Another, rather amusing, example was the use of a periscope by Middlemist et al (1976) to help study the effects of the invasion of personal space on stress levels. The setting was a three-urinal men's toilet and the subjects were toilet users whose personal space was 'invaded' by a confederate to the study. Subjects' stress levels were measured by delays in the onset of urination and a shortening of its duration. Urination was observed by a researcher in a nearby toilet stall with the assistance of the periscope! One wonders whether anything is considered private by some researchers!

5 ASSESSING THE VALIDITY OF OBSERVATIONS

As with other research data, we must always be concerned about the validity and reliability of observations. Validity refers to the extent to which observations accurately record the behaviour in which the researcher is interested. One aspect of validity is reliability. This refers to the consistency of observations, usually whether two (or more) observers, or the same observer on separate occasions, studying the same behaviour come(s) away with the same data. Of course, if observational techniques are unreliable they are highly likely to produce invalid data.

5.1 THREATS TO VALIDITY

The validity of observational data can be threatened in a number of ways. First, there is the possibility of reactivity — both personal and procedural. In this case, actual observations of behaviour may be accurate, but subjects do not behave in the way they normally behave. Personal reactivity occurs when subjects behave differently because of the personal characteristics or behaviour of the observer. They may behave in particular ways because the observer is male or female or belongs to a certain 'racial' or ethnic group, or because he or she dresses or conducts him- or herself in a particular way. What is important here is how the subjects perceive the observer, how they interpret his or her behaviour, and how they behave as a result. Where two or more observers are involved, as in many more-structured observations, and subjects react differently with different observers, then problems of reliability emerge too.

Procedural reactivity occurs when subjects behave differently because they know they are being studied or observed. They change their behaviour in response to the procedures involved in the process of observation itself. This form of reactivity is most marked in experiments where subjects are often placed in 'artificial' situations and what goes on is deliberately manipulated. However, it is eliminated altogether in covert research where the subjects are unaware that they are being observed. Where procedural reactivity is high the ecological validity of the observations — the extent to which they can be generalized to other settings — will be in doubt.

A second possible threat to validity comes from the inadequacies of the measuring instruments used in the observation. The preconceived categories of an observation schedule may be unsuitable or inadequate for describing the actual nature of the behaviour which occurs. They may ignore aspects of behaviour which are important given the aims of the research (see my answer to Activity 10, for instance), or may force the observer to code behaviour which is significantly different under the same category. As a result, the descriptions produced by the observation system may be invalid. A number of observers of school classrooms, for example, have been concerned to estimate the extent of unequal treatments by teachers of children from different social groups (see, for example, Green, 1983, French and French, 1984). One research technique has been to count the number.
of interactions teachers have with different pupils. Unfortunately, though, this may give an invalid measure of the extent of unequal treatment because it ignores qualitative differences between individual interactions. These are obviously important to any assessment of inequality.

Techniques used in ethnographic research may also be inadequate. Selectivity in note taking and the distortion which can occur when the researcher relies upon memory when writing field notes can be significant sources of error.

A third potential threat to validity comes from observer bias. All observers have particular cultural knowledge, and approach observation from particular theoretical and sometimes political standpoints. These subjectivities can affect what behaviour is selected for observation, and how this behaviour is interpreted and recorded. They may therefore result in invalid data.

In more-structured observation, precisely what is to be observed is set out in advance and is clearly a product of the researcher's theoretical ideas. But in this type of observation the aim is to minimize the effects of observer subjectivity by ensuring that all observers observe behaviour in the same way and follow the same coding rules. Unfortunately, there is sometimes inconsistency in the way rules are applied by different observers (and sometimes by the same observer on different occasions). This happens particularly when there are ambiguities in the coding system or where coding requires inferences about observed behaviour. As I mentioned in my answer to Activity 1, Scarth and Hammersley (1986) criticize aspects of the ORACLE research (Galton et al., 1980) on these grounds. They argue that the coding rules for categorizing teachers' questions are ambiguous and require the observer to infer teachers' intentions. They suggest that the variations in the questioning styles between the teachers discovered by the research could have been artificially produced by differences between observers. The data produced may therefore have been invalid (see also Croll and Galton, 1986, and Scarth and Hammersley, 1987).

Ethnographers have more often been accused of allowing their theoretical and political preconceptions to bias their observations. Perhaps the potential for this type of bias is greater with ethnography because what is to be observed, and how, is not systematically set out before the observation begins. The work of the anthropologist Margaret Mead (1943), on the island of Samoa, has been criticized on these grounds. Freeman (1984) has suggested that the conclusions she came to about the stress-free nature of adolescence there were to a considerable extent the product of her own preconceptions. His observations revealed rather different experiences and behaviour amongst young people. One of the weaknesses of ethnographic research is that it is largely the product of the ideas, choices and negotiative strategies of an individual researcher. As a result, the picture produced by one researcher of a group or institution may be very different from that produced by another.

A related threat to validity arises from misperception or misinterpretation of behaviour. The observer may simply misunderstand the 'real' nature of the behaviour he or she observes. This is more likely when behaviour is complex or when the observer is unfamiliar with the social situation and the meanings which are in play. For example, in my own research in a multi-ethnic school I observed a situation in which a white adolescent boy called an Afro-Caribbean boy a 'nigger'. I initially interpreted this as a clear instance of racial abuse. However, I discovered later that the two boys were the best of friends and regularly swapped racial insults as a way of reinforcing a sense of camaraderie. My initial observation was clearly invalid (if we define 'racial abuse' as implying negative evaluation of the person to whom it is directed).

5.2 WAYS OF ASSESSING VALIDITY

How, then, do observational researchers assess the validity of their observations? One method is to check the reliability of research techniques. The main way of doing this is replication. Sometimes whole studies are repeated using the same
procedures, with different observers. At other times, parts of a study may be repeated (as in the ORACLE research — see Croll, 1980) Replication is more feasible in experiments and studies involving more-structured observation where procedures are clearly specified and therefore can be fairly easily repeated.

One form of replication involves examining the extent of agreement between two observers of the same behaviour. This technique is more often used in more-structured observation. Here it usually involves comparing the individual coding decisions of two observers to see to what extent they agree. This is termed absolute agreement. Alternatively, it may involve comparing their overall coding results to see whether the total number of behaviours allocated to the categories of the schedule tally. This is called marginal agreement. Obviously, absolute agreement is a much stronger test of the reliability of observations, but it is not always possible to conduct such a test with all observation systems.

Inter-observer agreement is usually worked out for each variable in the observation system and can be expressed as the proportion of instances when the two observers agree on appropriate coding. So if they code 80 out of 100 instances of behaviour in the same way, we have 80 per cent agreement. This was the technique used in the ORACLE research you read about earlier, although, as I pointed out in my answer to Activity 1, the researchers do not give reliability figures for the separate categories.

Replication is not usually feasible in ethnographic research. Procedures are not prestructured, and the course of the research is very much a product of the ideas and personal idiosyncrasies of the researcher and the way he or she interacts with subjects. Moreover, procedures are not usually recorded in sufficient detail to allow another researcher to reproduce them in the same way. The nearest ethnographers come to replication is the re-study. The same group or institution is studied again some time after the original study (see, for example, Burgess, 1987). Re-studies provide limited information on reliability because, although the research setting may be the same, the research questions and procedures, the sub-settings, sub-groups and individuals studied are often very different. Moreover, in the period of time since the original study, the group or institution will probably have changed considerably and differences in data and conclusions may be the product of these changes.

Techniques such as reflexivity, triangulation and respondent validation are more often used in ethnographic research to assess validity. Reflexivity involves the continual monitoring of, and reflection on, the research process. During and after the fieldwork the researcher tries to assess the extent of his or her own role in the process of data production and how the data were affected by the social context in which they were collected. Reflection during data collection may influence the process of future data collection because it may throw up methodological hypotheses and suggest alternative ways of collecting data. When the fieldwork is complete it forms the basis of the researcher's own methodological assessment of the data.

Sometimes researchers provide readers with a reflexive account of their work in the form of a natural history. These are sometimes contained in methodological appendices to the main work (see, for example, Whyte, 1981) or are sometimes published separately (see, for example, Burgess, 1984). These accounts can be useful to the reader in assessing the researcher's role in the production of data and the conclusions of the study.

Triangulation is more a direct check on the validity of observations by cross-checking them with other sources of data. If a researcher's conclusion is supported by data from other sources then we can be more confident of its validity. Triangulation can involve comparing data on the same behaviour from different researchers (as in reliability checks in more-structured observation) who possibly adopt different roles in the field. Alternatively, it can involve comparing data produced by different methods — for example, observational data can be compared with interview data — or it can involve comparing data from different times, sub-settings, or subjects.
Comparing data from the researcher's observations of behaviour with data from the various subjects involved is one form of respondent validation. Here the aim is to check the validity of the researcher's observations by reference to the subjects' perceptions. This may take a number of forms. The researcher may discuss his or her observations with subjects, asking them whether they feel the observations are accurate and what their perceptions of a particular incident were. Alternatively, the researcher may ask participants to supply written accounts of a particular instance or period of behaviour. These can then be compared with the researcher's own observations. Ball (1981), for example, in his study of mixed-ability teaching in a comprehensive school, sometimes compared his account of lessons with those provided by teachers. And, finally, the researcher may feed back his or her observations to subjects and ask for their comments (again a technique used by Ball - see Ball, 1984).

The advantage of such techniques is that the researcher may be able to access important additional knowledge about the behaviour under consideration — for example, about the thoughts and motives of subjects, their perceptions of the behaviour of others, and about the social context in which the behaviour occurred — which is not available from observation. They may also provide a valuable alternative perspective on the behaviour which occurred, as well as useful information on subjects' perspectives. As Fielding (1982) found in his study of the National Front, they may also be a useful way of encouraging the involvement of subjects in the research.

However, respondent validation does not automatically ensure the accuracy of data. Subjects may be more concerned to manipulate the impression of behaviour which is contained in the data as a way of enhancing or protecting their own interests. Their accounts may therefore present behaviour in certain ways or may give particular interpretations of that behaviour. These accounts will, of course, be influenced by the social context in which they are delivered, and by their perception of the researcher and of the use to which the data will be put. Moreover, it is important to recognize that subjects' accounts of their actions and perceptions are reconstructions from memory which may not necessarily correspond to their thoughts or perceptions at the time the behaviour occurred. We must recognize and try to assess, too, the potential threats to validity in respondent accounts.

6 CONCLUSION

In this unit I have tried to introduce you to the different styles and techniques used in observational research. I identified two basic styles — more-structured observation, and less-structured or ethnographic observation. Although there are clear differences between the two styles, I do not want to leave you with the impression that they are opposed and mutually exclusive. I must emphasize that observational research often involves both styles and that, at times, both raise similar issues and problems.

All observers have to gain access to subjects. In a narrow sense this means getting to a physical position from which subjects can be observed. But it also often means developing relationships with subjects so that, as far as possible, the way they behave 'naturally' can be observed. Similar ethical problems — concerning deception, invasion of privacy and harm to subjects, for example — are involved in both styles, and in both styles decisions have to be made about what to observe. In more-structured observation those decisions are more likely to be made prior to the actual fieldwork, whereas in less-structured observation decisions are made during the course of fieldwork itself. Nevertheless, researchers must choose what to focus on and select appropriate samples. The major difference between the two styles is that more-structured observation requires observers to allocate behaviours to preconceived categories of an observation schedule, whereas less-structured observation involves written accounts in field notes describing the nature of behaviour in more detail.
In the final part of the unit I have considered various threats to the validity of observational data, and the main ways in which researchers check validity. Producers of observational research should always be concerned with the validity of their data. Observational research will only make a useful contribution to the bank of public knowledge that I mentioned in my introduction if we can be reasonably confident of its validity. Readers of observational research should also keep matters of validity in the forefront of their minds. A consideration of such matters is essential to any assessment of research. As producers and consumers of observational research, we can never, of course, be absolutely sure of the validity of observations, but we must decide and make clear to what extent confidence in that validity is justified.
ANSWERS TO ACTIVITIES

ACTIVITY 1

1(a) The observers focused on one pupil at a time (the target pupil) and recorded a number of aspects of behaviour — whether the pupil was waiting for the teacher, or paying attention to others, or engaging in disruptive behaviour, the extent of the pupil's involvement in work tasks, and the pupil's location. They also recorded the nature of the pupil's interaction with other pupils, and with the teacher or other adults.

1(b) When observing the teachers, the researchers concentrated on the different types of interactions teachers had with pupils. They distinguished between oral and silent interactions, and, in the former, between questions and statements (of different types).

2(a,b) The researchers selected three local education authorities (LEAs). It is not clear how these were chosen, but in part this reflected a wider concern of the study with pupil transfer to a later stage of education. Each LEA had a different transfer age. In each LEA two receiving schools, contrasting in their internal organization but similar in their catchment areas, were chosen with the help of LEA advisers. Once these schools agreed to take part in the wider study, their feeder schools became the sample of primary schools. One school decided not to take part and in one LEA an unspecified number of other schools were added. The teachers studied were those who taught pupils in the year (or two years) prior to their transfer.

The observers spent three days each term with each class over a one-year period. We are given no information on how the days were selected. They observed for six 55-minute sessions during their stays. In each session the teacher was observed for a continuous period of nineteen minutes — sometimes at the beginning, sometimes at the end, and sometimes in the middle of the session, and eight target pupils were selected by dividing each class into four ability groups — one high ability, one low ability, two of middle ability — and by choosing a boy and a girl at random from each group. Observations were conducted of both the teacher and the target pupils at 25-second intervals.

3 The observers underwent two weeks' 'intensive training' and refresher training during the study. This may help to improve reliability, but it does not, of course, ensure it. In order to check reliability, towards the end of the second year of the study four classrooms were observed by pairs of observers using the teacher and pupil records at synchronized times. The researchers checked the extent of agreement between the observers by comparing their coding of behaviour at each time unit. The observers' coding was the same on the pupil record in about 90 per cent of cases, and on the teacher record in 76 per cent of cases. The researchers maintain that these levels of reliability were acceptable for the study. Part of the ORACLE study was also replicated in the second and third year of the research and the consistency of certain data was checked (a discussion of this can be found in Croll, 1980)

4 When observing pupils the observers recorded the teacher's activity and location, and also the behaviour of any adult or pupil with whom the target pupil interacted. When observing the teachers they noted the type and composition of their pupil audience and also the curricular activity of pupil(s) with whom the teacher interacted. The observers also recorded more general information about the classes such as seating arrangements, classroom management, organization and grouping, curricular activities, and the use of materials, books and apparatus. They also seem to have noted basic information about the schools and their catchment areas.

5 I have identified the following advantages in the research methods used: The very fact that the ORACLE research used observation was an advance on other
studies of classrooms such as the one by Bennett (1976) which relied in the main on teacher-completed questionnaires for data on classroom organization and behaviour. The research also involved a relatively large amount of classroom observation compared to other observation studies of classrooms. More-structured observation allowed the researchers to provide a more rigorous, quantitative description of variations in teacher and pupil behaviour than is possible using a less-structured approach. It was also possible for the observations of several different researchers to be compared and combined to provide a broader picture. The researchers could assess the reliability of the observational techniques employed, and replicate aspects of the study. (In the second year of the research some of the children were observed again with different teachers and the same observation instruments were used.)

However, I can also see a number of problems. The sample of LEAs, schools and teachers was not selected in a way that made it likely to be representative of any wider population of LEAs, schools or teachers. Moreover, the size of the time samples of classroom behaviour (three days each term) was very small in comparison with the total amount of class time in a year and the samples do not appear to have been random. Their representativeness must also therefore be in doubt. It is difficult to know whether we can generalize on the basis of these samples about behaviour in these particular classrooms, or about classroom behaviour in a wider population of classrooms.

One thing that is not clear (at least from this extract) is why the ORACLE researchers were interested in the particular categories of behaviour that appear in their observation schedules. Presumably these were taken to be indicators of particular teaching styles or types of classroom behaviour. However, it is difficult to judge the validity of the categories as measures without some discussion of this.

Some of the categories in the observation schedules give me cause for concern. I think it must have been difficult to code behaviour unambiguously into certain categories. Although the researchers produced and utilized a coding manual (Boddell and Jasman, 1983), which provided detailed rules for coding behaviour, Scarth and Hammersley (1986) argue that this did not eliminate all potential problems. They consider the categories of 'teacher questions' and point out that coding sometimes required assessments by the observer of the teachers' expectations of pupils' answers and therefore inferences about teachers' thinking. This raises doubts in their minds about the validity of the observational findings.

Scarth and Hammersley (1986) also note that the reliability scores are aggregates for all the categories on the teacher and pupil records. They therefore may conceal large variations. It is possible, for example, that reliability was 100 per cent for some categories of behaviour, but much lower for others. Thus on certain crucial variables reliability may have been poor. Furthermore, the reliability scores were calculated on observations conducted in the second year of the study when the observers presumably had become fairly familiar with the observation schedules. Reliability might have been more of a problem in the earlier stages of the study.

Another issue raised by Scarth and Hammersley (1986) is that of reactivity. Given that observer visits were relatively short and infrequent, and that teachers (and maybe pupils) knew they were going to be observed by a relative stranger, it is likely that reactivity would have been high.

A final point worth bearing in mind, should you examine the results of the study, is that time sampling — at 25-second intervals — conflates frequency and duration. So, for example, we cannot say from the data produced whether the teacher asked a large number of short questions or a small number of long ones.

**ACTIVITY 2**

1. Punch initially tried to study police forces in Britain, but his proposals were rejected by the Home Office (he does not explain why). He decided to study the
Dutch police because he felt there was a better chance of his proposal being accepted, as individual forces in Holland can make their own decisions about research. He had also lived for a short period in Holland and had a Dutch wife. Punch is not very clear why he wanted to study an inner-city force in Amsterdam. He sees to have been attracted by the potential excitement and controversy which sometimes surrounds police work in inner-city areas, and he implies that Amsterdam was more typical of other large ‘cosmopolitan’ cities than were other Dutch cities. His contact and sponsor, Dr Rom Fris, seems to have chosen the specific station and the police officers with whom Punch spent his first period of fieldwork. It is difficult to judge to what extent the officers Punch studied were representative of any wider population.

2 Punch’s informal contacts with the International Police Association enabled him to spend two weeks with two ‘community relations officers’ (p 183) in Rotterdam. Here he got to know Dr Fris, a psychologist working for the Amsterdam police. Fris seems to have acted as a sponsor and arranged Punch’s access to a specific station and officers. Punch tells us little about how he developed these contacts and how Fris acted on his behalf. He does say that being a foreigner was an advantage in gaining access because he was perceived as less threatening. He also explains that after his first spell in the field (one month) he sent copies of his report to ‘strategically placed people’ (p 189) and was given permission to do a further three months’ fieldwork.

3 Punch tells us more about how he built up relationships with the officers he studied. The fact that he could speak Dutch (or was learning) obviously helped. He concentrated on one ‘shift’ of officers and demonstrated that he was willing to be with them during their work at all times (including the night shift) and in all weathers. Perhaps more importantly, he helped the officers with their work in a number of ways. He assisted on routine tasks such as making coffee and clearing up in the canteen or sweeping the road after an accident, but also on more important work such as translating during the interrogation of English suspects, and even the discovery and arrest of drug traffickers. The officers themselves sometimes found it useful for Punch to pose as a detective. In these ways, Punch demonstrated that he was willing to get involved in police work, that he had an empathetic approach to the officers, and that he was someone who could be trusted. He also mentions that he deliberately presented himself as anti-intellectual.

4 Punch seemed to take part in police work increasingly over the course of the fieldwork. His role changed from a relatively uninvolved observer/researcher to a more actively involved officer-assistant/observer/researcher.

5 Punch’s research was not covert as regards the police officers. They knew he was doing research, although how much they knew, or were told, about the aims and focus of the research is difficult to tell. But the ordinary citizens, with whom Punch observed officers dealing, did not know he was a researcher. Indeed, it is clear that many thought he was a police officer. These subjects, if we accept they were subjects, were deceived, or at least not informed, about the research. However, it is difficult to see how in many situations Punch could have consulted them about the research.

Punch’s active role as an assistant officer clearly had an adverse effect on some of these ‘subjects’. In the case of the drug traffickers, Punch’s actions led to their arrest. Some of these subjects were clearly harmed (in the sense that they were more likely to be arrested and convicted) by the research. Punch had to weigh up a number of competing commitments here — his commitment as a citizen to particular values and laws, a researcher’s commitment to protect his or her subjects’ interests, and a commitment to the importance of his research which benefited from an enhanced relationship with his police officer subjects. When considering his criminal subjects, Punch seemed to side with the first and third of these commitments. Indeed, he makes his position vis-à-vis the interests of many of the criminals with whom he came into contact pretty clear on pages 196–7. He is more ambivalent about what his position would have been had he discovered police malpractice or crime. ‘Fortunately’, he says, ‘this problem did not arise’.
in his research, although earlier in the extract (p 192) he describes being given considerable information, at a party, about police corruption

6 Punch argues that participant observation enabled him to ‘penetrate’ (pp 184–5) the occupational culture of police officers. He claims that the police often erect barriers which prevent public scrutiny and utilize strategies to present a favourable image of their work. Participant observation enabled him to see ‘the inner reality of police work’ (p 184), to get beneath public representations and see policing, both on the street and behind the scenes, as it really is. He argues that much police work consists of unique face-to-face encounters involving considerable discretion on the part of individual officers, and that observation is required to get at this complexity. He implies that ‘more-structured techniques’ (p 185), by which I think he means interviews and questionnaires, would not have given accurate information on what officers actually do.

7 During the fieldwork he obviously made use of informal conversations and discussions with officers. He also had access to documentary and statistical evidence, although he does not say much about what types. Following the fieldwork, he conducted interviews with an unspecified number of officers of different ranks.

8 I think the data he collected give us a more detailed and accurate account of police work (or the work of these particular officers) than we would obtain from other methods. To some extent Punch did gain an insider’s view of the occupational culture of this group of officers. But we must remember that Punch’s account is a constructed representation of police work. It is a representation based upon lengthy observation and close study and is therefore more likely to be valid than others which are not based on such extensive research. But nevertheless it is a representation coloured by Punch’s ideas and viewpoints, and therefore does not exhaust the ‘reality of police work’.

Punch’s account is also, of course, a product of what he was able to observe and of what the officers were prepared to tell him. Aspects of the ‘reality of police work’ may have been concealed from him. Indeed, in the section of the extract where he discusses police corruption, Punch concedes that ‘a subterranean police culture had largely escaped me’ (p 192). The account is also a product of what Punch selected to observe and record. My impression is that Punch focused on particular events which were exciting and exotic rather than routine.

**ACTIVITY 10**

Just after I wrote this activity I had a conversation with a female colleague who claimed that departmental meetings (I work in the education department in a ‘new’ university) were almost always dominated by men. I decided to observe the next departmental meeting to put her claim to the test. Of course, the meeting I observed may not have been typical and, as her claim was of a probabilistic nature, observing one meeting was not a real test. Nevertheless it provided the basis for a useful exercise. I have tried here to reproduce briefly my thoughts as I prepared this exercise.

I had to think first about behavioural indicators of ‘dominance’ One possibility I thought of was a simple count of the number of contributions made to the meeting by men and women. In this case I would simply have two categories of behaviour — contribution or no-contribution to the meeting — and all I would have to do would be to record whether a contribution was being made and if so whether it was made by a man or a woman.

There were two problems — I had to decide what sorts of behaviour made up ‘the meeting’, and what behaviour constituted ‘a contribution’. The former was relatively easy to resolve. I decided that the meeting began when the chairperson (a man) called it to order and raised the first item on the agenda, and that it ended when he formally closed it. Therefore, **excluded** informal talk before and after the meeting, which actually in part concerned the business of the meeting, and included certain informal/social talk, not relevant to the agenda, which occurred.
during the meeting. This did raise doubts in my mind about my definition of the meeting. The second problem was more difficult. How should I define 'contribution' and what rules would I have for deciding whether a particular instance could be coded as a contribution? I decided to exclude non-verbal behaviour, and single word utterances — such as 'yes', 'no', 'what', 'how', etc — and formulated the following definition: 'a contribution consists of a verbal utterance of a phrase or more, addressed to the general body of the meeting'. I decided that a phrase consisted of two or more words, but pondered over how I would decide whether or not utterances were 'addressed to the general body of the meeting'. I wanted to exclude comments which were made as asides to immediate neighbours or to sub-groups and concentrate on the formal talk which made up the meeting. (Was my definition of 'the meeting' changing?) I decided to see how things went during my observation.

Something else that occurred to me was that just counting contributions took no account of the length of those contributions. It was clearly possible for a person to make only a small number of contributions and yet take up a large proportion of the meeting time. I decided therefore to use a system of continuous recording which would enable me to time the length of individual contributions. I would have to indicate on a time schedule when a contribution started and finished and whether it was made by a man or a woman. I would then have data on the frequency and the duration of contributions made by men and women.

My preparations thus made, I set out to observe the meeting. I decided to focus on the first 30 minutes of the meeting, partly because I thought this would be long enough to try out the technique, and partly because I knew there were subjects in the second half of the agenda on which I wanted to contribute. This obviously further restricted the representativeness of my observations, but it also highlights the difficulty — in fact I think the impossibility — of conducting live more-structured observation in situations in which one is actively involved. I must confess that I did not seek the consent of the members of the meeting to my observation. I placed myself on the periphery of the meeting and observed covertly. I felt this was acceptable because this was just an exercise rather than a 'real' piece of research, and because seeking informed consent would have involved time-consuming negotiation which I could not afford. I also felt that no-one would be harmed, and that people might have behaved differently had they known I was observing them. You may disagree with my ethical judgements here.

The schedule, taken from the middle of the observation period, and shown in Figure 2, illustrates the type of data I collected.

You may be interested to know that men made 43 per cent and women 57 per cent of the contributions to the meeting. However, men's contributions tended to be longer — they took up 62 per cent of the total contribution time, compared to 38 per cent for women's. In considering the questions of dominance and equity, we would perhaps have to adjust these figures to take account of the different number of men and women present in the meeting. On this occasion there were fourteen men and eleven women present. (Thus we would expect men's total contribution to be a little higher, whether or not they were dominant.)

I had a number of problems collecting these data and I also have reservations about the data I produced. For the sake of brevity let me list these:

1. I found it difficult to record precisely the beginning and end of a contribution. This was mainly because I was using the second hand of my watch. I really needed a stop-watch.

2. I found concentrating on the recording quite hard work (especially as I was trying to listen to the subject matter of the meeting) and when the half hour was up I was quite relieved.

3. On several occasions there were overlaps when two (or more) people were talking at once which made accurate recording difficult.
There were also a number of ambiguous 'contributions', as I expected — for example, comments made which were not directed to 'the general body of the meeting'. I am not sure that I recorded these consistently. All these problems raise questions about the reliability of the method I used.

Perhaps more serious is the question of whether number and length of contributions are valid indicators of dominance. If men (or women) speak more often and for longer in meetings (given their numbers in the meeting), does this mean they have 'dominated' the meeting? If contributions are proportionate does this constitute parity? On reflection, I think these are fairly weak indicators of dominance. Dominance surely also involves what people say. It is possible for a person to say very little and yet control the conduct and content of a meeting by the nature of what they say or the way in which they say it. Dominance may also involve having a key influence over certain phases of a meeting — for example, when decisions are made. Again, a person may say very little during a meeting but when it comes to the key function of a meeting — to make a decision on some matter — theirs is the dominant voice. Dominance in this sense may be better indicated by who gets their way in terms of decisions. Moreover, dominance can also be a matter of setting and controlling the agenda of a meeting. Key decisions may be taken before the meeting even begins.

So perhaps I needed data on other indicators of dominance. For instance, I could have divided the nature of people's contributions into different categories and then coded each category — such categories might have included, for example, questions, information giving, expression of opinion, procedural comments, etc. However, I suspect you can imagine some of the problems this would raise. I perhaps also needed data on who made important decisions, and who set the agenda, amongst other things. (See Hargreaves, 1981, for a discussion of dominance in meetings based on a more ethnographic approach.)
ACTIVITY II

In my first attempt to do this activity I decided on a Chinese take-away shop as the location. I chose this setting because on previous visits I had been fascinated by some of the patterns of interaction occurring there. Thinking about possible lines of analysis before I went, I came up with three ideas: the meeting of different ethnic cultures (i.e., Chinese and English), the dramatic analogy between the front and the back of the shop, as frontstage and backstage (see below), and some idea of leisure cultures — patterns of association, interest and values associated with leisure activities.

When I arrived at the shop there was only one person there: a man sitting down reading The Sun who collected his food and left almost immediately after I had arrived. No-one else came into the shop while I was there. The man serving in the shop spent all the time in the back part of the shop until he brought out the food I had been waiting for. There was a Liberal Party Political Broadcast on the colour television.

My response to this situation was that nothing was happening. The only possibility was to use the stage–backstage idea, but I could not develop it any further than it had already been developed by Erving Goffman (1959) from whom I had borrowed it in the first place. He had observed that service establishments are frequently divided into two areas: a front area where those working in the establishment present a front to clients, and a back area to which clients are not admitted.

Given the failure of my first shot at the activity to produce anything, I decided to have another go, using the same setting. As you will see, not much more seemed to be happening on my second visit, but I made a little more of it.

I shall present my jotted notes first, then my written-up version of the field notes and then the outline of a possible analysis.

**Jotted notes**

- Middle-aged couple, well-dressed, woman in fur coat, man in a coat with fur collar.
- Man standing reading paper
- Woman (of the couple), reading extracts from paper. American accent. 'That's ours.' Collects Chinese food.
- TV faces just one side of room. Man who serves is in back of shop.
- Man standing glances at violence on TV, puts paper down and watches (filling in time).
- He is standing, arms folded.
- (When I first arrived there was a couple just collecting Chinese and leaving.
- Man: 'It's ready, come on, let's get home that film's on')
- (Goffman)
- (Activities which occur at standard times/those slotted in)

**Written-up field notes (written up on return from the take-away)**

I enter and sit down on the left hand side of the room (R).

A man and woman (A and B) sitting on the left of the room, in their twenties, casually dressed, jeans etc. The server comes through the curtain from the kitchen, holds a packet of food towards the young couple.

A to B: 'It's ready, come on, let's get home that film's on.'

The server goes back into the kitchen.

Middle-aged couple (D and E), well-dressed, woman wearing fur coat, man in a car coat with fur collar. The woman is reading out excerpts from the newspaper to the man in a 'have you seen this?' kind of way.

The server reappears.

D. 'That's ours?'
They leave

Man (C) standing in the middle of the room reading a newspaper. TV programme relates to the everyday lives of two American policemen and at this point a particularly violent scene seems to occur (by the sound of it, I can't see the picture) C glances at the TV screen a couple of times, then puts down the paper and stands watching it, arms folded.

Outline of a possible analysis

A fairly obvious idea which struck me during the course of the observation was that what these people were doing was filling in time (whereas on this occasion I was not). In other words, they had not come to this place specifically to perform these activities — chatting, reading, watching TV etc. Furthermore, these activities would be curtailed as soon as the food arrived — they are in that sense low status or low priority activities (I can imagine that this is not always the case. For example, if by chance one meets a friend in this setting, the conversation may be continued after the food arrives, though there is a cost to this — cold food!)

I then began to wonder if we could categorize activities in terms of those which are time- and place-bound and those which are 'portable' and can be fitted in anywhere, and between those which are high and low priority/status, a distinction which may or may not coincide with the first. However, even from the little bit of data collected here we can see that this is not the case. While watching TV can be a 'fill in', it can also be a high priority activity which people go to particular places at particular times to do — witness A's comment to B as he collects the food. Nevertheless, while the same activity can vary in its priority and portability, there do seem to be certain kinds of activity — smoking and chatting, for example — which are standard time fillers.

However, I wonder whether the notion of filling or killing time is really satisfactory. These people are in fact doing something, they are waiting. The point is better formulated in terms of the level of involvement associated with an activity on a particular occasion and thus the degree to which other activities can be performed simultaneously. Waiting requires a minimal level of involvement, presumably one must simply watch out for the food to arrive, and thus it can be combined or 'filled' with other activities.

This is as far as my analysis got in the time available.
REFERENCES


Becker, H S, Geer, B, Hughes, E C and Strauss, A L (1961) *Boys in White Student Culture in Medical School*, Chicago, IL, University of Chicago Press

Bell, PA (1990) *Environmental Psychology*, 3rd edn, Fort Worth, TX, Holt, Rinehart and Winston


Boyce, D and Jasman, A (1983) *The Pupil and Teacher Record: A Manual for Observers*, Leicester, University of Leicester


UNIT 12 OBSERVATIONAL RESEARCH


Glaser, B and Strauss, A (1968) Time for Dying, Chicago, IL, Aldine


83


Humphreys, L (1970) *Teenroom Trade*, Chicago, IL, Aldine


King, R. (1978) *All Things Bright and Beautiful?: A Sociological Study of Infant Classrooms*, Chichester, Wiley


Mead, M (1943) *Coming of Age in Samoa*, Harmondsworth, Penguin

UNIT 12 OBSERVATIONAL RESEARCH


Patrick, J (1973) A Glasgow Gang Observed, London, Eyre Methuen

Pellegrini, A D (1989) 'Elementary school children's rough and tumble play', Early Childhood Research Quarterly; vol 4, pp 245-60


Punch, M (1979) 'Observation and the police the research experience', in Hammersley, M (ed) (1993) (DEH313 Reader)


Rubenstein, J (1973) City Police, New York, Ballantine


Spencer, G (1973) 'Methodological issues in the study of bureaucratic elites: a case study of West Point', Social Problems; vol 21, no 1, pp 90-103


ACKNOWLEDGEMENTS

Grateful acknowledgement is made to the following sources for permission to reproduce material in this unit:

**TEXT**


**FIGURE**

Figure 1 adapted from Junker, B.H. (1960) *Field Work*, University of Chicago Press © 1960 by the University of Chicago All rights reserved

**TABLE**

Table 1 Flanders, N (1970) *Analysing Teaching Behavior*, Addison-Wesley, © Ned A Flanders.