies or involve multiple participants. We also discuss the ways in which integrate and

transcribe data from multiple sources. It is important to note that in video-based, multi-modal research of social interaction, the standing and use of transcripts is very different from more conventional forms of conversation and discourse analysis. The transcript is not used as independent representation of the action. During analysis the transcript routinely accompanies review and inspection of relevant sections of the videorecording. The transcript serves as analytic resource, as a way of exploring fragments of data and noting the interactional location and characteristics of particular actions and activities however slight or fleeting they might appear. The transcript is not used independently of the actual fragments, the relevant sections of data, either during analysis or in presenting material to colleagues or audiences. Due to the ethical and practical constraints of including fragments of video-recording with publications, it is not unusual to find highly simplified transcript, sometimes accompanied by still frames or drawings, in text. These are not examples of the transcripts developed and used within analysis, rather, highly selective and simplified characterisations of action to enable the reader to gain a gist of the action in question.

Do you think that in some cases it could be more favourable to make video recordings of social interactions without preventively informing the actors? Beyond the ethical question that encourages us to obtain the informed consent, don't you think that in some cases specific phenomena are only observable in a covert modality? In your research is the sociologist always present during the video recording or, once the video camera is positioned, can he leave the video camera filming alone?

The type of permission that is secured for collecting data, in particular video-recordings, varies significantly between different settings and types of activity that are being addressed. For example public and semi-public settings place very different constraints on data collection than more professional environments which personal information may be presented and discussed. Gaining 'informed consent' is a critical aspect of social science research and increasingly subject to the scrutiny of University and professional ethics committees. What constitutes 'informed' and 'consent' varies significantly between settings and activities. One of the concerns that is sometimes raised with regard to informed consent and the visibility of video-recording is that it can influence the conduct of the participants. In this regard, in reviewing any data corpus we have gathered, we scrutinise the materials to find, where possible, instances in which participants' themselves orient in some way to the recording and/or the presence of the researcher. These are then subject to analysis and used to reflect upon the robustness and naturalness of the data. We have developed techniques to avoid or lessen the obtrusiveness of record-

ing, one of which is not, where possible, to be seen operating the camera during data collection (see for example Heath *et al.* 2010).

References

- Heath C. (2013), *Dynamics of Auction: Social Interaction and the Sale of Fine Art and Antiques*, Cambridge University Press, Cambridge.
- Heath C.C. and Luff P. (2013), *Embodied Action and Multi-modal Interaction*, in Stivers T. and Sidnell J. (eds), *Handbook of Conversation Analysis*, Blackwell, Oxford.
- Heath C.C. and Luff P. (2000), *Technology in Action*, Cambridge University Press, Cambridge.
- Heath C.C., Hindmarsh J. and Luff P. (2010), *Video in Qualitative Research: Analysing Social Interaction in Everyday Life*, Sage, London.
- Heath C.C., Luff P. and Sanchez Svensson M. (2002), *Overseeing Organization*, in «British Journal of Sociology», 53, 2: 181-203.
- Luff P. and Heath C.C. (2015), *Transcribing Embodied Action*, in Schiffrin D., Saucier E., Tannen D. (eds.), *The Handbook of Discourse Analysis*, Second Edition, Wiley-Blackwell, Oxford.
- vom Lehn D., Gibson W., Webb H. and Heath C.C. (2013), *Assessing Distance Vision as Interactional Achievement: A Study of Commensuration in Action*, in «Soziale Welt», 64, 1-2: 115-136.
- Webb H., vom Lehn D., Heath C. and Gibson W. (2013), *Engendering Response: Professional Gesture and the Assessment of Eyesight in Optometry Consultations*, in «Symbolic Interaction», 36, 2: 137-158.