To judge rightly we need to bear both in mind, never to forget the numbers when thinking of the percentages, nor the percentages when thinking of the numbers. This last is difficult to those whose daily experience or whose imagination brings vividly before them the trials and sorrows of individual lives. . . . In intensity of feeling such as this, and not in statistics, lies the power to move the world. But by statistics must this power be guided if it would move the world aright.

Booth, *Poverty*, 1902–3

A man [*sic*] is but what he knoweth. The mind itself is but an accident to knowledge, for knowledge is a double of that which is. The truth of being, and the truth of knowing, is all one.

Bacon, *Essays*, 1625

It may be the World will judge it a fault in me, that I oppose so many eminent and ingenious Writers, but I do it not out of a contradictory or wrangling nature, but out of an endeavour to find out truth, according to that proportion of sense and reason Nature has bestowed upon me.

Cavendish, *Observations upon Experimental Philosophy*, 1666

Experiments in Knowing

Gender and Method in the Social Sciences

Ann Oakley

The New Press

New York

Originally published in the United Kingdom by Polity Press, 2000

Published in the United States by The New Press, New York, 2000
Distributed by W. W. Norton & Company, Inc., New York

LIBRARY OF CONGRESS CATALOGING IN PUBLICATION DATA
Oakley, Ann.
Experiments in knowing : gender and method in the social sciences / Ann Oakley.
p. cm.
Includes bibliographical references and index.
ISBN 1-56584-620-6 (hc.)
1. Feminist theory. 2. Social sciences—Methodology. I. Title.
HQ1190.025 2000
305.42′01—dc21 00-036163

The New Press was established in 1990 as a not-for-profit alternative to the large, commercial publishing houses currently dominating the book publishing industry. The New Press operates in the public interest rather than for private gain, and is committed to publishing, in innovative ways, works of educational, cultural, and community value that are often deemed insufficiently profitable.

The New Press, 450 West 41st Street, 6th floor, New York, NY 10036
www.thenewpress.com

Set in Ehrhardt
Printed in the United States of America

9 8 7 6 5 4 3 2 1

Contents

3

Hearing the Grass Grow

If we had a keen vision and feeling of all ordinary human life, it would be like hearing the grass grow and the squirrel's heart beat, and we should die of that roar which lies on the other side of silence.

Eliot, *Middlemarch*

To the professional positivist this seems like chaos. The voices and material lead the researchers in unpredictable, uncontrollable directions. This is indeed not a controlled experiment.

Okely, 'Thinking through fieldwork'

This chapter examines in more detail some central issues about the validity and reliability of 'qualitative' research. It picks up on some of the contentions of the case against 'quantitative' methods which were discussed in the previous chapter: is it true that a research method based on listening to the silent necessarily builds a more valid knowledge? How can we distinguish 'good' from 'bad' 'qualitative' research? What are the strengths and weaknesses of using limited, non-representative samples? Do multiple methods of data collection invariably help the search for valid research findings? Do 'qualitative' methods really dissolve power and preserve the privacy and integrity of research participants – are they really more 'ethical'?

The eye of the beholder

Antoine de Saint-Exupéry's classic work *The Little Prince* opens with an account of how as a child the author drew a picture of a boa constrictor digesting an elephant (figure 3.1). When he showed this to the adults in his social circle, they could only see a hat. But the young artist retained his flair for imagination as an adult, so that when he became a pilot and his plane broke down in the middle of the Sahara desert, and a little prince from asteroid B-612 appeared

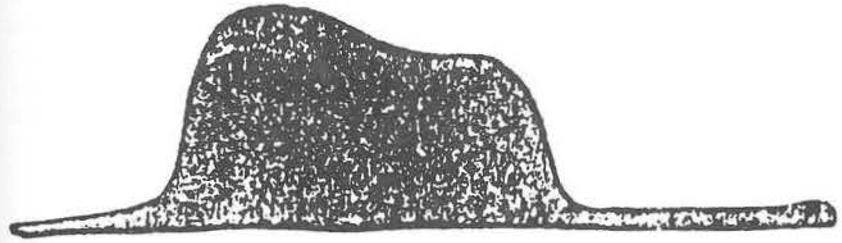

Figure 3.1 A boa constrictor digesting an elephant, or a hat.
From de Saint-Exupéry 1974:7, © Editions Gallimard.

and asked him to draw a sheep, he was not at all surprised. The little prince told him that asteroid B-612 had only once been seen through a telescope, by a Turkish astronomer in 1909. The astronomer had presented his finding to the International Astronomical Congress, but no one had believed him, because he was wearing Turkish costume at the time. When he repeated the presentation eleven years later minus the Turkish costume, it was well received.

Aside from being a classic of children's literature, the narrative of *The Little Prince* conveys two points about research rather well: first, that what is seen is shaped by the eye of the beholder; and second, that the credibility of research findings depends in part on the social standing of researchers. Figures 3.2 and 3.3 extend the point about the variability of perception: whose vision is the true one? The animal shown in figure 3.2 entered the epistemological literature via the reflections of Wittgenstein in his *Philosophical Investigations* (1997:194–5). The figure of 'duck-rabbit'[1] was used by Wittgenstein to convey the message that seeing and interpreting are different activities; that what people see is affected by what they already know about the underlying pattern. The figure appears quite unequivocally to be a duck until one's attention is drawn to the small indentation on the right which marks the rabbit's mouth. Now the duck's beak becomes the rabbit's ears. The third example, figure 3.3, is the well-known Müller–Lyer diagram. Most residents of a '"carpentered" culture' (Campbell 1988b:362) will say that the horizontal line in (a) is longer than the one in (b). But use of a ruler will reverse this perception, and (b) turns out to be longer than (a).

It could, of course, be argued that figures 3.1–3.3 *are* inherently ambiguous. It is reasonable to see a hat and a duck rather than a boa constrictor and a rabbit; the trick of vision that leads to line (a) seeming longer than line (b) would deceive most people. Other visual representations, such as photographs, offer a greater degree of verisimilitude by comparison. But this is not necessarily the case. In the early photographs of American Indians, for example, photographers clothed and posed their subjects according to how they thought Indians *ought* to

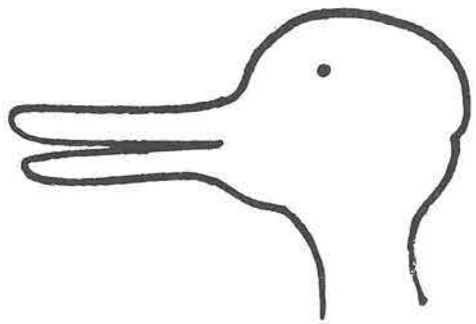

Figure 3.2 'Duck-rabbit'. From Wittgenstein 1997:194.
Reproduced by permission of Blackwell Publishers.

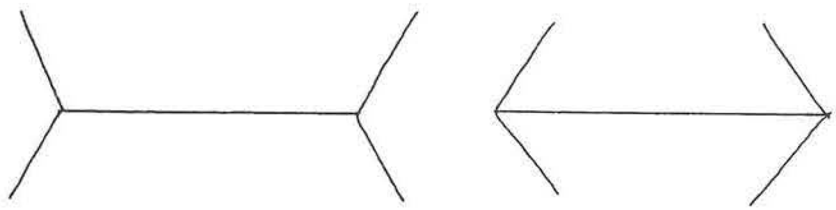

Figure 3.3 The Müller–Lyer diagram.
From D. T. Campbell 1988b:362. © 1988 by the University of Chicago. Reproduced by permission of the University of Chicago Press.

look (Scherer 1975; see also Becker 1979). Some degree of 'posing' is a feature of much modern photography. If we 'triangulated' (see below) the frequency of smiling suggested by photographs, on the one hand, with the conclusions of observational data, on the other, the two sets of results would probably be considerably out of line with one another: people smile because this is expected when photographs are taken. Methods of recording reality produce representations which are mediated by the recording activity itself. We are left with the question of how far these representations correspond to the 'truth' from which they were derived. While this is a question to be asked about all research, it is particularly pertinent to 'qualitative' enquiry, which does not aim to be 'representative' in the same way as 'quantitative' research.

As the aphorism goes, the findings of 'quantitative' research may be reliable but not valid, while the opposite may be the case for 'qualitative' studies: validity is all, and reliability very much a secondary issue. 'Validity' here means the extent to which the research 'findings' truly correspond to the 'reality' from which they were drawn; 'reliability' refers to the repeatability of the findings – will the same researcher obtain the same data again, and/or will different researchers treat the data in the same way? (Galtung 1967:121).

Hearing the silent

Giving voice to the silent has been a dominant feminist metaphor. Tillie Olsen celebrated the history of women writers in a book called *Silences* (1978). Feminist sociologists such as Dorothy Smith (1988) refer extensively to the metaphor of 'silencing'; one of the earliest collections of feminist social science essays in the USA was called *Another Voice* (Millman and Kanter 1975a). The quotation from George Eliot's *Middlemarch* at the head of this chapter (1872:189) makes the important link between the silencing of particular voices, on the one hand, and the material contexts in which silencing is embedded, on the other. Those of us whose experiences and labours are most closely bound up with responsibility for the everyday world are least likely to be considered as having something important to say. What George Eliot is commenting on in the quotation from *Middlemarch* is the state of the young, passionate Dorothea Brooke's feelings after six weeks of marriage to the dry, over-intellectual clergyman Mr Casaubon. The couple are honeymooning in Rome, whose delights provoke in Dorothea a wealth of fresh observations; these contrast with the 'lifeless embalmment of knowledge' she finds in her husband. Dorothea's feelings seem to her unauthenticated; the marriage acts as a metaphor for the clash of 'qualitative' and 'quantitative' approaches.

The methods of 'qualitative' research – interviews, observations, focus groups, life histories – are notable for the closeness they require between researcher and researched. The two sides of the research process exist in the same plane, face to face. In-depth interviews are the face-to-face method *par excellence*, and so have been the chosen method for feminist researchers.[2] Interviews imitate conversations; they hold out the promise of mutual listening. Many of the reasons for preferring a 'qualitative' approach centred on in-depth interviews are the obverse of the objections which feminist critiques have levelled against 'quantitative' methods: the advantages of 'connected' as distinct from 'separated' knowing, dissolution of the artificial boundaries between knower and known, the opportunity to ground knowledge in concrete social contexts and experiences.

Gender and generalization

Women's Ways of Knowing by Mary Belenky and colleagues, first published in 1986, rapidly rose to fame as an exemplary feminist research project. It is a study which meets many of the criteria for a feminist 'qualitative' project:

loosely structured interviewing, sensitivity to interviewees' concepts and forms of thought; a concern with grounding women's experiences in the material circumstances of their lives. The analysis provided is appealing. It resonates on the ideological level with feminist precepts, and thus also with those personal aspects of women's experiences which those precepts represent. But it does present us with a number of problems, many of which concern the general credibility of 'qualitative' research studies.

The book is, as the title suggests, a study of how women know and the ways of knowing that are valued by women. It was hailed as making a significant contribution to the growing literature on women's distinctive, but traditionally silenced, ways of being, knowing, thinking and feeling, a literature initiated by Carol Gilligan's study of moral choice, *In a Different Voice* (1982). Whereas Gilligan's argument was that mainstream theories of psychological development ignore the particular parameters and processes relevant to women's lives, Belenky and colleagues focused on the ways in which education as an institution has traditionally disregarded the special orientation of women to education. In their study they describe five epistemological perspectives from which they say the women they interviewed viewed and knew the world: silence, received knowledge, subjective knowledge, procedural knowledge, and constructed knowledge.

The interviews in the study were largely unstructured, allowing the women to talk about and around a range of topics broadly defined as of interest to the research team. Questions such as: 'What does being a woman mean to you?', 'What do you care about, think about?' and 'How do you know what is right/true?' were asked (Belenky et al. 1986:231, 234), along with others derived from previous work on moral and cognitive development. The interviews ranged from two to five hours long, and resulted in 5,000 pages of transcribed text. Parts of the transcripts were scored by coders 'blind' to who the women were;[3] the rest of the material was analysed contextually, using a laborious process involving multiple readings, copying of quotations and categorizing of themes: 'The very process of recopying the women's words, reading them with our eyes, typing them with our fingers, remembering the sounds of the voices when the words were first spoken,' say Belenky and colleagues, 'helped us to hear meanings in the words that had previously gone unattended' (Belenky et al. 1986:17). This process of 'immersing' oneself in the data is generally considered to be an important feature of 'qualitative' research.

There were 135 women in Belenky et al.'s sample, which was not random but intentionally chosen to represent certain groups of women. Ninety of the 135 were enrolled in six different kinds of educational institutions, and were selected after discussion with staff for being 'representative' in terms of age, interests, commitment and educational/academic performance; twenty-five of the ninety had previously been interviewed by a member of the research team for another project. The remaining forty-five women came from three different family agencies: an organization working with teenage mothers; a network of self-help groups for parents with a history of child abuse and family violence; and a children's health programme. How the interviewed women were

selected from the three agencies is not clear, and no information is given about the social backgrounds of the sampled women as a whole.

Most significantly, Belenky and colleagues chose to listen only to women. They approached their subjects with a disposition to search for certain themes: in this case, how the institutions of the family and education impose on women a particular kind of silencing, within which, like the youthful Dorothea Casaubon, women can find it difficult to hear the authenticity of their own voices. In defence of their research strategy, they offer the following: 'The male experience has been so powerfully articulated that we believed we would hear the patterns in women's voices more clearly if we held at bay the powerful templates men have etched in the literature and in our minds' (Belenky et al. 1986:9). They did have some comparative material available, however: a study carried out some years earlier by William Perry and colleagues on the epistemological development of predominantly male students at Harvard (Perry 1970). But despite the references to the Perry data, the analysis seems to have proceeded without much cross-reference to the male students' ways of knowing; these data were in any case gathered some 20 years earlier and from a single-institution sample. Belenky and colleagues do note that the women in their study were 'on the whole less privileged in terms of social class' (Belenky et al. 1986:44) than the men in Perry's study. The results of their study showed that 'almost half' of the sample fell into the 'subjective' knowledge category; there were 'only two or three' in the 'silence' category (Belenky et al. 1986:55, 23); and the distribution of the sample across the other types of knowing is not given. 'We recognize', say Belenky and colleagues, 'that these five ways of knowing are not necessarily fixed, exhaustive or universal categories . . . that similar categories can be found in men's thinking, and . . . that other people might organize their observations differently' (Belenky et al. 1986:15).[4]

Another book about women, *Social Origins of Depression* by George Brown and Tirril Harris, published a decade earlier in 1976, provides a different illustration of some of the same problems. The methodology of the Brown and Harris study had a very different orientation from that of the Belenky et al. study; the research described in the book is part of a major programme of work undertaken by Brown and his colleagues aimed at documenting the causal associations between stressful life events and circumstances, on the one hand, and mental and physical health, on the other. The core of their methodology is an approach to collecting data from individuals which separates, so far as is possible, the 'subjective' meanings of these data from the 'objective' meanings they are given by a systematic, standardized coding framework developed by researchers and applied to all similar events, circumstances and states, whatever their personal meaning. 'Subjective' meanings are seen as one of many possible sources of bias interrupting the identification of causal relationships rather than, as in the Belenky et al. study, the very stuff of the research itself.

Social Origins of Depression is a meticulous account, based on highly structured interviewing and psychiatric assessment, of the relationship between life events and emotional well-being in a sample of 458 working-class women living in south London. The 'vulnerability factors' which make depression more likely, as identified in the Brown and Harris study (death of a parent in childhood, three or more young children, lack of a confiding relationship, no employment outside the home), have made an important contribution to understanding the links between life circumstances and depression, and have become part of the technical vocabulary of the socio-psychiatric literature. However, Brown and Harris's claims to have established the social roots of depression are made on the basis of empirical data collected only from women. They decided to focus on women because women are more prone to depression than men, and they are also, crucially, more likely to be at home during the day and be willing to submit themselves to lengthy interviews (Brown and Harris 1976:21–2).

This approach leaves the door open to the crucial objection that, were Brown and Harris to have studied men, they might have come away with a different map of the relationships between people's lives and their mental health. There are strong reasons why this might be so. For example, there is much research which shows that men's responses to difficult situations *are* different from women's; they are less likely to 'internalize', to feel adverse events as an injury to self-esteem; the whole pattern of illness and mortality is gender-differentiated in ways and for reasons that, despite a great deal of research, remain largely mysterious (Briscoe 1985; Jenkins 1985; Meddin 1986; Verbrugge 1986; Viney et al. 1985).

These proved but ill-understood differences derive from the fact that gender is not simply a property of individuals, but a set of interactive processes whose influence needs to be accounted for in any research process. Part of the feminist academic project has been to demonstrate the biases embedded in research findings if the influence of gender is ignored. Gender operates by linking biology to social structure in systematically patterned and unequal ways. Men and women appear to be different because they occupy different social positions and 'enjoy' different life-chances. Thus, analyses of research 'findings' which control for variables related to social position generally result in the disappearance of most so-called sex or gender differences (Walker 1984; see also Crawford 1989). Where research is carried out with socially similar samples of males and females, the conclusions that can plausibly be drawn about gender differences may be very different from research which depends on single sex samples only. The best way to derive unwarranted conclusions about one sex (or about human beings in general) is to omit the other. This is similar to the conclusion discussed later in this book that the best way to prove that something works is to omit a comparison situation. What we have here is not a conflict between 'quantitative' and 'qualitative' ways of knowing, but an epistemological error that can be present in *both* ways of knowing.

Women's ways of ageing

The literature on the menopause, a uniquely female experience,[5] provides a third illustration of how misleading it can be to derive research findings from limited samples using research methods which give considerable power to the researcher's own interpretive disposition. For most of its history, the menopause has been socially stereotyped as a negative event. In eighteenth and nineteenth-century Europe, physicians took the view that the menopause was inevitably a period of decay (Grossman and Bart 1979). In the 1990s middle-aged women are thought to suffer from many health problems caused by the menopause, an 'oestrogen deficiency disease'; these symptoms include hot flushes, sweating, tiredness, joint and muscle pains, insomnia, nervousness, weight gain, headaches, back pain, irritability, mood swings, frequent and/or involuntary urination, depression, forgetfulness, low self-esteem, palpitations, dizziness, shortness of breath, loss of feeling in hands and feet, lack of energy, and a phenomenon known as 'restless legs' (Holte 1991; Kaufert et al. 1988; Oldenhave et al. 1993). Of course, the first thing to be said about this formidable list of symptoms is that it is promulgated by a medical profession and a pharmaceutical industry which has a vested interest in women's ill health – in defining women as sick when they may not be, and in prescribing medical remedies when they may not be needed. The medicalization of the menopause has, indeed, been significantly driven by the pharmaceutical industry in Europe and North America, which decided in the late 1930s to make middle-aged women the main targets for the sale of newly synthesized hormones (Oudshoorn 1994). Accounts of their experiences by women themselves, gathered by 'qualitative' researchers such as Emily Martin, present an alternative 'cultural grammar' from the one conveyed by medical representations. This stresses the release of energy and potential that often accompanies the menopause. Within such a grammar, hot flushes are best represented as an experience of embarrassment shaped by a culture in which there is a profound conceptual association between power, rationality and coolness (Martin 1987). Think, for instance, of the metaphors of 'losing one's cool' or not being 'real cool' or getting 'hot under the collar', all of which link emotionality and heat with lack of power.

The underlying epistemological error is potentially the same, however: most of the 'findings' on the experiences of middle-aged women come from studies of women without any comparison group of similarly aged men. The health problems experienced by women are attributed to their being women, without any alternative explanation being considered. In the few studies that do look at health in middle age in both sexes, all the above symptoms except for hot flushes and sweating have been found to be just as common among men (Bungay et al. 1980; Holte 1991; M. Hunter 1990).[6] It is also possible that a number of these symptoms are experienced relatively commonly in the pre-menopausal years, but until this question is asked in a research study, we will not know the answer.

There are patterned differences between groups of women in the extent to which they view the menopause as an illness. Women with manual occupations or those with no paid employment outside the home are most likely to perceive it as troublesome (Martin 1987:170). Table 3.1 illustrates the general point about gender and social circumstances. The data are taken from a large national survey of 9,000 people, the Health and Life Styles survey, carried out in England, Wales and Scotland in 1984–5. Looking at the last two columns of

Table 3.1 Social class and health, men and women aged 40–59

	Social class					
	I, II, IIINM		*IIIM, IV, V*			
	women %	*men* %	*women* %	*men* %	*all women* %	*all men* %
Good/excellent	32	37	25	32	29	35
Poor/very poor	28	23	30	27	29	25
Good but unfit	11	9	14	12	13	11
Good but poor psycho-social health	12	8	13	8	13	8
High illness without disease*	12	7	16	10	14	9
'Silent' disease**	8	15	6	10	7	13

* No disease declared, but a high rate of illness.
** Chronic disease declared without accompanying illness.
Source: Taken from Blaxter 1990:63, table 5.1 (based on 1,301 women, 1,070 men)

the table, one might conclude that women have generally poorer health than men. However, the previous four columns show that in most cases the *direction of the social class difference* is the same for women as for men. Health is related to occupation and to material circumstances, and the distribution of women between occupations is different from that of men. For example, in the UK in 1996 40 per cent fewer employed women than employed men worked full-time, and over half of all employed women, but less than a quarter of employed men, worked in clerical, secretarial, sales and 'personal and protective service' occupations (Macintyre and Hunt 1997; see also Arber and Ginn 1993). Unfortunately the existence of two separate research traditions, one looking at social class and health, the other at gender and health, has meant very little exploration of the ways in which socio-economic position, gender and health interact. Generally it seems that social class differences in indicators of health and illness are often stronger for men than for women. Sometimes the direction of these may be reversed; thus women's body mass index decreases with higher socio-economic position, while men's increases (Macintyre and Hunt 1997).

Looking and finding

It is generally true that what people look for they will find, and that what they are not looking for will probably escape them. Thus, Brown and Harris 'found' causal relationships between social adversity and mental health, and Belenky and colleagues 'found' distinctively female approaches to knowledge. This serendipitous nature of research findings is neatly illustrated by a subtopic which concerned both research teams: child abuse.

A later study by Brown and colleagues (Bifulco et al. 1991) 'found' that being sexually abused as a child emerged as an important vulnerability factor explaining a significant amount of the variation in depression found in the earlier study. Abuse had not been part of the structured interview agenda used in the first study; the later study was done at a time when child abuse and sexual abuse were both in the news. Had a less structured approach been used in the earlier study, it is possible, of course, that some women would have made some sort of reference to abuse. Like Brown and Harris, Belenky and colleagues lighted on the variable of child abuse only when it became part of the media political agenda. Half-way through their study they included a question about it: 'One of the things that we have been finding is that many women were sexually abused at some time in their lives, even as children. Studies have shown that a large percentage of women have been victims of sexual or physical abuse. Has this ever happened to you?' Seventy-five women answered this question, providing 'alarming' 'statistics': 52 per cent said they had been sexually abused; around one in four reported incest in childhood (Belenky et al. 1986:233, 58–9). These figures for sexual abuse are considerably higher than those found in other studies (see Astbury 1996). The relative 'over-estimate' may have had something to do with the 'leading' tone of the question asked and/or with the particular nature of the sample.[7] Such findings suggest the alarming possibility that at least some of 'hearing the silent' may consist of projecting someone's else's imagined reality.[8]

Trustworthiness and its enemies

The basic question is the credibility of research findings or, as one 'qualitative' researcher has put it: 'The soundness of qualitative methodology is the most urgent challenge for those researchers interested in the growth of qualitative inquiry' (Morse 1994a:4). There are three issues here: how 'qualitative' research is done, how it is described, and how its audience is able to decide whether or not its findings are trustworthy.

Reporting research

Judgements of the validity of research can only be made, of course, on the basis of published or otherwise publicly available accounts. One may therefore be judging not so much the quality of the study itself, but the quality of the way it is reported. John Lofland from the University of California carried out an analysis of the reporting of 'qualitative' studies in the early 1970s. His basic data were the papers submitted for publication to a journal and the reports of evaluators asked to assess these papers. Using these data, Lofland discerned five different dimensions and twelve different styles of evaluating and reporting which reflect different attitudes and practices in the field of 'qualitative' research as a whole. For example, the 'Protocol' style involves an author beginning a paper with a few paragraphs about his or her topic being an important social problem, or aspect of social life, or whatever, and then going on to provide a lengthy set of interview transcripts. The underlying assumption here is either that 'the significance of the raw materials is so obvious as to require no further work on the author's part or that the material is so interesting that it renders superfluous any framing by the author' (Lofland 1974:104). In the 'Then They Do This' style, on the other hand, reports are organized around multiple citations of direct observations with little attempt at analysis or synthesis but with much quotation from 'ungrammatical', 'free-flowing' field notes.

The point here is that the evaluated reports in Lofland's survey were alike in their 'qualitative' data collection methods, but they exhibited enormous diversity in styles of reporting and analysis. He concluded that 'qualitative' research is distinct among all forms of enquiry 'in the degree to which its practitioners lack a public, shared, and codified conception of how what they do is done, and how what they report should be formulated' (Lofland 1974:101). His explanation for this state of affairs refers to a dominant conception within social science of 'qualitative' methods as exploration and discovery devices – a general 'ideological celebration of creativity' which provides a mandate for reporting anything social in virtually any manner whatsoever. So 'qualitative' research can be considered 'organizationally and technologically the most individualized and primitive of research genres' (Lofland 1974:110). On this view, anyone can do it. All you need is yourself, some people to watch or talk to, and pen and paper. Another name for this is 'blitzkrieg ethnography': quick forays into research fields by people who use the terminology of 'qualitative' or ethnographic research but misuse its tools by acting as if a deep understanding of something can be gained and transmitted without too much time or difficulty (Rist 1980). It was, perhaps, bound to follow that as 'qualitative' research came increasingly into vogue in the 1970s, committed ethnographers would object that their trade was being plied by people who did not really understand it and were not trained to do it.

Whether training makes a difference is one of many unanswered questions. All research, however conducted, and within whichever paradigm, involves 'an imaginative and creative leap from observed data to synthesis, hypothesis, and generalization' (Dreher 1994:295). The problem lies in how we, the audience, are able to track the nature of that leap. It has been suggested that asking 'qualitative' researchers to account for the processes involved may be like asking a centipede to consider how it is that it is able to move all its legs at the same time: faced with such an epigenetic enquiry, the centipede is simply paralysed (Sandelowski 1994:47).

Magic and myth in ethnography

When Malinowski wrote his account of the trading system operated by South Sea Islanders, *Argonauts of the Western Pacific* (1922), he was well aware that he needed to give some account of how he arrived at the conclusions he did.

'What is then this ethnographer's magic', he asked rhetorically, 'by which he is able to evoke the real spirit of the natives, the true picture of tribal life?' (Malinowski 1922:6). The answer was that the 'magic' inhered in the 'patient and systematic' application of the rules of common sense and scientific principles. Any good ethnography, says Malinowski, has to be based on three principles: the pursuit of scientific aims, living among the natives, and using certain methodological procedures in setting down the rules and regulations of native cultures. Prime among these is the method of 'statistic documentation by concrete evidence' (Malinowski 1922:17), whereby the ethnographer attempts to infer underlying patterns from collected data, where necessary going beyond the surface impressions provided by native informants. Malinowski is in no doubt that the central problem is one of trustworthiness: 'an Ethnographer, who wishes to be trusted, must show clearly and concisely . . . which are his own direct observations, and which the indirect information that form the bases of his account' (Malinowski 1922:15).

Margaret Mead's *Coming of Age in Samoa*, published six years after Malinowski's *Argonauts*, is the most widely read of all anthropological books; it has been translated into sixteen languages, including Urdu and Serbo-Croatian (Tiffany and Adams 1985:27). When Mead sailed for Samoa in 1925 with six notebooks, a portable typewriter and a small Kodak camera, she knew, by her own admission, very little about fieldwork. Her professor and doctoral supervisor, Frank Boas, the intellectual leader of American cultural anthropology at the time, told her she should spend her time sitting around listening to people, but need not bother with any kind of study of the culture as a whole, since this had already been done (Mead 1972:156–7, 147–8). Mead went to Samoa to study female adolescence, because Boas wanted her to; he had a hypothesis that adolescent rebellion in America was due to modern conditions, and did not occur in 'primitive society'. Boas's theory was part of a wider debate about nature versus nurture which had reached one of its recurrent peaks in the 1920s. Other cultures constitute 'natural experiments' in the society–biology relationship, so that studying them can theoretically provide an answer to the question as to whether nature or nurture has the upper hand in shaping human conduct. As Mead herself put it in the case of Samoa, 'Here are the proper conditions for an experiment; the developing girl is a constant factor in America and in Samoa; the civilisation of America and the civilisation of Samoa are different' (Mead 1928:108–9).

Mead's fieldwork in Samoa was concentrated on three villages close to one another. She lived there, in a community of about 600 people, for six months as a member of a group of unmarried girls.[9] The community included sixty-eight girls aged between nine and twenty (Mead 1928:144–5), twenty-five of whom she got to know well. As a result, she concluded that adolescence in Samoa is not marked out or experienced as a difficult period in women's lives. But Mead's Samoan study is probably best known for her conclusions about the place of sex in Samoan adolescence. The image of easy love affairs under the palm trees for which the study is famous derives from Mead's description of a culture which she portrayed as untrammelled by the kind of moral constraints and double standards characteristic of American society at the time. For Samoan girls, as for Samoans generally, sex, said Mead, was 'a natural pleasurable thing' (Mead 1928:112); it was accepted and expected that unmarried girls would have active sex lives, and sexuality did not constitute any kind of moral battleground, either for them or for adults.

Mead's interpretation of Samoan adolescence has since been contested in a much-publicized book called *Margaret Mead and Samoa: the making and unmaking of anthropological myth* published in 1983 by Derek Freeman. In this book Freeman alleges that Mead acquired an incorrect view of Samoan culture because as a woman she was excluded from any participation in Samoan political life, because she was based in a Western household,[10] because she lacked systematic training, because she used an unsystematic, 'homespun' approach to studying the complex problem Boas had set her, and, most of all, because when she went to Samoa, she was already committed to the theory of cultural relativism: she found what she expected to find. Freeman highlights the different view of Samoa he felt was revealed by his own fieldwork. Freeman's Samoa is far from being the romantic paradise attributed to Mead's account; instead it is a competitive masculine society rife with violence and unacknowledged tensions of all kinds. Freeman's research was done in a different village on a different island. A third Samoan ethnographer, Lowell Holmes (1957, 1983) conducted a restudy of Mead's original research in the same village, and concluded that her findings were substantially correct. Further fieldwork in the early 1970s by a fourth anthropologist, Eleanor Gerber, seemingly uncovered a culture considerably stricter about sexual matters than the one Mead had described. This dissonance was explained by Gerber's informants as a consequence of their parents and grandparents simply telling Mead lies in order to tease (or please?) her (Gerber 1975; Freeman 1983:108).

Female anthropologists can penetrate aspects of unfamiliar cultures which are inaccessible to male anthropologists, and vice versa. Freeman and Holmes had privileged access to male informants and to the domain of politics and political rivalries, whereas Mead concentrated on the world of Samoan women. But there is more to it than the sex/gender of informants. As Sharon Tiffany and Kathleen Adams argue in their *The Wild Woman: an inquiry into the anthropology of an idea* (1985), the 'scientific' façade of anthropology hides anthropologists' participation in an ideology of 'the Other' and a view of women which together make some interpretations more likely than others. Forms of thought dominant in Europe in the nineteenth century conceived of women as closer to nature and of other cultures as verging on savagery; fragments from this world-view are liable to influence anthropological 'findings'. Mead's Samoa portrayed women as self-confident and actively and unashamedly sexual; Freeman's vision is inspired by a different view of women, one in which they are dehumanized sex objects. Another factor, of course, is that considerable time had passed between Mead's original observations and the later ones.[11] But whatever else the Mead–Freeman debate is, it is not an example of the deliberate falsification of results – a charge against which Mead's daughter, Mary Catherine Bateson, had to defend her mother after Freeman's book was published (Bateson 1983). It is, instead, best read as an example of the 'untrustworthiness' of 'uncontrolled findings' (Lincoln and Guba 1985:289). While the diversity of the world's cultures provides an experimental laboratory for the anthropologist, the relatively inaccessible processes of anthropological enquiry are quite different from those of a controlled experiment (as the second quotation which heads this chapter (Okely 1994:20) notes).

Ways of judging

What happens when we try to spell out criteria which might be used for judging the trustworthiness of 'qualitative' enquiries? Table 3.2 gives four different lists of criteria compiled by different researchers who have looked at this issue.

While some degree of overlap is clear from a quick reading of the table, there are also issues picked up in one list which are not reflected in the others. There is considerable variety in the language used to describe the criteria. Words such as 'clear', 'adequate', 'careful' and 'systematic' are often used, sometimes in an interchangeable sense, although they are not strictly interchangeable (one may be 'clear' without being 'systematic', for example). Unpacking the various standards suggested in table 3.2 into their component parts gives forty-six distinguishably different criteria. Twenty-eight of these occur in only one of the lists, ten in two, and six in three; only two are common to all four lists (clear description of the sample and how it was recruited, and an adequate description of how the findings/analytic framework are derived from the data).[12]

Reaching agreement on what criteria to include in assessing the validity of 'qualitative' research is clearly not an easy task. Many of the criteria in table 3.2 involve making judgements about whether or not a standard has actually been satisfied; for example, what constitutes an 'adequate' description of fieldwork methods or an 'adequate' description of data analysis, or an 'adequate' description of the context (column 3, points 7 and 12, and column 1, point 11); what is 'sufficient original evidence' (column 2, point 11, and column 4, point 17), and how 'clear' do criteria (column 4, point 3), definitions of problems and purposes (column 1, points 3 and 6) or descriptions (column 3, points 1,3 and 4, column 2, points 2,5 and 7) have to be in order to persuade the assessor that they are clear enough?

Audit and other trails

The lists of criteria in table 3.2 are predicated on an important assumption: that 'qualitative' research *is* the same sort of enquiry as 'quantitative' research, about which such questions of trustworthiness are regularly asked. There are those who would argue the opposite case:

Table 3.2 Four examples of 'quantitative' criteria for judging the trustworthiness of 'qualitative' research

Cobb and Hagemaster 1987	*Mays and Pope 1995*	*Boulton et al. 1996*	*Medical Sociology Group 1996*
1. Understanding of qualitive paradigm	1. Explicit account of theoretical framework and methods stated	1. Clear aim(s)	1. Research methods appropriate to research questions
2. Appropriate references cited	2. Clear description of context	2. Qualitative approach appropriate	2. Clear connection to existing body of knowledge
3. Problem clearly defined	3. Clear description and justification of sampling strategy	3. Clear description of sample	3. Clear criteria for sample selection and data collection and analysis
4. Scope of question manageable within study time frame	4. Theoretically comprehensive sampling strategy to ensure generalizability of conceptual analyses	4. Clear description of recruitment	4. Theoretical justification for selection of cases
5. Purpose is discovery/ description/theory building/ illustration	5. Clear description of fieldwork methods	5. Adequate description of sample characteristics	5. Sensitivity of methods matches needs of research questions
6. Study purpose clearly stated	6. Independent inspection of evidence possible	6. Adequate and appropriate final sample	6. Relationship between researcher and subjects considered, and research explained to 'subjects'
7. Inclusion of literature review if appropriate	7. Clear description and theoretical justification of data analysis procedures	7. Adequate description of fieldwork	7. Systematic data-collection and record-keeping
8. Literature review sufficiently comprehensive	8. Analysis repeated by more than one researcher to ensure reliability	8. Adequate description of data collection methods	8. Reference to accepted analytic procedures
9. Major concepts defined	9. Use of quantitative evidence to test qualitative conclusions where appropriate	9. Systematic data collection	9. Systematic analysis
10. Appropriate initial framework	10. Evidence of seeking out contradictory observations	10. Sensitive data collection	10. Adequate discussion of how findings derived from data
11. Context adequately described	11. Sufficient original evidence presented to satisfy reader of relation between interpretation and evidence	11. Careful records of data	11. Adequate discussion of evidence for and against researcher's arguments
12. Plan for gaining entrée given		12. Adequate description of data analysis	12. Measures taken to test validity of findings
13. Researcher–respondent relationship understood		13. Evidence provided to support analysis	13. Steps taken to see if analysis is comprehensible to participants
14. Role of researcher apparent		14. Sufficient original material presented	14. Clear contextualization of research
15. Issues of qualitative study sampling adequately addressed		15. Evidence that supporting material is representative	15. Systematic presentation of data
16. Characteristics of sample outlined		16. Evidence of efforts to establish validity	16. Clear distinction made between data and interpretation

Table 3.2 *cont'd.*

Cobb and Hagemaster 1987	*Mays and Pope 1995*	*Boulton et al. 1996*	*Medical Sociology Group 1996*
17. Knowledge of qualitative research strategies demonstrated		17. Evidence of efforts to establish reliability	17. Sufficient original evidence presented to satisfy reader of relationship between evidence and conclusions
18. Plan for organizing/ retrieving data outlined		18. Study located in broader context	18. Clear statement of author's own position
19. Framework for analysis stated			19. Credible and appropriate results
20. Problems of validity and reliability addressed			
21. Demonstration of how framework is derived from data			
22. Understanding of ethical issues			
23. Importance of study to subject area outlined			

> Except at a very high level of abstraction, *it is fruitless to try to set standards* for qualitative research per se. (*Howe and Eisenhart 1990:4; emphasis added*)

> The greatest concern today is that many qualitative researchers are using quantitative criteria to interpret, explain, and support their research findings without realizing *the questionable practice* or the inappropriateness of such efforts. Using quantitative criteria to evaluate qualitative studies is clearly inconsistent with the philosophy, purposes, and goals of each paradigm. *(Leininger 1994:97; emphasis added)*

According to this logic, the 'truth' that is sought through 'qualitative' research is a special kind of truth: as 'socially and historically conditioned agreement' (J.K. Smith 1984:380). What is true is simply what people at the time can agree is true or trustworthy. Such a position disputes the 'ontological creed' of the 'positivist' 'paradigm' – that the object of social research is to find out how things really are. 'Reality' can only be a property of a mental framework, and what counts as knowledge can only be a human construction. If 'reality' does not exist, then establishing how best to assess whether research findings adequately represent this must be a senseless task.

Some aspects of the argument about applying standards of trustworthiness across research 'paradigms' remind me of a discussion that took place in the early 1980s in a medical research unit where I was then working. The subject was a proposed study to investigate the accuracy with which pregnant women's experiences of fetal well-being predicted health problems in their babies. The method being suggested was one in which women would be asked to quantify their experiences of the ways in which babies moved in the womb by counting the number of movements per unit of time on a regular daily basis; these observations would be recorded, and the relationship between them and the babies' state of health would be looked at. Dissenting voices argued that mothers' feelings about the health of fetuses were 'qualitative' in nature and could not be quantified in this way. In seeking to impose one way of knowing on another, the proposed methodology was simply an improper translation between incommensurable languages. Moreover, as it was already known that there was a problem in getting health professionals to take mothers' anxieties about their fetuses seriously, this was the bit of the predictive chain that really needed attention.[13]

If criteria derived from a 'foundationalist' standpoint should not be used, then what alternatives are there? One suggested solution is that of the 'audit trail'. This model is derived from the process of checking financial accounts, and depends on access not only to the raw data but also to all data reduction and analysis products (field notes, notes about theories and concepts), processes and products of data synthesis (themes, definitions, findings), process notes (methodological procedures, etc.), material relating to 'intentions and dispositions' (the original proposal, personal notes about the research) and 'instrument-development information' (how the interview schedules etc. were developed). A fully developed audit trail is a detailed, laborious and time-consuming process involving a sequence of predetermined stages and resulting in an 'attestation' that the auditor found (or did not find) evidence that the results of the research can be trusted (see Halpern 1983; see also Lincoln and Guba 1985:319–27). But what may appear to be a relatively straightforward process in theory can prove more awkward in practice. For example, in an audit of an evaluation carried out by two auditors of a child care information and referral service in the USA, the two auditors could not agree as to whether the evaluation had been well done (Greene at al. 1988). Much depends on judgement. The audit concept does not itself, therefore, dispense with the need for a set of agreed standards.

Non-foundational criteria?

So what alternative criteria might be used to determine trustworthiness in 'qualitative' enquiry? Table 3.3 compares four sets of proposals made by those who argue that special criteria are needed to establish the trustworthiness of 'qualitative' research findings. Those shown in the first column come from the work of Egon Guba and Yvonne Lincoln in the USA, who have been at the forefront of the effort to establish standards for 'qualitative' research. Their work stemmed from the rise in 'qualitative' research papers published in education journals in the late 1970s; the editors of these journals were at a loss as to how to judge the rigour of these studies. The result was the set of 'trustworthiness' criteria shown in table 3.3. The other lists were produced by 'qualitative' researchers who have all given some thought to how the credibility of their enquiries might be judged.

Like table 3.2 above, the criteria proposed in the four lists in table 3.3 have some overlap and some differences. A total of twenty-five different criteria can be identified. Of these, thirteen appear once and eleven twice: one (the collection of 'thick' data) appears in three lists, but none in four.[14] 'Thick' data/description is a popular notion among 'qualitative' researchers. The term was devised by philosopher Gilbert Ryle, and was adapted to anthropology by Geertz (1973); what it means is a detailed and literal description of the entity being studied, including interpreting the meaning of descriptive data in terms of cultural norms and values (Guba and Lincoln 1981:119).

A comparison of the two tables does, however, suggest that the 'qualitative' criteria are not strikingly different from the 'quantitative' ones. That is, the former represent an adaptation of the latter, but the issues each are intended to address are the same.[15]

Checking with members

Two of the lists in table 3.3 suggest that taking data/interpretations back to the people from whom they came in the first place ought to be part of the attempt to establish trustworthiness. This is not a new idea; Alfred Schutz's 'postulate of adequacy' required that scientific propositions be understandable

to community members, because if scientific and common-sense thinking are not overlapping terrains, then science must have got something wrong (Schutz 1967). The implication is that if one's research participants agree with the interpretations which constitute the research 'findings', then more confidence can be had in their reliability. This kind of iterative process, in which the researched join forces with the researchers in a collaborative effort to give birth to the research product, is also a feature of the feminist argument for 'qualitative' research. It is an essential aspect of the case for decreasing the power differences between researcher and researched; neither party should dominate either the process of deciding who should take part in research or that of determining what the research 'means'.

An early experiment in 'member checking' which pre-dated the feminist call for more democratic practices is the study of 'dual-career families' carried out by Rhona and Robert Rapoport in the late 1960s. The book of this name, published in 1971, described the lives and experiences of five British families in which both spouses were employed in professional careers and were also parents. The intention of the research (which included interviews with a total of sixteen couples) was to look at similarities and differences in the ways such families cope with the demands of their careers and their family lives, at strengths and weaknesses of the dual-career family pattern, and at the extent to which it may serve the interests of men and women rather differently. Both partners were interviewed a minimum of four times over a period of two years. On one of these occasions the couples were given for discussion a write-up of their particular 'case'. The five couples whose cases were detailed in the book were given a further opportunity to comment on the material the Rapoports wanted to publish. 'The feedback of the reports to the couples', they say, 'was regarded partly as an ethical requirement and partly as a validity check. For populations of the kind studied . . . we assumed that their own perceptions together with ours would provide the most valid approximation of "the truth"' (Rapoport and Rapoport 1971:324). It was not a straightforward matter. The initial feedback proved 'a point of some tension'. The couples had spoken freely – many of them said more freely than they would have talked to their best friends – and this sometimes brought into the open elements that had not been explicit, either to themselves or to each other, before. Such tensions were of much interest to the researchers, and of course highly relevant to the purpose of the research. But in order to satisfy the research participants, most of them had to be excised. Getting the consent of the five couples who were the focus of the detailed case-studies was even more difficult. Partly because they felt they might be recognized, but also because of a general feeling of embarrassment, they exercised 'quite severely' their right to veto the inclusion of certain materials. Interestingly, the men wanted to excise more than the women did – in part, it seems, because of a general view that such research on family life is really women's business anyway (Rapoport and Rapoport 1976).[16] Two of the five case-studies ended up so truncated that they were nearly dropped from the final version, which certainly reads in places as an account which has been 'sanitized' by the omission of anything remotely conflictual or controversial.

Table 3.3 Four examples of 'qualitative' criteria for judging the trustworthiness of 'qualitative' research

Lincoln and Guba 1989	*Leininger 1994*	*Muecke 1994*	*Popay et al. 1998*
Credibility 1. Prolonged engagement (at the enquiry site to establish rapport and immerse the researcher in the culture to be studied)	1. *Credibility* – ensuring that the researcher uses active listening, reflection and empathic understanding to grasp what is 'true' to informants in their lived environment	1. The research interprets one social group to the large society or another society	1. The privileging of 'subjective meaning' – the research illuminates the subjective meaning, actions and context of those being researched
2. Persistent observation (sufficient observation to collect data relevant to the research topic)	2. *Confirmability* – repeated direct participatory and documented evidence observed or obtained from primary sources	2. The research participants would find the research an honest and caring description of them in their situation	2. Responsiveness to social context – the research design is adaptable/responsive to real-life settings
3. Peer debriefing (discussing the research with a disinterested peer)	3. *Meaning-in-context* – understanding data within holistic contexts (participants' environments)	3. The conceptual orientation of the researcher is acknowledged and coherently linked to the data	3. Purposive sampling – the sample produces the knowledge necessary to understand participants' location in structures and processes
4. Negative case analysis (revising hypotheses until they account for all known cases)	4. *Recurrent patterning* – using repeated experiences, events, etc. to identify patterns of sequenced behaviour	4. The relationship between the researcher and the researched is explicitly assessed for its influence on the data	4. Adequate description – the reader can interpret the meaning and context of what is researched
5. Progressive subjectivity (the researcher monitors her/his own construction/ biases	5. *Saturation* – full immersion by the researcher in the phenomena being studied; getting 'thick' data to know fully what is being studied	5. The anonymity and integrity of research participants are protected	5. Data quality – different sources of knowledge about the same issues are compared
6. Member checks (checking data with research participants)	6. *Transferability* – examining general similarities of findings in similar environmental situations	6. The sources of the data are sufficiently clear for the reader to assess the adequacy appropriateness and breadth of coverage of the data	6. Theoretical and conceptual adequacy – the research describes the process of moving from the data to their interpretation
Transferability 7. Using 'thick description' to establish the transferability of the findings to other settings		7. 'Thick description' is used to explore and contrast diverse sources of data	7. Typicality – claims are made for logical rather than probabilistic generalizations
Dependability 8. Ensuring that the research process is trackable and documentable		8. Data were obtained from a variety of sources	
Confirmability 9. Ensuring that data can be tracked back to their sources and that the logic connecting data and interpretations is explicit		9. Data were gathered accumulatively and cyclically leading to reformulation of questions	
		10. The research narrative is competent literature	

But, unsurprisingly, it is in the feminist research literature that we find the most fully developed examples of 'member checks'. These also uncover flaws in the logic underlying the process. A much-quoted example is the study by Joan Acker and colleagues of women's transitions into the labour market. 'We were convinced', they said, 'that middle-aged women who had spent most of their lives as wives and mothers had been ignored by much of the [women's] movement and we hoped that we might give voice to some of their perspectives' (Acker et al. 1983:426). Using unstructured in-depth interviews, Acker and colleagues studied sixty-five women, following a subgroup of thirty for up to five years. During this process they showed much of their written material to the women concerned. 'We have to admit to some reluctance', they acknowledged, 'to share our interpretations with those who, we expected, would be upset by them. There was a potential conflict between our feminist frame of reference and their interpretations of their own lives. Our solution to this conflict was not to include them as active participants in the analysis of our research.' A particular problem was women who were not in employment and who defined themselves as 'very independent' but whom the researchers viewed as both structurally and personally dependent. Acker and colleagues concluded that they had not solved the problem of how to do research democratically, in line with a feminist ethos, and that it is hard to avoid the position of researchers as those with 'the power to define'.

Perhaps it is always easier to preach than to practice. Ethnographic research, as Judith Stacey, among others, has observed, appears to be ideally suited to the requirements of feminist research, because the very closeness of researcher to researched suggests a dissolution of the traditional researcher–researched hierarchy. But after two and a half years of trying to do ethnographic fieldwork in a feminist way, Stacey found herself 'wondering whether the appearance of greater respect for and equality with research subjects in the ethnographic approach masks a deeper, more dangerous form of exploitation' (Stacey 1988:22). She frequently found herself in a position of knowing things about informants that they did not know she knew. How do you check with research participants the validity of data which they would not approve of you having in the first place? In one particular case, of the disclosure of a closet lesbian relationship, Stacey felt obliged to leave this out of her ethnographic account. While feminist ethical principles suggested that she respect her research participant who wanted it left out, deciding to do this resulted in a distortion of the 'truth' reported in the final write-up of the study (as well as colluding with the traditional homophobic silencing of lesbian experience).

Everything informants share with researchers is ultimately grist for the researchers' mill. All research represents an intrusion and intervention into a pre-existing system of relationships; thus, taking research data back to the researched is an example of a social event rather than a scientific test (Bloor 1997).

Triangulation

For many 'qualitative' researchers, another answer to the problem of establishing credibility for 'qualitative' research is called 'triangulation'. The term comes originally from broadcasting; radio triangulation means determining the point of origin of a radio broadcast by using directional antennae set up at the two ends of a known baseline. A triangle is then created by measuring the angle at which each antenna receives the most powerful signal; using geometry, the source can be pin-pointed (Lincoln and Guba 1985:305). Applied to research methods, the idea is usually that taking data from several sources will increase one's chances of being able to establish trustworthy results. As Michael Patton (1980:329) phrases it, 'There are basically two kinds of triangulation that contribute to verification and validation of qualitative analysis: (1) checking out the consistency of findings generated by different data-collection methods and (2) checking out the consistency of different data sources within the same method.' The aim is verification, not falsification (or both). 'Triangulation' is a term which makes many ill-specified appearances in research grant applications, creating the illusion that researchers have at hand a ready-made technique for dealing with disbelievers. But what happens in practice, when triangulation is tried, rather than merely appealed to, is that data from different sources or collected using different methods may conflict.

One example of this is a study of women's experiences of the menopause. A survey of 1,713 Finnish women carried out by Hemminki and colleagues provided the opportunity to compare the pictures gained of menopause symptoms in answers to a structured and an open-ended question (Hemminki et al. 1995). The structured question listed seventeen health problems, and asked the women to say which they had experienced in the last two weeks. The open-ended question enquired what symptoms the women had experienced *which they regarded as linked to the menopause*. The three symptoms most commonly ticked in answer to the structured question were tiredness, hot flushes and backache; the three most commonly reported answers to the open-ended question were perspiration, hot flushes and irritability. Table 3.4 shows the percentages of women reporting the most commonly mentioned symptoms in both sets of answers.

If one took the first set of answers only, menopause would appear to pose health problems for significant numbers of women. But, judged by the second set of answers, menopause is an altogether less troublesome experience. Even hot flushes are cited as menopause symptoms by less than a third of women. Which is the true answer?

A second example comes from the childbirth field. Health professionals and medical researchers are often sceptical about the extent to which women can be trusted to remember correctly aspects of their childbirth experiences. They assume that medical case-notes are more reliable. An experimental study of social support in pregnancy provided an opportunity to look at this and other methodological issues, since data were collected from medical notes and from mothers, using both self-administered questionnaires and semi-structured

Table 3.4 Women's answers to a structured and an open-ended question about menopause symptoms

	Structured question N = 436 (%)	*Open-ended question N = 303 (%)*
Tiredness	45	9
Hot flushes	41	40
Headache	35	4
Joint/muscle aches	24	6
Vertigo	24	7
Sleeplessness	22	15
Depression	21	13
Irritability	19	18

Source: Adapted from Hemminki et al. 1995:83 (women aged 50–4)

interviews. Table 3.5 shows the extent of 'agreement' between medical notes and mothers on certain key features of pregnancy and motherhood. The table can be read two ways: either that there is more than 90 per cent agreement between the two sources for most items, or that there is significant disagreement, particularly for the last item, neonatal problems (where the medical notes were more likely to record problems than the mothers were). Perhaps most surprising are the 3 per cent of cases where there was disagreement about the baby's sex; these represent fifteen instances in which mothers said the sex of the baby they had given birth to was different from the sex recorded in the hospital case-notes.

Table 3.5 Agreement between hospital records and mothers* on questions about pregnancy and the baby's condition

	Hospital and mother:	
	Agree (%)	*Disagree (%)*
No. of previous pregnancies	91	9
Bleeding in pregnancy	95	5
Baby's sex	97	3
Neonatal problems	81	19

* Mothers' information taken from home interviews for number of previous pregnancies, and from postal questionnaires for other items
Based on N = 467.
Source: Oakley et al. 1990:479

Table 3.6 Number of cigarettes smoked daily at the beginning of pregnancy as reported in hospital records and by mothers in home interviews and postal questionnaires

No. of cigarettes smoked	*Hospital records (%)*	*Home interviews (%)*	*Postal questionnaires (%)*
1–9	23	21	11
10–19	57	57	42
20 or more	20	22	47
Total	100	100	100

Based on N = 75 women
Source: Oakley et al. 1990:483

Table 3.6 comes from the same study. It shows data relating to the incidence of smoking in pregnancy from three sources – medical notes, interviews and questionnaires – for the same group of women. The figures derived from medical records and home interviews (with research midwives) are very similar. But estimates of cigarette smoking are significantly higher in the self-administered questionnaires. Actually, the mean number of cigarettes smoked went from 11.7 in medical notes to 13.0 in the home interviews to 16.6 in the postal questionnaires. What does this mean? Given the moral reprobation that smoking in pregnancy induces in some people, especially health professionals, it is reasonable to suppose that the answers mothers gave in the questionnaires were the 'true' ones. It is easier to tell the truth to an anonymous sheet of paper than face-to-face to someone who is likely to have a reaction (visible or hidden) to this piece of information. Such an explanation, however, does not work in relation to the discrepant instances of babies' sex in the previous table. Here, the most likely explanation is either incorrect recording by health professionals[17] or confusing layout of the questionnaire (or even incorrect data extraction by the researchers).

The third example comes from an investigation of the health status of older women. Moyra Sidell used three approaches to collect data on this: national mortality and morbidity statistics; data from two large-scale sample surveys; and in-depth interviews with thirty older women. What she found was 'a mass of paradox and downright contradictory evidence', with the 'hard' data giving a picture of older women living a long time with a high burden of disease and constantly going to doctors; the 'medium-textured' data yielding a more optimistic picture as regards subjective health perceptions; and the 'soft' data demonstrating much resilience and stamina, with women rarely taking to their beds and being extremely reluctant to use the health services. Sidell ends up viewing with nostalgia the days of 'methodological puritanism' when 'You were not required to perform intellectual contortions, and we qualitative types didn't have to do our sums' (Sidell 1993:117–18).

In other words, using multiple methods does not lead to sounder explanations in simple additive fashion. It may even be the case that 'the neat dovetailing of the pieces of a research puzzle should be cause for suspicion. Unanimity may be the hallmark of work in which the avenues to other explanations have been closed off prematurely' (Trend 1979:68). This conclusion was reached in an attempt to synthesize 'qualitative' and 'quantitative' data collected in a large-scale study conducted in the USA in the 1970s of the effects on low income families of housing allowances (see chapter 9). In this study, the two sorts of data could not easily be reconciled. 'Quantitative' impact data pointed to the programme being a success, but observational data suggested the opposite conclusion.

Differences between data sources have received a good deal of attention under the heading of 'validity' in questionnaire and interview research. For example, studies reported by William Belson used two interviews (an 'ordinary' and an 'intensive' one) separated by a short time period in order to find out how research participants actually interpreted particular questions. More than two-thirds did not interpret the questions as the questioner intended (Belson 1986). A famous example was the chocolate study. This was commissioned by confectionery manufacturers, who suspected that their market information was inaccurate. A sample of 295 adults in London were double interviewed (by market research and 'specially trained' interviewers) about their purchases of twelve chocolate products over the seven days preceding the interview. Comparing the two sets of data showed that the number of products bought in the week before the interviews was a fifth larger in the market research data; for particular products (notably Fry's Chocolate Cream), this made a 50 per cent difference to estimates of their market share. The discrepancies between the two sets of results were due to those questioned interpreting the time period idiosyncratically, being reluctant to say 'no' to the interviewer all the time, wanting to impress her/him, not taking the interview seriously, or simply forgetting what they had bought (Belson 1966).

Deceiving the sane

The interviewed can deceive their interviewers, intentionally or by simple omission. Deception is probably pervasive in all forms of research. Although the contentions of 'qualitative' researchers would have them possess the moral edge here – face-to-face methods supposedly lacking the objectifying distance of 'quantitative'/experimental ones – yet tensions exist between the principle of democracy and the very goal of research as producing warrantable knowledge.

Gellner's story about the man who could prove he was sane, whereas most people are unable to do so (see chapter 2), raises many questions, one of which concerns the ability of medical professionals to identify correctly sanity and insanity. D. L. Rosenhan (1973) put this to the test by conducting a small experiment in which eight sane people gained admission to twelve different hospitals on the East and West coasts of the USA. These 'pseudo-patients' included the researcher, psychologists, a painter and a housewife. They all contacted the hospitals complaining that they could hear voices telling them their lives were empty and hollow. Apart from inventing these voices and using pseudonyms, they gave medical staff accurate details of their life histories. Once admitted to hospital, they dropped the pretence of psychiatric symptoms and behaved normally. Part of the deal was that, if they were successful in getting into hospital, it was up to them to negotiate their own release.

Length of hospital stay ranged from seven to fifty-two days, with an average of nineteen days; around seven minutes a day on average was spent with medical staff; some 2,100 pills of many varieties were prescribed (only two were swallowed, the rest being pocketed or consigned to the lavatory, a common practice among 'real' patients). None of the medical staff correctly identified sanity, and all the pseudo-patients were released with diagnoses of schizophrenia.[18] The experiment was instructive in demonstrating how strong the bias is among doctors to make 'type II' errors – to call healthy people sick. But was the deception justified?

Laud Humphreys' study of homosexual encounters in public places, *Tearoom Trade* (1970), is often cited as the classic example of this particular dilemma in research practice. The preface to the book, by the sociologist Lee Rainwater, calls *Tearoom Trade* part of the great tradition of studies of city life going back to Henry Mayhew. Humphreys' study made an important contribution to criminological knowledge in correcting the view then prevailing that the anonymous sexual encounters of 'lavatory homosexuality' were the province of the single and lonely. Most of the men Humphreys observed were married and predominantly heterosexual. But the only way he was able to know this was because he noted down the registration numbers of cars belonging to men he observed in such encounters, and with the help of 'friendly policemen' tracked down their addresses. He then interviewed the men, ostensibly for a study of health which had nothing to do with his own study. Humphreys, in real life an Episcopalian minister, pretended to be a gay man in order to do his research:[19] 'I am convinced', he says, 'that there is only *one* way to watch highly discreditable behaviour and that is to pretend to be in the same boat with those engaging in it' (Humphreys 1970:25).

Most research using 'qualitative' methods does not give rise to quite the ethical dilemmas represented by *Tearoom Trade*. But participant observation, as Shulamit Reinharz found (see chapter 2) is likely always to carry an element of 'dishonest' social interaction, and methods such as life histories, interviews and focus groups do not *in themselves* guarantee greater participation of research 'subjects' in developing the product of the research.[20]

Conclusion

In the war between 'quantitative' and 'qualitative' paradigms, 'qualitative' research is, as we saw in chapter 2, the alternative paradigm fighting for a general democratization of ways of knowing. This chapter has looked at some of its key contentions: that listening intensively to the silent as subjects of research produces a more authentic, and hence reliable, knowledge; that generalizations about knowledge can proceed from highly limited research samples; that asking questions produces honest and trustworthy answers; that consensual standards for 'good', 'qualitative' research can be specified; that sharing research findings with the researched is an uncontentious exercise; that the products of 'qualitative' research can be verified using data from different sources or methods; that 'qualitative' research is inherently less deceptive than the 'quantitative' or experimental kind. All these contentions are problematic. The language of paradigms beguiles us into thinking that the alliance between 'qualitative' enquiry and the world of the social will somehow guarantee that such research is both ethically and scientifically 'better' at representing people's interests. But in-depth interviewing and ethnographic observations may only bring us nearer to the truths that flourish inside researchers' heads. The laudable goal of feminist research, to do away with the traditional 'objectification' of research participants, may itself be a contradiction in terms: however one looks at it, from within whichever paradigm, researchers *are* the ones with 'the power to define'. While 'qualitative' research may have problems of credibility, so does 'quantitative' research. Indeed, as I shall argue later, these are to a large extent the *same* problems. One might reasonably argue that the distinguishing mark of all 'good' research is the awareness and acknowledgement of error, and that what flows from this is the necessity of establishing procedures which will minimize the effect such errors may have on what counts as knowledge.