

## Practices of Seeing Visual Analysis: an Ethnomethodological Approach

CHARLES GOODWIN

### INTRODUCTION

A primordial site for the analysis of human language, cognition and action consists of a situation in which multiple participants are attempting to carry out courses of action together while attending to each other, the larger activities that their current actions are embedded within and relevant phenomena in their surround. Vision can be central to this process.<sup>1</sup> The visible bodies of participants provide systematic, changing displays about relevant action and orientation. Seeable structure in the environment can not only constitute a locus for shared visual attention, but can also contribute crucial semiotic resources for the organization of current action (consider, for example, the use of graphs and charts in a scientific discussion). For the past thirty years both conversation analysis and ethnomethodology have provided extensive analysis of how human vision is socially organized. Both fields investigate the practices that participants use to build and shape in concert with each other the structured events that constitute the lifeworld of a community of actors. Phenomena investigated in which vision plays a central role range from sequences of talk, through medical and legal encounters, to scientific knowledge.

The approach taken by both ethnomethodology and conversation analysis to the study of visual phenomena is quite distinctive. At least since Saussure proposed studying *langue* as an analytically distinct sub-field of a more encompassing science of signs, different kinds of semiotic phenomena (language, visual signs, etc.) have typically been analysed in isolation from each other. However in the work to be described here neither vision, nor the images or other phenomena that participants look at, are treated as coherent, self-contained domains that can be subjected to analysis in their own terms. Instead it quickly becomes apparent that visual phenomena can only be investigated by taking into account a diverse set of semiotic resources and meaning-making practices that participants deploy to build the social worlds that they inhabit and constitute through ongoing processes of action (see Chapter 9). Many of these, such as structure provided by current talk, are not in any sense visual, but the visible phenomena that the participants are attending to cannot be properly analysed without them. The focus of analysis is not thus representations or vision *per se*, but instead the part played by visual phenomena in the production of meaningful action.

Both the methodology and the forms of analysis used in this approach can best be demonstrated through specific examples.

### GAZE BETWEEN SPEAKERS AND HEARERS

In formulating the distinction between competence and performance, Chomsky (1965: 3–4) argued that actual speech is so full of performance errors, such as sentence fragments, restarts and pauses, that both linguists and parties faced with the task of acquiring a language should ignore it. Investigating a corpus of conversation recorded on video, C. Goodwin (1980, 1981: Chapter 2) indeed found precisely the ‘false starts’ and ‘changes of plan in mid-course’ that Chomsky describes. In the following, instead of producing an unbroken grammatical sentence, the speaker says:<sup>2</sup>

Cathy: En a couple of girls- *One* other girl from there,

However, as shown in Figure 8.1, when the video is examined it is found that the restart occurs at a specific place: precisely at the point where the speaker brings her gaze to her addressee and finds that her addressee is looking elsewhere. Moreover, the restart acts as a request for the hearer’s gaze. Thus immediately after the restart the hearer starts to move her gaze to the speaker. Paradoxically, if the speaker had not produced a restart at this point she could have said something that would appear to be an unbroken grammatical unit if one examined only the stream of speech (for example, ‘En a couple of girls from there . . .’), but which would in fact be interactively a sentence fragment since her addressee attended to only part of it.

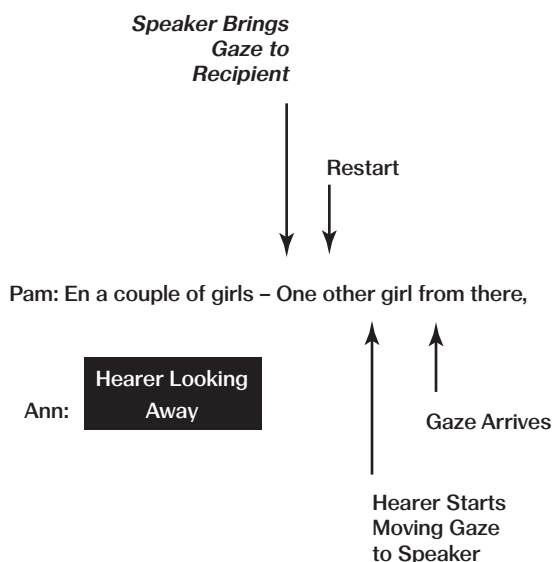


Figure 8.1 Gaze between speakers and hearers: transcript of Pam and Ann.

The identities of speaker and hearer are the most generic participant categories relevant to the production of a strip of talk. The phenomena examined here (which occur pervasively in conversation) provide evidence that the work of being a hearer in face-to-face interaction requires situated use of the body, and gaze in particular, as a way of visibly displaying to others the focus of one's orientation. Moreover speakers not only use their own gaze to see relevant action in the body of a silent hearer, but actively change the structure of their emerging talk in terms of what they see.

What relevance do processes such as these have to the other issue raised by Chomsky (1965: 3), that of determining 'from the data of performance the underlying system of rules mastered by the speaker-hearer'? Many repairs involve the repetition, with some significant change, of something said elsewhere in the utterance:

We wen t- I went to  
If he could- If you could

Such repetition has the effect of delineating the boundaries and structure of many different units in the stream of speech. Thus, by analysing what is the same and what is different in these examples one is able to discover the following: first, where the stream of speech can be divided into significant sub-units; second, that alternatives are possible in a particular slot; third, what some of these alternatives are (here different pronouns); and fourth, that these alternatives contrast with each other in some significant fashion, or else the repair would not be warranted. Repairs in other examples not only delineate basic units in the stream of speech (noun phrases, for example), but also demonstrate the different forms such units can take and the types of operations that can be performed upon them (see Goodwin, 1981: 170-3). Repairs further require that a listener learn to recognize that not all of the sequences within the stream of speech are possible sequences within the language, for example that 'I' does not follow 'to' in 'We went t- I went to . . .'. In order to deal with such a repair a hearer is thus required to make one of the most basic distinctions posed for anyone attempting to decipher the structure of a language: to differentiate what are and are not possible sequences in the language, that is between grammatical and ungrammatical structures. The fact that this task is posed may be crucial to any learning process. If the party attempting to learn the language did not have to deal with ungrammatical possibilities, if for example she was exposed to only well-formed sentences, she might not have the data necessary to determine the boundaries, or even the structure of the system. Chomsky's argument that the repairs found in natural speech so flaw it that a child is faced with data of very 'degenerate quality' does not appear warranted. Rather it might be argued that if a child grew up in an ideal world where she heard only well-formed sentences she would not learn to produce sentences herself because she would lack the analysis of their structure provided by events such as repair. Crucial to this process is the way in which visual phenomena, such as dispreferred gaze states, can both lead to repair and demonstrate that the participants are in fact attending in fine detail to what might appear to be quite ephemeral structure in the stream of speech.

What has just been described provides one example of the methodologies and forms of analysis used to investigate visual phenomena within conversation analysis.

Several observations can be made. First, the focus of analysis is not visual events in isolation, but instead the systematic practices used by participants in interaction to achieve courses of collaborative action with each other – in the present case the interactive construction of turns at talk and the utterances that emerge with those turns. Visual events, such as gaze, play a central role in this process but their sense and relevance is established through their embeddedness in other meaning-making tasks and practices, such as the production of a strip of talk that is in fact heard and attended to by its addressee. This links vision to a host of other phenomena, including language and the visible body, as an unfolding locus for the display of meaning and action.

Second, what the analyst seeks to do is not to provide his or her own gloss on how visual phenomena might be meaningful, but instead to demonstrate how the participants themselves not only actively orient to particular kinds of visual events (such as states of gaze), but use them as a constitutive feature of the activities they are engaged in (for example, by modifying their talk in terms of what they demonstrably see).

Third, in addition to the spatial dimension that is naturally associated with vision, these processes also have an intrinsic temporal dimension as changes in visual events are marked by, and lead to, ongoing changes in the organization of emerging action. If one had only a static snapshot, or measured only a single structural possibility, such as mutual gaze, instead of looking at the temporally unfolding interplay of different combinations of participant gaze, the type of analysis being pursued here would be impossible.

Fourth, such analysis requires data of a particular type, specifically a record that maintains as much information as possible about the setting, embodied displays and spatial organization of all relevant participants, their talk, and how events change through time. In practice no record is completely adequate. Every camera position excludes other views of what is happening. The choice of where to place the camera is but the first in a long series of crucial analytical decisions (see Chapter 9). Despite these limitations a video or film record does constitute a relevant data source, something that can be worked with in an imperfect world (see Chapter 3).

Fifth, crucial problems of transcription are posed. The task of translating the situated, embodied practices used by participants in interaction to organize phenomena relevant to vision poses enormous theoretical and methodological problems. Our ability to transcribe talk is built upon a process of analysing relevant structure in the stream of speech and marking those distinctions with written symbols, that extends back thousands of years and is still being modified today (for example, the system developed by Gail Jefferson (Sacks et al., 1974: 731–3) for transcribing the texture of talk-in-interaction, including phenomena, such as momentary restarts and sound stretches, that are crucial for the analysis being reported here). When it comes to the transcription of visual phenomena we are at the very beginning of such a process. The arrows and other symbols I've used to mark gaze on a transcript (see Goodwin, 1981) capture only a small part of a larger complex constituted by bodies interacting together in a relevant setting. The decision to describe gaze in terms of the speaker–hearer framework is itself a major analytic one, and by no means simple, neutral description. Moreover a gazing head is embedded within a larger postural configuration, and indeed different parts of the body can simultaneously display orientation to different

participants or regions (see Kendon, 1990b; Schegloff, 1998), creating 'participation frameworks' of considerable complexity. Thus on occasion a transcriber wants some way of indicating on the printed page posture and alignment. In addition, not only the bodies of the participants, but also phenomena in their surround, can be crucial to the organization of their action. To try to make the phenomena I'm analysing independently accessible to the reader so that she or he can evaluate my analysis, I've experimented with using transcription symbols, frame grabs, diagrams and movies embedded in electronic versions of papers. Multiple issues are involved and no method is entirely successful. On the one hand the analyst needs materials that maintain as much of the original structure of the events being analysed as possible, and which can be easily and repetitively replayed. On the other hand, just as a raw tape recording does not display the analysis of segmental structure in the stream of speech provided by transcription with a phonetic or alphabetic writing system, in itself a video, even one that can be embedded within a paper, does not provide an analysis of how visible events are being parsed by participants. The complexity of the phenomena involved requires multiple methods for rendering relevant distinctions (e.g. accurate transcription of speech, gaze notation, frame grabs and diagrams; see also Ochs, 1979). Moreover, like the two-faced Roman god Janus, any transcription system must attend simultaneously to two separate fields, looking in one direction at how to accurately recover through a systematic notation the endogenous structure of the events being investigated, while simultaneously keeping another eye on the addressee/reader of the analysis by attempting to present relevant descriptions as clearly and vividly as possible. In many cases different stages of analysis and presentation will require multiple transcriptions. There is a recursive interplay between analysis and methods of description.

Work in conversation analysis has provided extensive study of how the gaze of participants toward each other is consequential for the organization of action within talk-in-interaction. Phenomena investigated include the way in which speakers change the structure of an emerging utterance, and the sentence being constructed within it, as gaze is moved from one type of recipient to another, so that the utterance maintains its appropriateness for its addressee of the moment (Goodwin, 1979, 1981); how speakers modify descriptions in terms of their hearer's visible assessment of what is being said (M.H. Goodwin, 1980); how genres such as stories are constructed not by a speaker alone, but instead through the differentiated visible displays of a range of structurally different kinds of recipients (speaker, primary addressee, principal character, etc. – see Goodwin, 1984); the organization of gaze and co-participation in medical encounters (Heath, 1986; Robinson, 1998); the interactive organization of assessments (Goodwin and Goodwin, 1987); gesture (Goodwin, 2000; Streeck, 1993, 1994); the use of gaze in activities such as word searches (Goodwin and Goodwin, 1986); and so on. Though not strictly lodged within conversation analysis the work of Kendon (1990a, 1994, 1997) on both the interactive organization of bodies as they frame states of talk and on gesture is central to the study of visible behaviour in interaction. Haviland (1993) provides important analysis of the interactive organization of gesture within narration (for extensive analysis of gesture from a psychological perspective, see McNeill, 1992).

## SCIENTIFIC IMAGES

The visible, gazing body, and the orientation of participants toward each other as they co-produce states of talk, is central to the work in conversational analysis just examined. By way of contrast much work within ethnomethodology has focused not on the bodies of actors, but instead on the images, diagrams, graphs and other visual practices used by scientists to construct the crucial visual working environments of their disciplines. As noted by Lynch and Woolgar:

Manifestly, what scientists laboriously piece together, pick up in their hands, measure, show to one another, argue about, and circulate to others in their communities are not 'natural objects' independent of cultural processes and literary forms. They are extracts, 'tissue cultures,' and residues impressed within graphic matrices; ordered, shaped, and filtered samples; carefully aligned photographic traces and chart recordings; and verbal accounts. These are the proximal 'things' taken into the laboratory and circulated in print and they are a rich repository of 'social' actions. (1990: 5)

Despite important differences in subject-matter and methodology, both fields emphasize the importance of focusing not on representations or other visual phenomena as self-contained entities in their own right, but instead on how they are constructed, attended to and used by participants as components of the endogenous activities that make up the lifeworld of a setting. Thus, in introducing their important volume *Representation in Scientific Practice*, Lynch and Woolgar define their enquiry as follows: 'Instead of asking "what do we mean, in various contexts, by 'representation'?" the studies begin by asking, "What do the participants, *in this case*, treat as representation?"' (1990: 11). Note that what must be investigated is specified both in terms of the orientation of the participants and with respect to the features of the relevant local setting (for example, '*in this case*'). This leads to a distinctive ethnomethodological perspective on *reflexivity*: "Reflexivity" in this usage means, not self-referential nor reflective awareness of representational practice, but the inseparability of a "theory" of representation from the heterogeneous social contexts in which representations are composed and used' (Lynch and Woolgar, 1990: 12).

In a classic article Lynch formulates the task of analysing scientific representations as that of describing the publicly visible 'externalized retina' that is the site for the practices implicated in the social constitution of the objects that are the focus of scientific work:

This study is based on the premise that visual displays are more than a simple matter of supplying pictorial illustrations for scientific texts. They are essential to how scientific objects and orderly relationships are revealed and made analyzable. To appreciate this, we first need to wrest the idea of *representation* from an individualistic cognitive foundation, and to replace a preoccupation with images on the retina (or alternatively 'mental images' or 'pictorial ideas') with a focus on the 'externalized retina' of the graphic and instrumental fields upon which the scientific image is impressed and circulated. (1990: 153–4)

Using as data images from scientific journal articles and books, Lynch describes two families of practices used to constitute the visible scientific object: 'selection'

and ‘mathematization’. Selection, illustrated through double images in which a photograph and a diagram of entities visible in the photograph are presented side by side, is described as a host of practices that iteratively transform one image of an entity into another (for example, the photograph to the diagram) while simultaneously structuring and shaping what it is that is being represented. Crucial to this process is the fact that different selective/shaping practices, including filtering, uniforming, upgrading and defining, can be repetitively applied creating not just a single image, but a linked, directional chain of representations. Indeed much of the work of actually doing science consists in building and shaping what Latour (1986) has called *inscriptions* in this fashion (see also Latour and Woolgar, 1979). ‘Mathematization’ refers not simply to the use of numbers, but instead to the host of practices used to transform recalcitrant events into mathematically tractable visual and graphic displays, for example graphs, charts and diagrams. Thus an image showing a map of lizard territories is assembled through, among other operations, driving stakes into the lizards’ environment to create a grid for measurement (and thus injecting a scientifically relevant Cartesian space into the very habitat being studied), repetitively capturing lizards, distinguishing them from each other by cutting off a different pattern of toes on each lizard, recording each capture on a paper map of the staked out territory, and finally drawing lines around collections of points to create the map. As noted by Lynch the product of these practices, for example the published map, ‘is a hybrid object that is demonstrably mathematical, natural and literary’ (1990: 171). Note how in all of these cases the focus of analysis is on the contextually based practices of the participants who are assembling and using these images to accomplish the work that defines their profession.

Though emerging from psychological anthropology, rather than ethnomethodology, Hutchins’s (1995) ground-breaking study of the cognitive practices required to navigate a ship outlines a major perspective for the analysis of both images and seeing as forms of work-relevant practice. Hutchins demonstrates how the practices required to navigate a ship are not situated within the mental life of a single individual, but are instead embedded within a distributed system that encompasses visual tools, such as maps and instruments for juxtaposing a landmark and compass-bearing within the same visual field and actors in structurally different positions who use alternative tools and, in part because of this, perform different kinds of cognitive operations, many of which have a strong visual component (e.g. locating landmarks and plotting positions on a map).

## IMAGES IN INTERACTION

All of the work discussed so far takes as its point of departure for the investigation of visual phenomena the task of describing and analysing the practices used by participants to construct the actions and events that make up their lifeworld. Rather than standing alone as a self-contained analytic domain, visual phenomena are constituted and made meaningful through the way in which they are embedded within this larger set of practices. However, within this common focus, two quite different orders of

visual practice have been examined. Research in science studies has investigated the images produced by scientists, and the way in which they visually and mathematically structure the world that is the focus of their inquiry, without however looking in much detail at how scientists attend to each other as living, meaningful bodies, or structure what they are seeing through the organization of talk-in-interaction. By way of contrast, studies of the interactive organization of vision in conversation looked in considerable detail at how participants treat the visual displays of each other's bodies as consequential, and how this is relevant to the moment-by-moment production of talk, but did not focus much analysis on images in the environment. Clearly all of the phenomena noted – the visible body, participation, gesture, the details of talk and language use, visual structure in the surround, images, maps and other representational practices, the public organization of visual practice within the worklife of a profession, and so on – are relevant. The question arises as to whether it is possible to analyse such disparate phenomena within a coherent analytic framework.

Before turning to studies that have probed such questions several issues must be noted. First, it is clearly not the case that the only acceptable analysis is one that includes this full range of all possible visual phenomena. Both participants and the structures that provide organization for action and events use visible phenomena selectively. Parties speaking over the telephone can see neither each other's bodies nor events in a common surround. A scientific journal can be read in the absence of the parties who constructed its text and diagrams. More interestingly within face-to-face interaction participants can continuously shift between actions that invoke, and perhaps require, gaze towards specific events in the surround and those that make relevant gaze towards no more than each other's bodies. Even in this more limited case there may be a real issue as to whether it is relevant to attend to everything that a body does; for example, some gestures made by a speaker may not require gaze toward them from an addressee. There is thus an essential contingency, not only for the analyst but more crucially for the participants themselves, as to what sub-set of possible visual events are in fact relevant to the organization of the actions of the moment. Moreover, this means that in addition to investigating how different kinds of visible phenomena are organized, the analyst must also take into account how participants show each other what kinds of events they are expected to take into account at a particular moment, for example to indicate that a participant, gesture or entity in the surround should be gazed at. There is thus not only communication through vision, but also ongoing communication about relevant vision (Goodwin, 1981, 1986, in preparation; Streeck, 1988).

Second, visual events are quite heterogeneous, not only in what they make visible, but more crucially in their structure. Consider for example the issue of temporality. Both gestures and the displays of postural orientation used to build participation frameworks are performed by the body within interaction. However, while gestures, like the bits of talk they accompany, are typically brief (for example, they frequently fall within the scope of a single utterance) and display semantic content relevant to the topic of the moment, participation displays frame extended strips of talk and typically provide information about the participants' orientation rather than the specifics of what is being discussed. Bodily displays with one kind of temporal



duration (and information content) are thus embedded within another class of visual displays being made by the body which have a quite different structure.

Third, the structure of visual signs, including their possibilities for propagation through space and time, can be intimately tied to the medium used to construct them. A major theme of Shakespeare's sonnets focuses on the contrast between the temporally constrained human body, condemned to inevitable decay, and the (limited) possibilities for transcending such corruption provided by language inscribed on the printed page which can remain fresh and alive long after its author and subject have passed into dust. This contrast between the temporal possibilities provided by alternative media (e.g. the body and documents) constitutes an ongoing resource for participants in vernacular settings as they build, through interaction with each other, the events that make up their lifeworld. In addition to the displays made by a fleeting gesture or local participation framework, participants also have access to images and documents which can encompass multiple interactions and quite diverse settings. This arises in part from the specific media used to constitute the signs they contain. Rather than being lodged within an ever-changing human body, such documents constitute what Latour has called immutable mobiles, portable material objects that can carry stable inscriptions of various types from place to place and through time (1987: 223).

However, despite the way in which crucial aspects of the structure of images and documents remain constant in different environments, they are not self-contained visual artefacts that can be analysed in isolation from the processes of interaction and work practices through which they are made relevant and meaningful. The same image or document can be construed in quite different ways in alternative settings. For example, a schedule listing all arriving and departing flights was a major tool for almost all work groups at the airport studied by the Xerox PARC workplace project (Brun-Cottan et al., 1991; Goodwin and Goodwin, 1996; Suchman, 1992), and indeed it linked diverse workers throughout North America into a common web of activity. However while baggage loaders carefully structured their work to anticipate arriving flights, so that planes could be speedily unloaded, these same arrival times were almost ignored by gate agents looking at the same schedule, but concerned with the departure of passengers. Each work group highlighted the common document in ways relevant to the specific work tasks it faced.

Similarly, on the oceanographic ship reported in Goodwin (1995), a map showing where samples would be taken in the Atlantic at the mouth of the Amazon was a major document at all stages of the research project. Before the ship sailed the places where samples could be taken were the focus of intense political debate between different groups of scientists and the Brazilian and American governments. After the project was completed the map provided an infrastructure for graphic displays that could be used in published journal articles to show what the scientists had found about how the waters of the Amazon and the Atlantic interacted with each other, that is, a way of making visible relevant scientific phenomena. During the voyage itself the map not only provided a common framework for the quite different work of various teams of scientists and the crew navigating the ship, but could also be looked at by lab technicians, not able to go to bed for days at a time because of the map's incessant

sampling demands, to locate places where stations were far apart and rest was possible. In brief, though the material form of images and documents gives them an extended temporal scope, and the ability to travel from setting to setting, they cannot be analysed as self-contained fields of visually organized meaning, but instead stand in a reflexive relationship to the settings and processes of embodied human interaction through which they are constituted as meaningful entities. To explicate such events analysis must deal simultaneously with the quite different structure and temporal organization of both local embodied practice and enduring graphic displays.

Finally, the visual (and other) properties of settings structure environments that shape, on an historical time-scale, the activities systematically performed within those settings. A very simple example is provided by the bridge of the oceanographic ship which not only had a window facing forward so the helmsman could steer the ship and watch for trouble, but also a window facing backwards. This was used by a winch operator who had the task of lifting heavy instrument packages in and out of the sea. Though being used here to do science, this arrangement is in fact a systematic solution to a repetitive problem faced by sailors, such as fishermen using nets, who have to manoeuvre heavy objects while at sea. Solutions found to these tasks, such as the rear-facing window with the visual access it provides (as well as the forward window facilitating navigation), are built into the tools that constitute the work environments used by subsequent actors faced with similar tasks. See Hutchins (1995) for illuminating analysis of this process, including tools that visually structure complex mathematical calculations, as well as maps. Both work environments and many of the tools used within them (computer displays, etc.) structure in quite specific ways the embodied visual practices of those who inhabit such settings.

In an attempt to come to terms with such issues Goodwin (2000) has proposed that images in interaction are lodged within endogenous activity systems constituted through the ongoing, changing deployment of multiple 'semiotic fields' which mutually elaborate each other. The term semiotic field is intended to focus on signs-in-their-media, that is the way in which what is typically been attended to are sign phenomena of various types (gestures, maps, displays of bodily orientation, etc.) which have variable structural properties that arise in part from the different kinds of materials used to make them visible (for example, the body, talk, documents, etc.). Bringing signs lodged within different fields into a relationship of mutual elaboration produces locally relevant meaning and action that could not be accomplished by one sign system alone. Consider for example a place on a map indicated by a pointing finger which is being construed in a specific fashion by the accompanying talk. Not the map as a whole, that is a self-sufficient representation, nor the pointing finger in isolation from (a) its target (the spot on the map) and (b) the construal being provided by the talk, nor the talk alone, would be sufficient to constitute the action made visible by the conjoined use of the three semiotic fields, each of which provides resources for specifying how to relevantly see and understand the others (see the brief discussion of the Rodney King data below for a specific example; see Goodwin (in preparation) for more detailed analysis of pointing). The particular sub-set of semiotic fields available in a setting that participants orient to as relevant to the construction of the actions of the moment constitutes a 'contextual configuration'. As interaction unfolds

contextual configurations can change as new fields are added to, or dropped from, the specific mix being used to constitute the events of the moment. Thus, as contextual configurations change there is both unfolding public semiotic structure and contingency (and, indeed, in some circumstances actions can misfire when addressees fail to take into account a relevant semiotic field, such as the sequential organization provided by a prior unheard utterance – see Goodwin (in preparation) for an example).

## PROFESSIONAL VISION

Work settings provide one environment in which the interplay between situated, embodied interaction and the use of visual images of different types can be systematically investigated. In many work settings participants face the task of classifying visual phenomena in a way that is relevant to the work they are charged with performing. Frequently they must also construct different kinds of representations of visual structure in the environment that is the focus of their professional scrutiny. I will now briefly examine how such vision is socially organized in two tasks faced by archaeologists: first, colour classification and, second, map-making. Then I will look at how such professional vision was both constructed and contested in the trial of four policemen charged with beating an African-American motorist, Mr Rodney King, where the key evidence was a videotape of the beating.

### Colour classification as historically structured professional practice

As part of the work involved in excavating a site, archaeologists make maps showing relevant structure in the layers of dirt they uncover. In addition to *artefacts*, such as stone tools, archaeologists are also interested in *features*, such as the remains of an old hearth or the outlines of the posts that held up a building. Such features are typically visible as colour differences in the dirt being examined (for example, the remains of a cooking fire will be blacker than the surrounding soil, while the holes used for posts will also have a different colour from the soil around them). Field archaeologists thus face the task of systematically classifying the colour of the dirt they are excavating. The methods they use to accomplish this task constitute a form of professional visual practice. As demonstrated by the discussion of Lynch's analysis of scientific representation, and the brief description of the oceanographers, crucial work in many different occupations takes the form of classifying and constructing visual phenomena in ways that help shape the objects of knowledge that are the focus of the work of a profession (for example, architects, sailors plotting courses on charts, air traffic controllers and professors making graphs and overheads for talks and classes). Such professional vision constitutes a perspicuous site for systematic study of how different kinds of phenomena intersect to organize a community's practices of seeing.

Goodwin (1996, 2000) describes how archaeologists code the colour of the dirt they are excavating through use of a Munsell chart. Figure 8.2 shows two archaeologists

performing this task: the Munsell page that they are using and the coding form where they will record their classification. Within this scene are a number of different kinds of phenomena relevant to the organization of visual practice, including tools that structure the process of seeing and classification and documents that organize cognition and interaction in the current setting while linking these processes to larger activities and other settings. These archaeologists are intently examining the colour of a tiny sample of dirt because they have been given a coding form to fill out. That form ties their work at this site to a range of other settings, such as the offices and lab of the senior investigator, where the form being filled in here will eventually become part of the permanent record of the excavation and a component of subsequent analysis. The multivocality of this form, the way in which it displays on a single surface the actions of multiple actors in structurally different positions, is shown visually in vivid fashion by the contrast between the printed coding categories and the hand-written entries of the field workers.

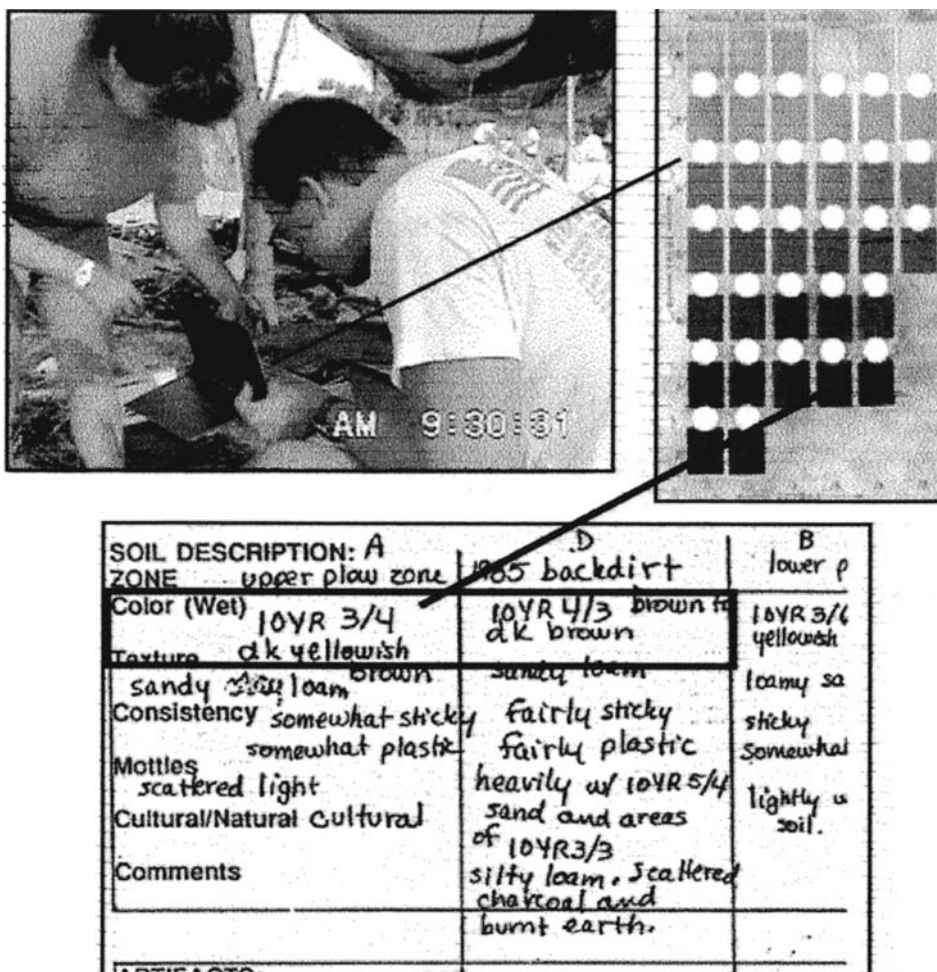


Figure 8.2 Transcript of archaeologists coding and recording the colour of the dirt they are excavating through use of a Munsell chart.

The use of a coding form such as this to organize the perception of nature, events or people within the discourse of a profession carries with it an array of perceptual and cognitive operations that have far-reaching impact. Coding schemes distributed on forms allow a senior investigator to inscribe his or her perceptual distinctions into the work practices of the technicians who code the data. By using such a system a worker views the world from the perspective it establishes. Of all the possible ways that the earth could be looked at, the perceptual work of field workers using this form is focused on determining the exact colour of a minute sample of dirt. They engage in active cognitive work, but the parameters of that work have been established by the classification system that is organizing their perception. Insofar as the coding scheme establishes an orientation toward the world, a work-relevant way of seeing, it constitutes a structure of intentionality whose proper locus is not the isolated, Cartesian mind, but a much larger organizational system, one that is characteristically mediated through mundane bureaucratic documents such as this form.

Rather than standing alone as self-explicating textual objects, forms are embedded within webs of socially organized, situated practices. In order to make an entry in the slot provided for colour an archaeologist must make use of another tool, the set of standard colour samples provided by a Munsell chart. This chart incorporates into a portable physical object the results of a long history of scientific investigation of the properties of colour.

The Munsell chart being used by the archaeologists contains not just one but three different kinds of sign systems for describing each point in the colour space it provides:

- 1 A set of carefully controlled *colour samples* arranged in a grid to demonstrate the changes that result from systematic variation of the variables of Hue, Chroma and Value used to define each colour (each page displays an ordered set of Value and Chroma variables for a single hue).
- 2 *Numeric co-ordinates* for each row and column; their intersection specifies each square as a pair of numbers (such as 4/6 on the 10YR Hue page).
- 3 *Standard colour names* such as 'dark yellowish brown' (these names are on the left-facing page which is not reproduced here). Moreover these systems are not precisely equivalent to each other. For example, several colour squares can fall within the scope of a single name.

Why does the Munsell page contain multiple, overlapping representations of what is apparently the same visual entity (for example, a particular choice within a larger set of colour categories)? The answer seems to lie in the way that each representation as a semiotic field with its own distinctive properties makes possible alternative operations and actions, and thus fits into different kinds of activities. Both the names and numbered grid co-ordinates can be written, and thus easily transported from the actual excavation to the other work sites, such as laboratories and journals, that constitute archaeology as a profession. The numbers provide the most precise description and do not require translation from language to language. However locating the colour indexed by the co-ordinates requires that the classification be read with a Munsell book at hand. By way of contrast the colour names can be grasped in a way

that is adequate for most practical purposes by any competent speaker of the language used to write the report. The outcome of the activity of colour classification initiated by the empty square on the coding form is thus a set of portable linguistic objects that can easily be incorporated into the unfolding chains of inscription that lead step by step from the dirt at the site to reports in the archaeological literature. However, as arbitrary linguistic signs produced in a medium that does not actually make visible colour, neither the colour names nor the numbers allow direct visual comparison between a sample of dirt and a reference colour. This is precisely what the colour patches and viewing holes make possible. Thus, rather than simply specifying unique points in a larger colour space, the Munsell chart is used in multiple overlapping activities (comparing a reference colour and a patch of dirt as part of the work of classification, transporting those results back to the lab, comparing samples, publishing reports, etc.), and so represents the 'same' entity, a particular colour, in multiple ways, each of which makes possible different kinds of operations because of the unique properties of each representational system.

In addition to its various sign systems the chart also contains a set of circular holes, positioned so that one is adjacent to each colour patch. To classify colour the archaeologist puts a small sample of dirt on the tip of a trowel, puts the trowel directly under the Munsell page and then moves it from hole to hole until the best match with an adjacent colour sample is found. With elegant simplicity the Munsell page, with its holes for viewing the sample of dirt on the trowel, juxtaposes in a single visual field two quite different kinds of spaces: (a) actual dirt from the site at the archaeologists' feet is framed by (b) a theoretical space for the rigorous, replicable classification of colour. The latter is both a conceptual space, the product of considerable research into properties of colour, and an actual physical space instantiated in the orderly modification of variables arranged in a grid on the Munsell page. The pages juxtaposing colour patches and viewing holes that allow the dirt to be seen right next to the colour sample provide an historically constituted architecture for perception, one that encapsulates in a material object theory and solutions developed by earlier workers at other sites faced with the task of colour classification. By juxtaposing unlike spaces, but ones relevant to the accomplishment of a specific cognitive task, the chart creates a new, distinctively human, kind of space. It is precisely here, as bits of dirt are shaped into the work-relevant categories of a specific social group, that 'nature' is transformed into culture.

How are the resources provided by the chart made visible and relevant within talk-in-interaction? At line 17 Pam moves her hand to the space above the Munsell chart and points to a particular colour patch while saying 'En this one'. Within the field of action created by the activity of colour classification, what Pam does here is not simply an indexical gesture, but a proposal that the indicated colour might be the one they are searching for. By virtue of such conditional relevance (Schegloff, 1968) it creates a new context in which a reply from Jeff is the expected next action. In line 19 Jeff rejects the proposed colour. His move occurs after a noticeable silence in line 18. However that silence is not an empty space, but a place occupied by its own relevant activity. Before a competent answer to Pam's proposal in line 17 can be made, the dirt being evaluated has to be placed under the viewing hole next to the sample she

indicated, so that the two can be compared. During line 18 Jeff moves the trowel to this position. Because of the spatial organization of this activity, specific actions have to be performed before a relevant task, a colour comparison, can be competently performed. In brief, in this activity the spatial organization of the tools being worked with and the sequential organization of talk-in-interaction interact with each other in the production of relevant action (e.g. getting to a place where one can make an expected answer requires rearrangement of the visual field being scrutinized so that the judgement being requested can be competently performed). Here socially organized vision requires embodied manipulation of the environment being scrutinized.

It is common to talk about structures such as the Munsell chart as ‘representations’. However exclusive focus on the representational properties of such structures can seriously distort our understanding of how such entities are embedded within the organization of human practice. With its viewholes for scrutinizing samples, the page is not simply a perspicuous representation of current knowledge about the organization of colour, but a space designed for the ongoing production of particular kinds of *action*.

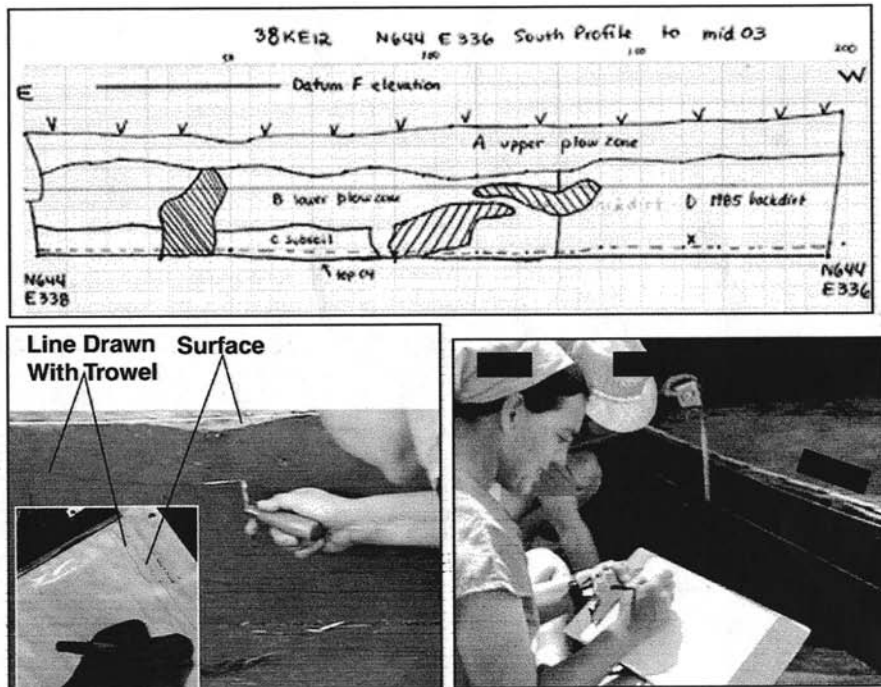
We will now look at how a group of archaeologists make a map. This process will allow us to examine the interface between seeing, writing practices, talk, human interaction and tool use (see Goodwin, 1994 for more detailed analysis).

## Map-making and the practices of seeing it requires

Maps are central to archaeological practice. The professional seeing required to produce and make use of a visual document, such as a map, encompasses not only the image itself but also the ability to competently see relevant structure in the territory being mapped, mastery of appropriate tools and, on occasion, the ability to analyse the work-relevant actions of another’s body. These different kinds of phenomena can be brought together within the temporally unfolding process of human interaction used to accomplish the activity of making a map.

In the following example, two archaeologists are making a map to record what they have found in a profile of the dirt on the side of one of the square holes they have excavated. Before actually setting pen to paper some relevant events in the dirt, such as the boundary between two different kinds of soil, are highlighted by outlining them with the tip of a trowel. The structure visible in the dirt is then mapped on a sheet of graph paper. Typically this task is done by two participants working together. One uses a pair of rulers (one laid horizontally on the surface and the other a hand-held tape-measure used to measure depth beneath the surface) to measure the length and depth co-ordinates of the points in the dirt that are to be transferred to the map, and then speaks these co-ordinates as pairs of numbers (e.g. ‘at fifteen three point two’). The second person plots the points specified on the graph paper, and draws lines between successive measurements. What we find here is a small activity system that encompasses talk, writing, tools and distributed cognition as two parties collaborate to inscribe events they see in the earth on to paper. In the transcript in Figure 8.3, Ann, the party drawing the map, is the senior archaeologist at the site, while Sue, the person making

measurements, is her student. The sequence to be examined begins with a directive: Ann, the writer, tells Sue, the measurer, to 'Give me the ground surface over here to about *ninety*'. However before Sue has produced any numbers, indeed before she has said anything whatsoever, Ann in lines 4 and 5 challenges her, telling her that what she is doing is wrong: 'No- No- Not *at* ninety. From you *to* about ninety'.



- 1 Ann: Give me the ground surface over here  
 2 to about *ninety*.  
 3 (1.6)  
 4 Ann: No- No- Not *at* ninety.  
 5 From you to about ninety.  
 6 (1.0)  
 7 Sue: °Oh.  
 8 Ann: Wherever there's a change in slope.  
 9 (0.6)  
 10 Sue: °Mm kay.  
 11 Ann: See so if its *fairly flat*  
 12 I'll need one  
 13 where it *stops* being fairly flat.  
 14 Sue: [Okay.  
 15 Ann: Like right there.

**Figure 8.3** Transcript of Ann (the senior archaeologist at the site) drawing a map, and Sue (her student) making measurements.



Directives are a classic form of speech action that sociolinguists have used to probe the relationship between language and social structure, and in particular issues of power and gender. Here Sue formats both her directive and her correction in very strong, direct 'aggravated' fashion. No forms of mitigation are found in either utterance, and Ann is not given an opportunity to find and correct the trouble on her own. Directives formatted in this fashion have frequently been argued to display a hierarchical relationship: Ann is treating Sue as someone that she can give direct, unmitigated orders to. And indeed Ann is a professor and Sue is her student.

Issues of power do not however exhaust the social phenomena visible in this sequence. Equally important are a range of *cognitive processes* that are as socially organized as the relationships between the participants. For example, in that Sue has not produced an answer to the directive, how can Ann see that there is something wrong with a response that has not even occurred yet? Crucial to this process is the phenomenon of *conditional relevance* first described by Schegloff (1968). Basically a first utterance creates an interpretative environment that will be used to analyse whatever occurs after it. Here no subsequent talk has yet been produced. However, providing an answer in this activity system encompasses more than talk. Before speaking the set of numbers that counts as a proper next bit of talk, Sue must first locate a relevant point in the dirt and measure its co-ordinates. Both her movement through space, and her use of tools such as a tape-measure, are visible events. As Ann finishes her directive, Sue is holding the tape-measure against the dirt at the left or zero end of the profile. However, just after hearing 'ninety' Sue moves both her body and the tape-measure to the right, stopping near the '90' mark on the upper ruler. By virtue of the field interpretation opened up through conditional relevance, Sue's movement and tool use can now be analysed by Ann as elements of the activity she has been asked to perform, and found wanting. Sue has moved immediately to ninety instead of measuring the relevant points between zero and ninety. The sequential framework created by a directive in talk thus provides resources for analysing and evaluating the visible activity of an addressee's body interacting with a relevant environment.

Additional elements of the cognitive operations and kinds of seeing that Ann requires from Sue in order to make her measurements are revealed as the sequence continues to unfold. Making the relevant measurements presupposes the ability to locate where in the dirt measurements should be made. However Sue's response calls this presupposition into question and leads to Ann telling her explicitly, in several different ways, what she should look for in order to determine where to measure. After Ann tells Sue to measure points between zero and ninety, Sue does not immediately move to points in that region but instead hesitates for a full second before replying with a weak 'Oh' (line 7). Ann then tells her what she should be looking for 'Wherever there's a change in slope' (line 8). This description of course presupposes Sue's ability to find in the dirt what will count as 'a change in slope'. Sue again moves her tape-measure far to the right. At this point, instead of relying upon talk alone to make explicit the phenomena that she wants Sue to locate, Ann moves into the space that Sue is attending to and points to one place that should be measured while describing more explicitly what constitutes a change in slope: 'See so if it's *fairly flat* I'll need one where it *stops* being fairly flat like right there.'

One of the things that is occurring within this sequence is a progressive expansion of Sue's understanding as the distinctions she must make to carry out the task assigned to her are explicated and elaborated. In this process of socialization through language there is a growth in intersubjectivity as domains of ignorance that prevent the successful accomplishment of collaborative action are revealed and transformed into practical knowledge, a way of seeing, that is sufficient to get the job at hand done, such that Sue is finally able to understand what Ann is asking her to do (that is to see the scene in front of her in a manner that permits her to make an appropriate, competent response to the directive). It would however be wrong to see the unit within which this intersubjectivity is lodged as simply these two minds coming together in the work at hand. Instead the distinction being explicated, the ability to see in the very complex perceptual field provided by the landscape they are attending to, those few events that count as points to be transferred to the map, are central to what it means to see the world as an archaeologist and to use that seeing to build the artefacts, such as this map, which are constitutive of archaeology as a profession. Such seeing would be expected of any competent archaeologist. It is an essential part of what it means to *be* an archaeologist, and it is these professional practices of seeing that Sue is being held accountable to. The relevant unit for the analysis of the intersubjectivity at issue here – the ability of separate individuals to see a common scene in a congruent, work-relevant fashion – is thus not these individuals as isolated entities, but instead archaeology as a profession, a community of competent practitioners, most of whom have never met each other, but who nonetheless expect each other to be able to see and categorize the world in ways that are relevant to the work, scenes, tools and artefacts that constitute their profession.

The phenomena examined so far provide some demonstration of how what is to be seen in a map, scene, human body or image stands in a reflexive relationship to other semiotic structures that participants are using to constitute visual phenomena as a relevant component of the events and activities that make up their lifeworld. These structures include language, the constitution of action and context provided by sequential organization and ways of seeing events and using images of different types that are lodged within the practices of particular social communities, such as the profession of archaeology.

## Professional vision in court

Parties who are not competent members of relevant social communities can lack the ability and/or the social positioning to see and articulate visual events in a consequential way. These issues were made dramatically visible in the trial of four Los Angeles policemen who were recorded on videotape administering a beating to an African-American motorist, Mr Rodney King, whom they had stopped after a high-speed pursuit triggered by a traffic violation. When the tape of the beating was shown on national television there was outrage, and even the head of the Los Angeles police department thought that conviction of the officers was almost automatic. However, at their first trial (they were later tried again in a federal rather than a state court for violating







Mr King's civil rights) all four policemen were acquitted, a verdict that triggered an uprising in the city of Los Angeles, with neighbourhoods being burned, federal troops being called in, and so on. The crucial evidence at the trial was a visual document: the videotape of the beating. Rather than transparently proving the guilt of the policemen who were seen on it beating a man lying prone on the ground, the tape in fact provided the policemen's lawyers with their evidence for convincing the jury that their clients were not guilty of any wrongdoing. They did this by using language, pointing and expert testimony to structure how the jury saw the events on the tape in a way that exonerated the policemen. In essence they used the tape of the beating to demonstrate that Mr King was the aggressor, not the policemen, and that the policemen were following proper police practice for subduing a violent, dangerous suspect (see Goodwin, 1994 for a more detailed analysis of such professional vision). Crucial to their success was their use of another policeman, Sergeant Duke, as an expert witness. It was argued that laypeople could not properly see the events on the tape. Instead, the ability to legitimately see what the body of a suspect was doing, such as Mr King's as he lay on the ground being beaten, and specifically whether the suspect was being aggressive or compliant, was lodged within the work practices of the social group charged with arresting suspects: the police. The ability to see such a body, and code it in terms of its aggressiveness, was a component of the professional practices that policemen use to code the events that are the focus of their work. Insofar as such vision is a public component of the work practices of a particular social group, someone who wasn't present but who is a member of the profession, a policeman, can make authoritative statements about what can be legitimately seen on the tape. However, while policemen constitute a socially organized profession, suspects and victims of beatings don't. Therefore there is no one with the social standing, that is membership and mastery of the practices of a relevant social group, to act as an expert witness to articulate what was happening from Mr King's perspective.

What was to be seen on the tape was structured through the way in which different semiotic fields, such as structure in the stream of speech, pointing which highlighted specific places and phenomena in the image being looked at and events in the image itself mutually elaborated each other to provide a construal of events that served the purposes of the party articulating the image. The following transcript (Figure 8.4) provides an example. At the point where we enter this sequence the prosecutor has noted that Mr King appears to be moving into a position appropriate for handcuffing him, and that while one officer is in fact reaching for his handcuffs the suspect is being co-operative.

By noting the submissive elements in Mr King's posture, and the fact that one of the officers is reaching for his handcuffs, the prosecutor has shown that the tape demonstrates that Mr King is being co-operative. If he can establish this point, hitting Mr King again would be unjustified and the officers should be found guilty of the crimes they are charged with. The contested vision being debated here has very high stakes.

To rebut the vision proposed by the prosecutor, Sgt Duke uses the semantic resources provided by language to code as aggressive extremely subtle body movements of a man lying face down beneath the officers (lines 7–11). Note, for example, not

only line 13's explicit placement of Mr King at the very edge, the beginning, of an aggressive spectrum introduced in earlier testimony, but also how very small movements are made much larger by situating them within a prospective horizon through the repeated use of 'starting to' (lines 6, 8, 11). The events visible on the tape are structured, enhanced and amplified by the language used to describe them.

- 1 Prosecutor: So uh would you,  
 2 again consider this to be:  
 3 a nonaggressive, movement by Mr. King?  
 4 Sgt. Duke: At this time no I wouldn't. (1.1)  
 5 Prosecutor: It is aggressive.  
 6 Sgt. Duke: Yes. It's starting to be. (0.9)  
 7  This foot, is laying flat, (0.8)  
 8  There's starting to be a bend. in uh (0.6)  
 9  this leg (0.4)  
 10  in his butt (0.4)  
 11  The buttocks area has started to rise. (0.7)  
 12 which would put us,  
 13 at the beginning of our spectrum again.
-  *indicates that Sgt. Duke is pointing on the screen at the body part described in his talk.*

**Figure 8.4** Transcript of use of video in the trial of Rodney King.

This focusing of attention organizes the perceptual field provided by the videotape into a salient figure, the aggressive suspect, who is highlighted against an amorphous background containing non-focal participants, the officers doing the beating. Such structuring of the materials provided by the image is accomplished not only through talk, but also through gesture. As Sgt Duke speaks he brings his hand to the screen and points to the parts of Mr King's body that he is arguing display aggression. The pointing gesture and the perceptual field which it is articulating mutually elaborate each other. The touchable events on the television screen provide visible *evidence* for the description constructed through talk. What emerges from Sgt Duke's testimony is not just a *statement*, a static category, but a *demonstration* built through the active interplay between coding scheme and the image to which it is being applied. As talk and image mutually enhance each other a demonstration that is greater than the sum

of its parts emerges, while simultaneously Mr King, rather than the police officers, becomes the focus of attention as the expert's finger articulating the image delineates what is relevant within it.

By virtue of the category system erected by the defence, the minute rise in Mr King's buttocks noted on the tape unleashes a cascade of perceptual inferences that have the effect of exonerating the officers. A rise in Mr King's body becomes interpreted as aggression, which in turn justifies the escalation of force. Like other parties, such as the archaeologists, faced with the task of coding a visual scene, the jury was led to engage in intense, minute cognitive scrutiny as they looked at the tape of the beating to decide the issues at stake in the case. However, once the defence coding scheme is accepted as a relevant framework for looking at the tape, the operative perspective for viewing it is no longer a layperson's reaction to a man lying on the ground being beaten, but instead a micro-analysis of the movements being made by that man's body to see if it is exhibiting aggression.

In the first trial, though the prosecution disputed the analysis of specific body movements as displays of aggression, the relevance of looking at the tape in terms of such a category system was not challenged. A key difference in the second trial, which led to the conviction of two of the officers, was that there the prosecution gave the jury alternative frameworks for interpreting the events on the tape. These included ways of seeing the movements of Mr King's body that Sgt Duke highlighted as normal reactions of a man to a beating rather than as displays of incipient aggression. In the prosecution's argument Mr King 'cocks his leg', not in preparation for a charge, but because his muscles naturally jerk after being hit with a metal club.

The study of the practices used to structure relevant vision in scientific and workplace environments, what Hutchins (1995) has called 'Cognition in the Wild', has become the focus of considerable research. A major initiative for such studies was provided by Lucy Suchman in the early 1990s when she initiated the workplace project while at Xerox PARC. The site chosen for research was ground operations at a mid-sized airport. Documents and images of many different types, and the ability of actors in alternative structural positions to see and analyse events in relevant ways, were crucial to the work of the airport. Phenomena that received extensive study included work-relevant seeing of documents, aeroplanes and events (Goodwin and Goodwin, 1996), the constitution of shared workspaces (Suchman, 1996), the study of how a common document co-ordinated different kinds of work in different work settings, and the practices involved in seeing and shaping phenomena in collaborative work (Brun-Cottan, 1991; Suchman and Trigg, 1993). In part because of the central role played by visual phenomena in the work being analysed, the project's final report was submitted as a videotape (Brun-Cottan et al., 1991). Subsequent analysis growing from this project has focused on the organization of both documents and visual phenomena in a range of occupational settings, such as law firms and the work of architects. In England Christian Heath and his collaborators have investigated the structuring of vision within interaction in a range of settings including the control room for the London Underground, centres for the production of electronic news, art classes, and so on (Heath and Luff, 1992b, 1996; Heath and Nicholls, 1997). In much of this research there is a focus on how core practices for the organization of talk,

reference, gesture and other phenomena central to the production of action within human interaction can encompass not only talk but also embodiment in a world populated by work-relevant objects.

Hindmarsh and Heath (in press) investigate reference within such a framework. LeBaron (1998), Streeck (1996) and LeBaron and Streeck (in press) examine how gesture emerges from the interaction of working hands acting in the world in settings such as architects' meetings and auto body shops. Robinson (1998) has provided analysis of how participants in medical interviews organize their interaction by attending to how gaze is shifted from other participants to relevant visual materials in the setting, such as medical records. Whalen (1995) analysed how the talk of operators responding to emergency 911 calls was organized in part by the task of filling in required information on a computer screen with a specific visual organization. Rogers Hall and Reed Stevens have investigated visual practices in a range of school, scientific and occupational settings (Hall and Stevens, 1995; Stevens and Hall, in press). Research in computer-supported co-operative work has focused specifically on new forms of visual access created by electronic media. Heath and Luff (1992b, 1993) have done considerable research on interaction mediated through video, demonstrating the crucial ways in which resources available to parties who are actually co-present to each other are not available in media spaces. Yamazaki and his colleagues have explored the systematic problems that arise when particular kinds of directives, such as instructions for how to use CPR to start a heart attack victim's heart, are given through talk alone, for example over the phone, without access to a relevant visual environment. Patients usually die since the novice is not able to place his or her hands at the appropriate spot on the patient's body. To remedy some of these issues technologies that incorporate basic resources available for doing reference in face-to-face interaction, such as pointing, have been developed. These include a remote-controlled car with a laser that has the ability to move while clearly marking the specific places being pointed at in a remote environment (Yamazaki et al., 1999). Nishizaka (in press) has investigated how participants co-ordinate gaze both spatially and temporally on electronic documents such as computer screens. Kawatoko (in press) has investigated how lathe workers organize perceptual fields so to make visible the invisible movements of their cutting tools. Both Kawatoko and Ueno (in press) have examined the organization of vision on many different levels (from documents to systematic placement of objects on the warehouse floor as part of its work flow) in the work practices of a large warehouse. In all of this, work practices for seeing relevant phenomena are systematically embedded within processes of social organization, structures of mutual accountability and the organization of activity.

## CONCLUSION

Within both conversation analysis and ethnomethodology visual phenomena have been analysed by investigating how they are made meaningful by being embedded within the practices that participants in a variety of settings use to construct the events and actions that make up their lifeworld. This has led to the detailed study of a range

of quite different kinds of phenomena, from the interplay between gaze, restarts and grammar in the building of utterances within conversation, to the construction and use of visual representations in scientific practice, to how the ability of lawyers to shape what can be seen in the videotape of policemen beating a suspect can contribute to disruption of the body politic that leaves a city in flames, to the part played by visual practices in both traditional and electronic workplaces. Visual phenomena that have received particular attention include:

- 1 The body as a visible locus for displays of intentional orientation through both gaze and posture.
- 2 The body as a locus for a variety of different kinds of gesture, from iconic elaboration on what is being said in the stream of speech, to pointing, to the hand as an agent engaged with the world around it.
- 3 Visual documents of many different types used in both scientific practice and the workplace, for example, maps, graphs, Munsell charts, coding forms, schedules, television screens providing access to distant sites, architectural drawings and computer screens.
- 4 Material structure in the environment where action and interaction are situated.

This perspective brings together within a common analytic framework both the details of how the visible body is used to build talk and action in moment-to-moment interaction and the way in which historically structured visual images and features of a setting participate in that process (see Chapter 4). Rather than standing alone as self-contained, self-explicating images, visual phenomena become meaningful through the way in which they help elaborate, and are elaborated by, a range of other semiotic fields – sequential organization, structure in the stream of speech, encompassing activities, etc. – that are being used by participants to both construct and make visible to each other relevant actions. The focus of analysis is always on how the participants in a setting themselves display a consequential orientation to visual phenomena (for example, by shifting gaze after a restart, focusing their work on a Munsell chart or building images as a core component of the practices used to make visible scientific phenomena). A variety of different methodologies are employed. However a basic component of many research projects includes going to the site where the activities being investigated are actually performed and examining what the participants are doing there as carefully as possible. Videotape records are frequently most useful because of the way in which they preserve limited but crucial aspects of the spatial and environmental features of a setting, the temporal unfolding organization of talk, the visible displays of participants' bodies and changes in relevant phenomena in the setting as relevant courses of action unfold (see Chapter 3). Analysis typically requires not only viewing the tape, ethnographic records and documents collected in a setting, but also the construction of new visual representations such as transcripts of many different types (note how some in this paper incorporate both detailed transcription of the talk and a variety of different kinds of graphic representations). While this analysis sheds much important new light on how visual phenomena are organized through systematic discursive practice, it is not restricted to vision *per se* but is instead investigating the more general practices used to build action within situated human interaction.

## NOTES

- 1 Vision is not, however, essential as both the competence of the blind and telephone conversations demonstrate. Below it will be argued that situated action is accomplished through the juxtaposition of multiple semiotic fields, only some of which make vision relevant.
- 2 Talk is transcribed using the system developed by Gail Jefferson (see Sacks et al., 1974: 731–3). Talk receiving some form of emphasis is marked with *underlining* or *bold italics*. Punctuation is used to transcribe intonation: a period indicates falling pitch, a question mark rising pitch and a comma a falling contour, as would be found for example after a non-terminal item in a list. A colon indicates lengthening of the current sound. A dash marks the sudden cut-off of the current sound (in English it is frequently realized as glottal stop). Comments (e.g. descriptions of relevant non-vocal behaviour) are printed in italics within double parentheses. Numbers within single parentheses mark silences in seconds and tenths of a second. A degree sign (°) indicates that the talk that follows is being spoken with low volume. Left brackets connecting talk by different speakers mark the point where overlap begins.

## REFERENCES

- Brun-Cottan, F. (1991) 'Talk in the work place: occupational relevance', *Research on Language and Social Interaction*, 24: 277–95.
- Brun-Cottan, F., Forbes, F., Goodwin, K., Goodwin, C., Harness, M., Jordan, B., Suchman, L., Randall, L. and Randall, T. (1991) *The Workplace Project: Designing for Diversity and Change*, video produced by Xerox Palo Alto Research Center.
- Chomsky, N. (1965) *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Goodwin, C. (1979) 'The interactive construction of a sentence in natural conversation', in G. Psathas (ed.), *Everyday Language: Studies in Ethnomethodology*. New York: Irvington Publishers.
- Goodwin, C. (1980) 'Restarts, pauses, and the achievement of mutual gaze at turn-beginning', *Sociological Inquiry*, 50(3–4): 272–302.
- Goodwin, C. (1981) *Conversational Organization: Interaction Between Speakers and Hearers*. New York: Academic Press.
- Goodwin, C. (1984) 'Notes on story structure and the organization of participation', in Max Atkinson and John Heritage (eds), *Structures of Social Action*. Cambridge: Cambridge University Press.
- Goodwin, C. (1986) 'Gesture as a resource for the organization of mutual orientation', *Semiotica*, 62 (1/2): 29–49.
- Goodwin, C. (1994) 'Professional vision', *American Anthropologist*, 96 (3): 606–33.
- Goodwin, C. (1995) 'Seeing in depth', *Social Studies of Science*, 25: 237–74.
- Goodwin, C. (1996) 'Practices of color classification', *Ninchi Kagaku* (Cognitive Studies: Bulletin of the Japanese Cognitive Science Society), 3 (2): 62–82.
- Goodwin, C. (2000) 'Action and embodiment within situated human interaction', *Journal of Pragmatics*, 32.
- Goodwin, C. (in preparation) 'Pointing as situated practice', in Sotaro Kita (ed.), *Pointing: Where Language, Culture and Cognition Meet*.
- Goodwin, C. and Goodwin, M.H. (1987) 'Concurrent operations on talk: notes on the interactive organization of assessments', *IPRA Papers in Pragmatics* 1, no. 1:1–52.
- Goodwin, C. and Goodwin, M.H. (1996) 'Seeing as a situated activity: formulating planes', in Y. Engeström and D. Middleton (eds), *Cognition and Communication at Work*. Cambridge: Cambridge University Press.
- Goodwin, M.H. (1980) 'Processes of mutual monitoring implicated in the production of description sequences', *Sociological Inquiry*, 50: 303–17.



- Goodwin, M.H. and Goodwin, C. (1986) 'Gesture and coparticipation in the activity of searching for a word', *Semiotica*, 62 (1/2): 51–75.
- Hall, R. and Stevens, R. (1995) 'Making space: a comparison of mathematical work in school and professional design practices', in S.L. Star (ed.), *The Cultures of Computing*, pp. 118–145.
- Haviland, J.B. (1993) 'Anchoring, iconicity, and orientation in Guugu Yimidhirr pointing gestures', *Journal of Linguistic Anthropology*, 3 (1): 3–45.
- Heath, C. (1986) *Body Movement and Speech in Medical Interaction*. Cambridge: Cambridge University Press.
- Heath, C. and Luff, P. (1992a) 'Media space and communicative asymmetries', *Human Computer Interaction*, 7: 315–46.
- Heath, C. and Luff, P. (1992b) 'Crisis and control: collaborative work in London Underground Control Rooms', *Journal of Computer Supported Cooperative Work*, 1 (1): 24–48.
- Heath, C. and Luff, P. (1993) 'Disembodied conduct: interactional asymmetries in video-mediated communication', in G. Button (ed.), *Technology in Working Order*. London and New York: Routledge.
- Heath, C. and Luff, P. (1996) 'Convergent activities: line control and passenger information on the London Underground', in Y. Engeström and D. Middleton (eds), *Cognition and Communication at Work*. Cambridge: Cambridge University Press.
- Heath, C. and Nicholls, G. (1997) 'Animating texts: selective readings of news stories', in L.B. Resnick, R. Säljö, C. Pontecorvo and B. Burge (eds), *Discourse, Tools and Reasoning: Essays on Situated Cognition*. Berlin, Heidelberg, New York: Springer.
- Hindmarsh, J. and Heath, C. (in press) 'Embodied reference: a study of deixis in workplace interaction', *Journal of Pragmatics*.
- Hutchins, E. (1995) *Cognition in the Wild*. Cambridge, MA: MIT Press.
- Kawatoko, Y. (in press) 'Organizing multiple vision', *Mind, Culture and Activity*, 7 (1–2).
- Kendon, A. (1990a) *Conducting Interaction: Patterns of Behavior in Focused Encounters*. Cambridge: Cambridge University Press.
- Kendon, A. (1990b) 'Spatial organization in social encounters: the f-formation system', in A. Kendon (ed.), *Conducting Interaction: Patterns of Behavior in Focused Encounters*. Cambridge: Cambridge University Press.
- Kendon, A. (1994) 'Introduction to the special issue: gesture and understanding in social interaction', *Research on Language and Social Interaction*, 27 (3): 171–4.
- Kendon, A. (1997) 'Gesture', *Annual Review of Anthropology*, 26: 109–28.
- Latour, B. (1986) 'Visualization and cognition: thinking with eyes and hands', *Knowledge and Society: Studies in the Sociology of Culture Past and Present*, 6: 1–40.
- Latour, B. (1987) *Science in Action: How to Follow Scientists and Engineers through Society*. Cambridge, MA: Harvard University Press.
- Latour, B. and Woolgar, S. (1979) *Laboratory Life: The Social Construction of Scientific Facts*. London: Sage.
- LeBaron, C. (1998) 'Building communication: architectural gestures and the embodiment of ideas', PhD dissertation, University of Texas, Austin.
- LeBaron, C. and Streeck, J. (in press) 'Gestures, knowledge, and the world', in D. McNeill (ed.), *Gestures in Action, Language, and Culture*. Cambridge: Cambridge University Press.
- Lynch, M. (1990) 'The externalized retina: selection and mathematization in the visual documentation of objects in the life sciences', in M. Lynch and S. Woolgar (eds), *Representation in Scientific Practice*. Cambridge, MA: MIT Press.
- Lynch, M. and Woolgar, S. (1990) 'Introduction: sociological contributions to representational practice in science', in M. Lynch and S. Woolgar (eds), *Representation in Scientific Practice*. Cambridge, MA: MIT Press.
- McNeill, D. (1992) *Hand & Mind: What Gestures Reveal about Thought*. Chicago: University of Chicago Press.
- Nishizaka, A. (in press) 'Seeing what one sees: perception, emotion and activity', *Mind, Culture and Activity*, 7 (1–2).

- Ochs, E. (1979) 'Transcription as theory', in E. Ochs and B.B. Schieffelin (eds), *Developmental Pragmatics*. New York: Academic Press.
- Robinson, J.D. (1998) 'Getting down to business: talk, gaze, and body orientation during openings of doctor-patient consultations', *Human Communication Research*, 25 (1): 97–123.
- Sacks, H., Schegloff, E.A. and Jefferson, G. (1974) 'A simplest systematics for the organization of turn-taking for conversation', *Language*, 50: 696–735.
- Schegloff, E.A. (1968) 'Sequencing in conversational openings', *American Anthropologist*, 70: 1075–95.
- Schegloff, E.A. (1998) 'Body torque', *Social Research*, 65 (3): 535–96.
- Stevens, R. and Hall, R. (1998) 'Disciplined perception: learning to see in technoscience', in M. Lampert and M. Blunk (eds), *Mathematical Talk and Classroom Learning: What, Why and How*. pp. 107–49. Cambridge: Cambridge University Press.
- Streeck, J. (1988) 'The significance of gesture: how it is established', *IPRA Papers in Pragmatics*, 2 (1): 60–83.
- Streeck, J. (1993) 'Gesture as communication I: its coordination with gaze and speech', *Communication Monographs*, 60 (4): 275–99.
- Streeck, J. (1994) 'Gesture as communication II: the audience as co-author', *Research on Language and Social Interaction*, 27 (3): 223–38.
- Streeck, J. (1996) 'How to do things with things', *Human Studies*, 19: 365–84.
- Suchman, L. (1992) 'Technologies of accountability: of lizards and airplanes', in G. Button (ed.), *Technology in Working Order: Studies of Work, Interaction and Technology*. London: Routledge.
- Suchman, L. (1996) 'Constituting shared workspaces', in Y. Engeström and D. Middleton (eds), *Cognition and Communication at Work*. Cambridge: Cambridge University Press.
- Suchman, L. and Trigg, R. (1993) 'Artificial intelligence as craftwork', in S. Chaiklin and J. Lave (eds), *Understanding Practice: Perspectives on Activity and Context*. Cambridge: Cambridge University Press.
- Ueno, N. (in press) 'Technologies of mutual accountability of society, social organization, and activity for collaborative activity', *Mind, Culture and Activity*, 7 (1–2).
- Whalen, J. (1995) 'A technology of order production: computer-aided dispatch in public safety communications', in P. Have and G. Psathas (eds), *Situated Order: Studies in the Social Organization of Talk and Embodied Action*. Washington, DC: University Press of America.
- Yamazaki, K., Yamazaki, A., Kuzuoka, H., Oyama, S., Kato, H., Suzuki, H. and Miki, H. (1999) 'GestureLaser and GestureLaser Car: development of an embodied space to support remote instruction', in S. Bødker, M. Kyng and K. Schmidt (eds), *Proceedings of the European Conference on Computer-Supported Cooperative Work, 12–16 September 1999, Copenhagen, Denmark*. Dordrecht/Boston/London: Kluwer Academic.