Sampling in qualitative research has had a hard time. On the one hand, it has been long neglected by many qualitative researchers as a mere positivistic worry; on the other hand, it has been undervalued by survey researchers because of the use of non-probability methods. Qualitative researchers often maintain that qualitative research does not need to sample or to consider seriously sampling issues, arguing that the most theoretically significant and important studies in field research (accomplished by Gouldner, Dalton, Becker, Goffman, Garfinkel, Cicourel, Sudnow and so on) were based on opportunistic samples. This argument may lead to the idea that thinking about issues of sampling, representativeness and generalizability is a waste of time. However, defining sampling units clearly before choosing cases is essential in order to avoid messy and empirically shallow research. As a matter of fact, in contemporary organizational research the problem of representativeness is a constant and growing concern of many researchers. In addition, traditional qualitative researchers often forget that sampling is an unavoidable consideration because it is, first of all, an everyday life activity deeply rooted in thought, language and practice.

On the other hand, survey researchers have often disqualified qualitative research because it is based on non-probability samples. Their reasoning, expressed in numerous research methods texts, lies in the following logic:

1. Only surveys and polls use representative samples;
2. To be representative a sample needs to be drawn up using probability methods;
3. As qualitative research is based on non-probability samples, its samples are not representative;
4. Only findings from a representative sample are automatically generalizable to the population;
5. (Therefore) representativeness leads to generalizability;
6. If research is not carried out on a representative sample, its findings are not generalizable;
7. Findings of qualitative researchers are not generalizable.

These sentences have become such commonplaces that they form an undisputed part of most researchers’ background assumptions. However, survey researchers do not realize that in social science (as I shall show later) probability samples are achieved very rarely because of a number of theoretical and practical reasons (such as the nature of social or cultural objects, the lack of a population list for most variables, the phenomenon of non-response, and so on). This implies that it is in survey research as well that issues of sampling, representativeness and generalizability need to be reframed in a new perspective.

For decades, field and survey researchers avoided each other and failed to respect the scientific credentials of their respective works. However, recently a conciliatory offer has been made by some qualitative researchers, legitimating both ways of doing research by stating that there are two kinds of generalizations: a generalization about a specific group or population (which aims at estimating the distribution in a population) and a generalization about the nature of a process. Sampling requirements are completely different in the two cases. The former generalization, which is implemented in a survey or a poll, is based on statistical logic; the latter, applied in field research, is based on the notion of ‘theoretical sampling’ (Glaser and Strauss, 1967).
To make clearer the difference between these two kinds of generalization, some authors have called the second one ‘transferability’ (Guba, 1981), ‘analytical generalization’ (Yin, 1984), ‘extrapolation’ (Alasuutari, 1995) and ‘moderate generalization’ (Williams, 2002).

While this offer was welcome and undoubtedly wise from a political point of view, it seems not to solve completely the theoretical and practical problems involved in notions of representativeness and generalizability.

In this chapter I shall try to follow an alternative path, in search of more adequate answers to the two main linked questions usually raised in social research:

1. How do we know to which extent our cases (sample) are representative of all members of the population from which the cases were selected?
2. Can we generalize from a few cases (a sample) to a population without following a purely statistical logic?

To understand this complex matter, several progressive analytical steps will be followed. First, some (misleading) commonplaces on which statistical sampling is grounded will be deconstructed, showing how probability samples are rarely achievable in survey research. Second, the theoretical legitimacy of non-probability samples will be described. As a matter of fact, experiments (reckoned by many scientists as the best possible examples of scientific procedure) are not based on probability samples. Third, the research practice of disciplines (such as palaeontology, ethnology, biology, astronomy, anthropology, cognitive science, linguistics, and so on) whose scientific work is based on few cases, will be considered to see if we can learn something useful for qualitative social research.

In order to answer the first question, we need to go back to the core problem of representativeness: the variance of the phenomenon under study. In social studies representativeness is often a practical matter, hardly ever an outcome of automatic (statistical) procedures, which are often useless (as well as difficult to implement) because in social research we look at the social significance of samples instead of a statistical logic.

In order to answer the second question as listed above, we need to distinguish two analytically different problems that are usually confused with one another: the representativeness of samples and the generalizability of findings. Offering these issues as two sides of the same coin neglects the existing social space between these two activities. In the social process that starts with the creation of a representative sample and ends with the generalization of findings, the researchers’ activity is constantly driven by biases and organizational obstacles.

Issues of sampling, representativeness and generalizability need to be faced in a practical way, quite differently from the abstract way usually suggested in textbooks where methodological principles and rules stand on their own with only a weak relation to practice. On the contrary, it is necessary to approach the whole question of sampling sequentially because it is misleading to plan the whole strategy before starting. In order to obtain representativeness, the sampling plan needs to exist in dialogue with field incidents, contingencies and discoveries.

Some of these issues will be discussed through an (invented) dialogue between a teacher and a class. The content of this conversation is a collection of questions, doubts, observations and objections that I encountered several times when teaching research methods over a long period of time.

Setting: a professor (P) with students (S) in a class

P: Today we'll talk about how to sample.
S1: Are we going to taste some wine?
Class: ha ha ha (laughing)
P: What? … Oh yeah, I get it. If I think it over … in fact it's not a silly joke ... it has some truth.

I told you in earlier weeks that there are many links between ordinary language and scientific language, or between common sense and scientific knowledge. This is just another case.

SAMPLING IN EVERYDAY LIFE

P: Before becoming a scientific procedure sampling probably was, and still is, a practical activity of daily life: the cook takes one macaroni out of the pot to know if the pasta is ready to be served; the buyer tastes to choose a wine or a cheese; the teacher asks a student some questions to assess his or her knowledge on the whole syllabus. In everyday life social actors sample constantly. If we refer to traditional
classifications of knowledge (see Frege, 1892; Peirce, 1902; Gomperz, 1905; Ogden and Richards, 1923; Morris, 1938; Whorf, 1956, Popper, 1972; Marradi, 1994), which divide it in three worlds (see Figure 28.1), we notice that there are sampling activities in all three worlds.

With regard to thought, cognitive psychologists Kahneman and Tversky showed that human beings use particular ways of reasoning called 'heuristics'. One of these is the heuristic of representativeness (Kahneman and Tversky, 1972; Tversky and Kahneman, 1974), that is, people's tendency to generalize based upon a few observed characteristics or events. With regard to the world of language, the same function is performed, as Becker (1998) has stated, by 'synecdoche, a rhetorical figure in which we use a part for something to refer the listener or reader to the whole it belongs to' (1998: 67). Finally, in the world of action the representative function is performed by the sample. The seller shows a sample of cloth to the customer; in a paint shop the buyer skims through the catalogue of shading colours in order to choose a paint.

**Figure 28.1** The three worlds of knowledge.

SAMPLING, REPRESENTATIVENESS AND GENERALIZABILITY

or grave goods specimens, and so on. As Becker states:

Archeologists and paleontologists have this problem to solve when they uncover the remnants of a now-vanished society. They find some bones, but not a whole skeleton; they find some cooking equipment, but not the whole kitchen; they find some garbage, but not the stuff of which the garbage is the remains. They know that they are lucky to have found the little they have, because the world is not organized to make life easy for archeologists. So they don’t complain about having lousy data. (Becker, 1998: 70–1)

S2: Yes, but still they do not know if these samples are representative …

P: This concerns the nature of the outcome (the sample) and the associated process for obtaining it, not the outcome itself. We’ll face this issue in a few minutes. For now the only thing you have to remember is that the process or procedure employed to select cases (e.g. a random one) has nothing to do with the concept of sampling itself. They are two separate issues.

**FROM EVERYDAY LIFE TO SCIENCE**

P: Yes, please (seeing a student with a raised hand).

S2: But we don’t know if they are representative … it could be a fake or damaged piece of cloth]

P: Wait, wait … you are going too fast. Before arriving at this point we first need to define a sample.

After having specified a population, the researcher decides if he or she will collect information on all its individuals (complete enumeration) or on a sub-set only (sample). So the sample is just a sub-set of cases. Nevertheless, it might also be one case only. Various scientific disciplines use one case as sample: geologists with a fossil or a fragment of stone, palaeontologists with fragments of skull, bones or a skeleton, archaeologists with ornaments or grave goods specimens, and so on. As Becker states:

Archeologists and paleontologists have this problem to solve when they uncover the remnants of a now-vanished society. They find some bones, but not a whole skeleton; they find some cooking equipment, but not the whole kitchen; they find some garbage, but not the stuff of which the garbage is the remains. They know that they are lucky to have found the little they have, because the world is not organized to make life easy for archeologists. So they don’t complain about having lousy data. (Becker, 1998: 70–1)

S2: Yes, but still they do not know if these samples are representative …

P: This concerns the nature of the outcome (the sample) and the associated process for obtaining it, not the outcome itself. We’ll face this issue in a few minutes. For now the only thing you have to remember is that the process or procedure employed to select cases (e.g. a random one) has nothing to do with the concept of sampling itself. They are two separate issues.

**SAMPLING IN SOCIAL SCIENCE**

P: One important question is that of why we sample.

Sampling was not always dominant in social sciences. Once it was common to
select the whole population of residents (a census), car owners, magazine subscribers, people in the phone directory, and so on. The main problem of this method was not the difficulty of reaching the whole population, but the difficulty of obtaining feedback because many of them did not send back mailed questionnaires. However, at that time researchers trusted in the ‘law of large numbers’, that is, if you have collected a great number of cases (it doesn’t matter how big is the related population) you are all right and do not have to worry. However, today we can state that those studies were not, in a strict sense, studies of populations but of … samples, that is, sub-sets formed by people who participated through their answers.

S1: So what’s new about the proposal about sampling from the statisticians George Gallup and Elmo Roper, the two ‘bad guys’ who forecast correctly the 1936 US presidential election? Samples were, even though unconsciously and against researchers’ intentions, used also before 1936!

P: Yes and no. You are right in pointing out that the change was not from the population method to the sampling method. However, the change was in the nature of samples. In the population studies, samples were always an output of the failure of recruiting participants; in the subsequent studies, samples were a rational and deliberate input. Researchers wanted to ‘rule the game’ (that is, to decide who to contact and how many) instead of waiting for participants’ to decide whether to answer a mailed questionnaire.

S1: That’s it?

P: No, of course. There were other major advantages to this new method of recruiting participants, such as lower costs, saved time, and fewer human resources needed to contact them. However, in the whole business there is one enormous disadvantage, which troubled statisticians for long time: a sample will always be a sub-set of a population.

S2: What does that mean?

P: It means that a sample always has a drawback. The idea of a sample was born from the material impossibility of examining the whole population (see Pinto, 1964: 623; Galtung, 1967: 51; Perrone, 1977: 71). Also, Popper’s refutability and falsification principles are based on such a recognition of the impossibility of the ideal of checking and controlling all members of a specified population. We wish we could do it but we cannot. Also, when this is needed, as in a census, some residents are always missing. For example, the 1990 US census failed to include about 10 million residents and double counted about 6 million others – for a total net undercount of 4 million people. So for the 2000 census the US Census Bureau decided to use scientifically proven statistical methods to account for non-respondents. A paradox!

S3: How can we overcome this problem?

P: There is no way to get out of it even though natural scientists believe that it is possible through the idea of the representativeness of a sample.

S1: Thank God! Finally we arrive at the point of representativeness…

P: Yes, but you will have many surprises. The sampling procedures in use in social sciences come from biology. This was made possible by the positivistic climate of the beginning of the twentieth century when social disciplines, in order to be scientific, imported language and procedures from natural sciences. It was a bad period and its consequences unfortunately are still present. Anyway the first problem, long and welldebated, arises from this emulation of natural science. Many authors have observed that biology or physics differ from social science in the nature of their units of analysis (Goode and Hatt, 1952: 327): a Russian chemist has no doubt that his or her atom of nitrogen acts in the same way as the atom of nitrogen of his or her Japanese or French colleague; however, no human being is similar to another one. So the presupposition of homogeneity, on which sampling in natural science is based, does not work in social
science because individuals are not interchangeable in the same way as atoms of nitrogen or little balls in a box.

S2: Is that all?
S3: It is an old criticism based on an abstract argument and has been solved anyway by survey researchers!
P: I will show later on that these problems have never been solved but only swept under the carpet. For the next few minutes I want to stay on the abstract level, as you said, because statistical principles are abstract and there are many other theoretical pitfalls in treating social sciences as if these were a branch of mechanics.

DO SURVEYS AND POLLS USE REPRESENTATIVE SAMPLES?

P: We previously made the important distinction between a representative sample and a sample. Let us go back to the cook, the buyer and the teacher.

They focus on a part to infer information about the whole because they think that the part (the sample) is representative of the whole (the population). While one macaroni drawn from the boiling pot may look like the others, answering some questions correctly does not necessarily mean the student knows the whole syllabus. As a matter of fact it is crucial to differentiate between a representative sample (a sub-set that is a miniature of the population) and a sample (a sub-set of cases the researcher focuses on during any research). While one macaroni is a representative sample, the student’s answers might be only a sample.

Strictly speaking, using representative samples is plausible if there is no doubt that they mirror the characteristics of the population.

S3: Let us say that researchers aim at obtaining generalizable findings: how can they be sure that the sample they are focusing on is also representative?
P: There are two different ways (or criteria) to check this: inductive and deductive. But … before describing those I need to make a digression …
S1: Oh boy! How long it is taking to see the end of this tunnel!
P: Hey hey hey … It's not my fault if in the past methodologists have complicated things instead of solving them.

My digression relates to the concept of probability. In many textbooks of methodology it has been written that for a sample to be representative it needs to be based on probability methods. On this basis they have disqualified qualitative research.

Usually such textbooks distinguish between probability samples (simple random, systematic, proportional stratified, non-proportional stratified, multistage cluster) and non-probability samples (haphazard or convenience, purposive sampling, snowball) used by qualitative research. As for the latter, Bailey (1978: 115) says that the evident drawback of non-probability sampling is that, as the researcher does not know each subject's probability of being selected, usually she or he cannot state that his or her sample is representative of the population. This reduces the possibility of generalizing the findings beyond the sample in use; non-probability sampling can only be adequate if the researcher does not aim at generalizing his or her findings beyond the sample.

S3: It seems to me quite reasonable what Bailey said.

P: Yes, it could be … but only if it is possible in practice to work also with a probability sample. Let's see an example. You did this calculation several times in maths or statistics classes: when rolling a dice, what is the probability of the number two appearing in one throw only?
S2: One in six.
P: Right! To make this calculation you need to know in advance the whole range of alternatives, that is, the population of the number of the die. In social science it is the same. Firstly, in a simple random sample, to know the probability if each person really has the same probability of being included in the sample the
researcher has to know this probability (no matter whether each individual will be chosen or discarded). Secondly, to make sure someone is not missed during the selection, the researcher needs a complete list of the population, that is, to know exactly the number of cases within the population.

The first drawback: the list of population

According to statistical criteria, we now realize we cannot go on without the complete population list. It means that in order to know if the sample actually mirrors the characteristics of the population, ‘you need to know empirically (for these characteristics) both the sample and the population’ (Marradi, 1989: 53). As Bailey (1978: 108–14) says, random selection is possible if we have accurate information about the distribution of the statuses among the characteristics of the population, which we try to re-create proportionally in the sample using probability criteria.

As a matter of fact, finding a complete population list is often difficult and sometimes impossible because the list is not available, so that simple random sampling is seldom applied in surveys. At registry offices, at town halls or in electoral registers, complete lists of the population are available, but they record only a small number of characteristics, mainly socio-demographic. For most of the sub-sets of population or the characteristics sociology is interested in (emotions, attitudes or behaviours), these lists are not available. As Corbetta points out: ‘how can we obtain a random sample of unemployed people…if we do not have the whole list of unemployed people?’ (1999: 332). Many unemployed people register for work at employment agencies, yet not all of them have registered. Thus:

- a population list is not available for most of studies of special groups of people. Let us think about studies of blue collar workers, the unemployed, people who work at home, artists, immigrants, housewives, pensioners, sports team supporters, churchgoers, members of political movements, volunteers who do social work, elderly people who live alone, cohabitants and so on. (Corbetta, 1999: 333)

TROUBLE WITH SYSTEMATIC SAMPLING

Survey researchers say that the alternative to simple random sampling is systematic sampling.

It consists of choosing an individual (even without a population list) at pre-set $k$ intervals, obtaining random samples. There is only one condition: the whole population must be gathered in one place only, so that the researcher can choose at regular intervals the individuals who will be included in the sample. This procedure is in widespread use in areas such as the quality control of handmade products, exit polls, market research, and surveys of visitors to museums, cinemas and theatres. Even this is quite difficult as in social sciences it is not always possible to gather all the individuals with the characteristic (or characteristics) observed in one place to select them. Therefore, when there is no population list (or an equivalent that allows the researcher to reach the whole population theoretically), we have to give up using probability sampling procedures. As a matter of fact, each individual of the population cannot have the same probability of being selected.

The significance of this argument is even greater when we think of other more dynamic units: as we do not know how characteristics concerning emotions, attitudes, opinions and behaviour are distributed in the population, aiming at statistical representativeness of samples is technically groundless. Is there a population list on authoritative or unselfish behaviour?

The argument in favour of the implementation of probability sampling in social sciences is theoretically weak because it works only when the population is known. But what do we do when the population is infinite or unknown?

Three problems in statistical sampling theory

Statisticians state that probability sampling is applied only when the size of the population is unknown. The method helps to find quantitative estimates (parameters) that are unknown in the population and to determine the sample size, even though we do not know the magnitude of the population. It estimates the variance of a variable (for example churchgoers) in the population. This method comes from classical statistics and is based on sampling theory, involving the theorem of the central limit. However, in social sciences the application of this method involves problems.

SAMPLE SIZE ASSOCIATED WITH THE NUMBER OF VARIABLES

First, such a sample size is valid only when doing a univariate analysis (an analysis focused on one variable only). If we want to carry out a bivariate (two variables) or multivariate (three or more variables) analysis, to determine
the sample size we should make a calculation that takes account of each variable. The sample is therefore bound to increase, otherwise when analysing the data some cells might lack cases. However, according to Sudman (1983: 157), this occurs often: while it is rare that published studies have an insufficient sample if considered in their totality, in multivariate analysis samples have to be subdivided and in most of such subdivisions samples are inadequate (see also Capecchi, 1972: 50–1).

**The problem of the non-transferability of the logic of representativeness** Secondly, following the previous argument, representativeness has to be achieved on each variable of the research. Capecchi (1972: 50–1) exemplified clearly that in an interview with 80 questions (consequently with 80 variables), representativeness has to be achieved for each of them. However, if we follow the statistical method recommended in sampling theory we would need an enormous sample. But nobody does it. What survey researchers really do (in order to construct a sample that claims to be a miniature of the population) is to estimate the variance of a few variables only, usually socio-demographic ones (such as gender, age, education, political vote, residency and a few others) of which we know truly the population parameters from previous censuses. But what about the remaining 70 or so variables regarding attitudes, behaviours, feelings where we do not know their distribution in the population? As stated by Marradi (1989: 60), representativeness cannot be transferred from one variable to another.

Let us look at a trivial example: if survey researchers aim to study sexual behaviour (a variable whose distribution in the population is unknown), they usually construct a (claimed) representative sample based upon the ‘gender’ variable, that is, they choose some men and some women. They calculate the variance of sexual behaviour by looking at the two values (male and female) of the gender variable. But this is not **behaviour**; it is gender! In other words, we cannot study behaviour indirectly through a representative sample constructed on the basis of the gender variable. Indeed, the category ‘sexual behaviour’ has at least six sub-categories: heterosexual, homosexual, bisexual, abstinence, mostly heterosexual and mostly homosexual. The idea according to which a sample that is representative of some socio-demographic characteristics is automatically representative of psychological features (or behaviours or opinions) is highly problematic.

**Society is not random** Another assumption of statistics is that cases are distributed at random in the population. However, as Gilli (1971: 230ff.) reminds us, society is not random. There is no evidence that the sampling assumptions underlying natural sciences (i.e. that cases are interchangeable because they are equal and distributed at random in the population) works well in social sciences. On the contrary, in society almost nothing is at random: for example, there are social inequalities affecting people’s position in the population.

Probability sampling equalizes the chance of every case, including the odd ones, turning up. This seems, sociologically speaking, nonsense because not every person has the same relevance in society. We need a probability sample if we study a topic such as the political vote, where everybody counts: one head, one vote. However, when we study political behaviours not everybody counts in the same way: leaders, some politicians and their advisers, lobbies, minorities, financial supporters count more than voters. If we study black minorities, we do not consider Martin Luther King or Malcolm X to be on the same level as others; however, random sampling does this. More than thirty years ago Capecchi said: ‘we can state without any doubt that no national or regional sociological research based on interviews has ever used a representative sample that follows the rigorous consequences of statistical logic’ (1972: 53).

**The phenomenon of non-response**

Following the logic of statistical theory it would seem that, at least, polls on a few socio-demographic variables might be done with representative samples. If this is true in theory, practice shows that another drawback always lies in wait: non-response. As Marradi points out:

> the concept of random selection is theoretically very simple and, thanks to the ideal-typical image of the box, quite clear to public opinion. However this clarity is misleading…human beings differ from balls in the box on two features: they are not at the researcher’s hand… and they are free to decide not to answer. (1989: 78)

We have to take into account the gap (which varies according to the research project) between the initial sample (all the individuals about whom we want to collect information) and the final sample (the cases we managed to get information on): ‘the two sets may correspond but usually a part of objects of the first sample is not gathered’ (Gasperoni and Marradi, 1996: 628).
The phenomenon of non-response is composed of three different aspects:

1. Lack of contact with the person who had been selected (because she or he has moved, is unknown at that address, not there at the moment, ill, a prisoner, dead, or for other reasons).
2. Refusal to be interviewed.
3. Refusal to fill in the whole questionnaire if there are questions considered too sensitive by the respondent. In Italy these missing answers may vary from 5 per cent up to 60 per cent, especially on questions concerning income (Gobo, 1997: 177).

Usually there is about a 30 per cent non-response with peaks of 50–60 per cent. This would not be a serious problem if the 70 per cent taking part in the poll were similar to the people who refused; in this case the representativeness of the sample would not be jeopardized. But in practice quite the reverse often occurs: those who refuse to take part in the poll are not a random sub-set of the sample (Marradi, 1989: 73–6). As Kish (1965: 558) states, replacing non-responses is often a mistake because substitutes are like those who accepted to be interviewed rather than those who refused to answer. Non-responses may have systematically different statistical distributions compared to interviewed people concerning their main socio-demographic characteristics (Castellano and Herzel, 1971, 302; Marradi, 1989; Brehm, 1993: 17). Non-response makes non-random a sample chosen at-random. This brings about serious distortions of the univariate characteristic values such as mean and variance and of bivariate and multivariate coefficients.

However, survey researchers have found a remedy: weighting. As a matter of fact, in order to obtain a representative sample, some criteria have been developed to weight the answers of the interviewed people who, on some socio-demographic characteristics, are similar to those who have not been reached by interviewers or have refused the interview. However, this process seems too artificial and quite arbitrary (Pitrone, 1984: 149–50; Marradi, 1989: 68–78). Indeed, statisticians avoid it and suggest doing the analysis on actually collected data only.

If according to survey methodologists Converse and Schuman (1974: 40) the 20 per cent is a reasonable amount of missing data and does not jeopardize the representativeness of the sample, it has to be stressed that we are far from working on representative samples. As survey methodologists Groves and Lyberg (1988: 191) pointed out, non-response error threatens the survey’s unique characteristic compared to other research methods: the statistical inference from sample to population. If the sample strays from the probability model, nothing can be said about its general representativeness, that is about the fact that it truly reproduces all the characteristics of the population.

**REPRESENTATIVENESS IN QUALITATIVE RESEARCH**

S3: This detailed argument leads to a sort of nihilism because it seems impossible to do research with representative samples …

P: Don’t be too pessimistic. The aim was only to show that the criticisms of qualitative research are exaggerated because in most of their studies not even probability samples are what they are claimed to be. In addition, I rejected the forced choice, as some scholars portray it, between an (approximately) random sample and a totally subjective one, or between a partial probability sample and a sample whose representativeness we cannot be sure of. This false dilemma does not take into account an alternative path, which can be developed by thinking radically differently about the problem of representativeness in social science …

S2: In which way?

P: First, admitting that the most important studies in qualitative research, which produced significant theories, were based on non-probability samples: Gouldner (1954) observed one small gypsum extraction and refining factory located close to his university, so this was a haphazard sample; Dalton (1959) did a study at Milo and Fruhling, two companies in a highly industrialized area in the US; Becker (1951) studied many dance musicians; anthropologist De Martino (1961) observed 21 people suffering from tarantism disease; ethnomethodologists Sudnow (1967) and Cicourel (1968) observed two hospitals and two police districts respectively; Goffman (1963) analysed various interactions between various people.

S3: Does this mean that thinking about issues of sampling, representativeness and generalizability is a waste of time?
P: That is a good point! Many qualitative researchers think as you. However, defining the sampling unit clearly (which comes before choosing the cases and thus picking the sample) is important in order to avoid messy and empirically shallow research. During their analysis of some Finnish studies of ‘artists’, Mitchell and Karttunen (1991) noticed different findings according to the definition of artist employed by the researchers. Some studies included in the category ‘artist’ only those who defined themselves as artists; in other cases, only those who created durable works of art were included; in other cases, only those who were considered artists by the whole community; and in yet other studies, those who were registered by artists’ associations. A comparison is therefore nonsense.

S1: So we need to sample with accuracy …

P: Yes, and fortunately the problem of representativeness is a constant and growing concern of many young qualitative researchers.

S4: But how do we know the extent to which our cases (the sample) are representative of all members of the population from which the cases are selected? S5: Can we generalize from cases-study findings to a population without following a purely statistical logic?

P: You have raised two important questions that until now have not been solved by qualitative methodologists. First of all, in order to produce persuasive answers, we need to keep these two questions separate. Putting them together belongs to an old way of framing the problem. Methodologists too often equate representativeness and generalizability, forgetting that the former characterizes samples only and the latter concerns findings. We shall see this clearly later.

**SAMPLING UNITS**

P: Let us start with sampling units and then look at their representativeness.

In sociology and political science the main trend is choosing clearly defined and easily detectable individual or collective units: persons, households, groups, associations, movements, parties, institutions, organizations, regions and states. The consistency of these collective subjects is not very real. In practice, members of these groups are interviewed individually: the head of the family, the human resources manager, the statistical department manager, and so on. This means that the sampling unit is different from the observational unit (i.e. the respondent). Only a focus group can (at least in part) preserve the integrity of the collective subject. On the other hand, as Galtung (1967: 37) stated, choosing individuals implies an atomistic view of society, whose structural elements are taken for granted or reckoned to be mirrored in the individual (which means neglecting the sociological tradition that gives priority to relations instead of individuals).

S1: What is the alternative?

P: Other, more dynamic units are wrongly neglected:

- beliefs, attitudes, stereotypes, opinions;
- emotions, motivations;
- behaviours, social relations, meetings, interactions, ceremonies, rituals, networks;
- cultural products (such as pictures, paintings, movies, theatre plays, television programmes);
- rules and social conventions;
- documents and texts (historical, literary, journalistic);
- situations and events (wars, elections).

The researcher should focus her or his investigation on these kinds of units, not only because social processes are more easily detectable and observable, but also because these units allow a more direct and deeper analysis of the observed characteristics. When Corsaro, for instance, tried to analyse systematically children’s behaviour in kindergartens, the unit he chose was ‘the interactive episode’. He defined this as
sequences of behavior which begin with two actors in an ecological area and their try/tries to reach a shared meaning of an ongoing or starting activity. The episodes end with a physical movement of the actors who leave the area where the behavior had started' (Corsaro, 1985: 24), disappearing from the stage. Corsaro emphasizes that the most important thing is not his definition (which may be questionable) but the need to create and employ explicitly units based on the researcher’s theoretical and practical aims.

SOCIAL VERSUS STATISTICAL POPULATIONS

P: After having chosen the sampling unit we turn to the representativeness of samples. In order to evaluate this, it is crucial to distinguish between logical versus social universes.

S1: What is this?
P: The former belongs to statistics, the latter to social sciences. The consequence is that the former aims at statistical significance, the latter aims at the social significance or the sociological relevance of the population. As Goode and Hatt (1952: 339) pointed out, if in a study of marriage you extract a sample from the ‘universe of marriages that happened in New York in the last ten years’, from a statistical point of view this procedure is correct because for statistics the actual reality of such a universe is irrelevant. However, from a sociological view this is nonsense because such a universe simply doesn’t exist, it doesn’t have any social substance, it isn’t a collective subject as a social movement or an organization. It is a researcher’s invention only.

S2: So how can we proceed?
P: First, you look at the variance of the phenomenon under the study. If it is high, you need many cases in order to include in your sample each category or class of your phenomenon. If its variance is low you need few cases; sometimes one case could be enough …

S1: I can scarcely believe my ears …
P: The conversation analyst Harvey Sacks (1992, vol. 1: 485, quoted in Silverman, 2000: 109) reminds us of anthropologist and linguist Benjamin Lee Whorf, who managed to reconstruct Navajo grammar by interviewing extensively only one native Indian speaker. Usually grammars have low variance. It would probably have been different if he wanted to study how Navajos bring up children or how they have fun. In his famous study, anthropologist Clifford Geertz (1972) attended 57 cockfights and ethnologist Ernesto De Martino (1961) observed 21 people (socially labelled as) suffering from tarantism disease. They selected a sample with so many or so few cases probably because this reflected the variance. This is the only very important rule for selecting the sample, and it is a pity that many qualitative and quantitative researchers seem to forget it.

S2: Can you give us another example?
P: Let us take my recent study of call centres. In Italy it has been estimated (this means that nobody knows the real number) that there were 1020 call centres in 2000. I was interested in ‘customer relationship management’ (the sampling unit) done by call takers in order to solve clients’ requests. What is the variance of this sampling unit? If I followed the traditional way of sampling, focusing on the call centre as organization, I would have selected three types of call centres: private or marketing oriented; public, as for example medical emergency dispatch centres; and non-profit. However, if I am interested in ‘customer relationship management’ practices, to make a representative sample one has to consider their variance in the population instead of the variance of call centres. In other words, the reference is not the variance of the call centre but the variance of ‘customer relationship management’ practices within call centres. Reviewing literature, talking with experts and doing ethnographic research, I found that there are mainly four types of such practices based on the nature and structure of tasks: counselling, marketing, interviewing and
advertising. In addition, looking closely at each of them, while the three latter seem to have low variance, counselling has an higher variance due to the nature of counselling itself, which ranges over different kinds of assistance: technical, medical, psychotherapeutic, and so on.

S1: This is quite complicated.
P: Yes, but it is a good way to obtain representativeness. I did not have enough resources to select four samples; I could study only one type of practice using this approach. This meant that when attempting to generalize the findings, one will be conscious about the extent to which the sample is representative. In addition, this is a way of getting away from many sociological stereotypes. We sample leftists versus rightists to know what they think about immigrants. Why not do the reverse? Sample attitudes about immigrants and then look at where leftists and rightists are located. This procedure avoids the effects of conventional theories.

S3: Maybe I am too stupid with statistics but I still cannot believe that a few cases, even selected with the accuracy you describe, can make a representative sample …
P: There is some evidence for it. At the end of the 1980s I did a study of interviewing, particularly on rituals and rhetorical strategies used by one interviewer in his telephone contacts in order to obtain the consent for the interview, with ten Italian respondents (Gobo, 1990, 2001). I tape-recorded the verbal exchanges and analysed them. Some years later I read a similar study conducted in the US by Maynard and Schaeffer (1999) and I found in it more or less the same patterns. This happened not because I was lucky but because interviewers receive a similar training and use in their work similar cognitive artefacts (scripts and questionnaires).

SAMPLING, REPRESENTATIVENESS AND GENERALIZABILITY

P: As I have already said, the precision of a sample, its being an accurate miniature of the universe, is better when the population from which it is drawn is homogeneous, and it is lower when it is more heterogeneous. This is the key issue for legitimating qualitative research sampling.

We already know that an inductive procedure used to check the correspondence between sample and population is almost impossible for practical reasons. This implies thinking through whether the social phenomenon under study is pervasive. In other words, we may expect there are not significant differences between the population and the sample. For this reason a few cases, mirroring a pervasive phenomenon in society, may be enough if its population is quite homogeneous. Following this reasoning is perfectly acceptable in many qualitative studies. For example, anthropologist Vincent Crapanzano (1980) studied Moroccan social relations through the experience of Tuhami, a furnace-workman. Anthropologist Marcel Griaule (1948) reconstructed the cosmology of the Dogon, a population from Mali, questioning a small group of informants only; Bourdieu’s book (1993) on professions is based on 50 interviews with policewomen, temporary workers, attorneys, blue-collar workers, civil servants and the unemployed. In general, biographical method relies on this methodological assumption (Javeau, 1987: 176ff.) Referring to conversation analysis, Harvey Sacks (1984: 22, quoted in Silverman, 2000: 109) has provocatively argued: 'tap into whomsoever, wheresoever and we get much the same things'.

STUDYING SIMILARITY OR DIFFERENCE?

Because probability sampling has many theoretical problems and is also almost impossible to apply in social research, the only practical and sensible way is to work with non-probability samples. We have two broad criteria in building a non-probability sample depending on a
QUALITY AND CREDIBILITY

researcher’s interest: looking at similarities or differences among cases. The former criterion maximizes the probability of extracting odd cases, but focuses on their similarities. The latter selects only odd (deviant) cases.

In the 1700s in biology there was a great controversy about this, associated with the two fundamental aims of natural history: disposition and denomination. This controversy has been marvellously described by Foucault (1966: 150–79). To accomplish these two tasks, naturalists of that time used one of two different and opposite techniques: the System (Linnaeus) and the Method (Buffon, Adanson, Bonnet). On the one hand, Carolus Linnaeus’s taxonomic technique was concerned with finding similarities among animals and plants; on the other hand Buffon, criticizing Linnaeus, maintained that our general ideas are relative to a continuous scale of objects of which we can clearly perceive only the middle rungs and the extremities increasingly flee from and escape our considerations. The more we increase the number of divisions in the productions of nature, the closer we shall approach to the true, since nothing really exists in nature except individuals, and since genera, orders, and classes exist only in our imagination. (quoted in Foucault, 1966, 146–7)

Bonnet seems no less resolute:

There are no leaps in nature: everything in it is graduated, shaded. If there were an empty space between any two beings, what reason would there be for proceeding from the one to another? Above and below every human being we find other beings that are close to him by some characters and differentiated from him by others.

It is therefore always possible to discover ‘intermediate organisms’, such as the polyp between the animal and the vegetable, the flying squirrel between the bird and the quadruped, the monkey between the quadruped and the man. Consequently, our divisions into species and classes ‘are purely nominal’; they represent no more than ‘means relative to our needs and to the limitations of our knowledge’ (1764, quoted in Foucault, 1966: 147).

This historical digression is epistemologically important, as in social research these two methods have been largely used. Most studies and associated samples follow the former criterion in search of dominant characteristics of pervasive phenomena: biographical method, conversation and discourse analysis, organizational ethnography, and so on. However, there are some important studies that have focused on deviant cases in order to understand standard behaviour [e.g. Goffman (1961) on ceremonies and rituals in psychiatric clinics; Cicourel and Boege (1972) on deaf children’s interpersonal communications; Garfinkel (1967) on achievement of sex status in an ‘intersexed’ person; Pollner and McDonald Wikler (1985) on interactions in a family with a mentally retarded child] or to explore subculture or emergent or avant-garde phenomena that could become dominant or significant in the future [e.g. Festinger, et al. (1956) on Jehovah’s Witnesses; Becker (1963) on marijuana smokers; Hebdige (1979) on style groups such as mods, punks and skinheads; Fielding (1981) on right-wing political movements]. The deviant case can be used also as a proof of refutability and falsification of a well-known and received theory, as Rosenhan (1973) did against the psychiatric theory that locates mental disease in a person’s head, or De Martino (1961) did against the medical theory that considered the tarantism syndrome as being caused by the sting of a spider. The latter procedure (widespread in palaeontology) does not consider to what extent the study’s finding is distributed in the population, but only notes that the phenomenon exists and must therefore be reviewed and understood theoretically.

THEORETICAL SAMPLING

Theoretical sampling concerns both. The expression was coined by Glaser and Strauss (1967), and refined by Schatzman and Strauss (1973), Strauss (1987) and Strauss and Corbin (1990). In the following refinements the notion has been expanded, producing different versions and loosening it from its initial meaning. Theoretical sampling means ‘sampling on the basis of concepts that have proven theoretical relevance to the evolving theory. [Relevance means] that concepts are deemed to be significant because they are repeatedly present or notably absent when comparing incident after incident, and are of sufficient importance to be given the status of categories’ (Strauss and Corbin, 1990: 176). In addition, as Mason writes, ‘theoretical sampling is concerned with constructing a sample which is meaningful theoretically because it builds in certain characteristics or criteria which help to develop and test your theory and explanation’ (1996: 94).

Theoretical sampling was and still is a tremendously important way of reasoning in order to select cases for your sample. However, at least in its first versions, it does not explicitly and clearly take account of the problem of representativeness, but simply avoids it. In such a useful concept there is nevertheless no suggestion on how
to consider the representativeness of selected categories. The selection of cases is driven only by their relevance for the phenomena under study and not by the need for representativeness. How, then, should we address Denzin’s concern that ‘the researcher has to show that these units, compared to all other similar phenomena, are representative’ (1971: 232), recently pointed out also by Becker (1998: 67)? For past generations of qualitative researchers the notion of theoretical sampling (beyond being extremely useful) was also pedagogically harmful because (wrongly) it was received as a slogan and an implicit invitation not to care about representativeness.

**PRACTICAL ADVICE ON SAMPLING**

What are the criteria for selecting cases in order to construct a representative sample in social research? Some criteria have been already described, but I shall summarize them together here even though there is not a precise logical itinerary because methodological principles and rules do not have to stand on their own, having only a weak relation to practice. On the contrary, it is necessary to approach the whole question of sampling sequentially and it would be misleading to plan the whole strategy before starting. In order to obtain representativeness, the sampling plan needs to be set in dialogue with field incidents, contingencies and discoveries. Here is an example of changing or adding to the sampling plan on the basis of something the researcher learns in the field:

Blanche Geer and I were studying college students. At a certain point, we became interested in student ‘leaders’, students who were heads of major organizations at the university (there were several hundred of them). We wanted to know how they became leaders and how they exercised their powers. So we made a list of the major organizations (which we could do because we had been there for a year and knew what those were, which we would not have known when we began) and interviewed twenty each of men and women student leaders. And got a great result – it turned out that the men got their positions through enterprise and hustling, while the women were typically appointed by someone from the university! (Becker, 13 July 2002, personal communication)

Afterwards consistency must be reached in the sampling reasoning, not through just applying procedural steps. The reasoning could be as follows:

1. The researcher usually starts from his or her research questions. Dalton’s ones were:

   Why did griever managers and managers form cross-cliques? Why were staff personnel ambivalent toward line officers? Why was there disruptive conflict between Maintenance and Operation? If people were awarded posts because of specific fitness, why the disparity between their given and exercised influence? Why among executives on the same formal level, were some distressed and some not? And why were there such sharp differences in viewpoint and moral concern about given events? What was the meaning of double talk about success as dependent on knowing people rather than on possessing administrative skills? Why and how were ‘control’ staffs and official guardians variously compromised? What was behind the contradictory policy and practices associated with the use of company materials and services? Thus the guiding question embracing all others was: what orders the schism and ties between official and unofficial action? (Dalton, 1959: 274)

Research questions contain concepts and categories (behaviours, attitudes and so on) the researcher is interested in studying.

2. Then she or he carries out primary (or ‘provisional’ and ‘open’, following Strauss and Corbin, 1990: 193) sampling in order to collect cases in accordance with concepts.

3. Because not every concept can be directly studied, in crafting the provisional sample she or he reasons about some aspects:

   a. specificity (focusing on precise social activities, with clear contours as a ritual or a ceremony);
   b. degree of openness of the field (open or closed places);
   c. intrusiveness (the will to reduce the researcher’s visibility);
   d. institutional accessibility (free-entry versus limited-entry situations within the organizations);
   e. significance (frequent and high organizational relevance of social activities).

4. It is recommended to sample types of actions or events, ‘incidents’ and not persons per se!' (Strauss and Corbin, 1990: 177, bold in the original text), in contrast with the common habit of sampling bodies and of seeking information from these bodies about behaviour and events that are never observed directly (Cicourel, 1996). This important recommendation has two reasons: first, we do not want to replicate the survey sampling mistake about the transferability of ideas about representativeness second, the same person can engage in several activities. For example, Dalton (1959), exploring power struggles in companies, found five ‘types of cliques’: vertical (symbiotic and parasitic), horizontal (defensive and aggressive) and random. If we
sample individuals we will notice that they stay in more than one clique according to the situation, intentions and so on. If we look at activities, everything is simpler.

(5) Until now in social research four main types of sampling have been invented: purposive, quota, emblematic and snowball. We will see more of these later. In selecting cases you need to pay attention to the variance of concept in order to include in the sample different voices or cases.

(6) During the study the researcher will certainly refine his or her ideas, categories and concepts, or come up with new ones. The important thing is to make connections among them, thus formulating working hypotheses. Even if some qualitative researchers believe that field research is carried out without hypotheses, making hypotheses is as much an everyday life activity as comparing, sampling, making inference, finding causes, and so on. A hypothesis is a conjectural statement or assertion about the relation between two or more properties. Not every hypothesis is testable; indeed the most interesting ones often aren’t. However, if we want to persuade the reader we need to formulate them in a testable way.

(7) When the researcher has set up some hypotheses, then she or he restarts sampling in order to collect systematically cases that relate each hypothesis, trying to make his or her analysis consistent. Strauss and Corbin call this second sampling ‘relational and variational: it is associated with axial coding. It aims to maximize the finding of differences at the dimensional level’ (1990: 176). The authors frame the research process as a funnel-shape: through three better and better focused steps (open, axial and selective) the researcher clarifies his or her statements because ‘consistency here means gathering data systematically on each category’ (Strauss and Corbin, 1990: 178). When the researcher finds an interesting aspect she or he would always have to control for it if this finding happens in other samples (Perrone, 1977: 27).

(8) To guarantee representativeness she or he needs to collect cases on recurrent behaviour at different moments. As the researcher cannot observe the population 24 hours a day, she or he has to decide at what time and where she or he will observe the population (Schatzman and Strauss, 1973: 39–41; Corsaro, 1985: 28–32). Social practices always recur in certain places and at certain times of the day. If the researcher knows all the different rituals of the organization observed she or he can draw a representative sample. We do not aim to know the distribution of such behaviours (how many times), a purpose that surveys cannot succeed in, as we saw, but only if they are quite recurrent and significant in the organization under study. In addition, ‘our concern is with representativeness of concepts’ (Strauss and Corbin, 1990: 190).

(9) The researcher can sample new incidents or she or he can review incidents already collected: ‘Theoretical sampling is cumulative. This is because concepts and their relationships also accumulate through the interplay of data collection and analysis…until theoretical saturation of each category is reached’ (Strauss and Corbin, 1990: 178, 188).

(10) This interplay between sampling and testing hypotheses is needed because

(a) representative samples are not predicted in advance but found, constructed and discovered gradually in the field;
(b) it reflects the researcher’s experience, previous studies and literature regarding the topic. In other words, the researcher will know the variance of a phenomenon cumulatively, study by study;
(c) representative samples are used to justify the researcher’s statements.

STRATEGIES OF SAMPLING

There are four important sampling procedures in qualitative research.

Purposive sampling

Purposive sampling consists of detecting cases within extreme situations as for certain characteristics or cases within a wide range of situations in order to maximize variation, that is, to have all the possible situations. We can choose two different elementary schools where, thanks to the press, to previous studies or interviews or personal experiences, we know we can find two extreme situations: in the first school there are enormous integration difficulties among natives and immigrants, while in the latter there are virtually none. We can also pick three schools: the first with huge integration difficulties, the second with average difficulties, and the third with rare cases. In the 1930s and 1940s the American anthropologist W. Lloyd Warner (1898–1970) and his team of colleagues and students carried out some studies on various communities in the United States. When he had to choose the samples, he decided to select different communities whose social structure mirrored some important features of American society. Four communities...
Quota sampling

Quota sampling is employed for objects that contain a wide range of statuses. The population is divided up into as many sub-sets as the characteristics we want to observe and the proportion of each sub-set in the sample is the same as in the population. This sampling method is in widespread use in market research and polls and is usually associated with the use of questionnaires. Curiously, even though it is not a probability sample, it is largely employed in quantitative research because it helps obtain good findings while cutting down on costs. Even though statisticians do not consider it to be scientific, because cases are not selected at random but by a human being (always the bearer of some bias), in the last few decades quota samples have been shown to produce findings as precise (and sometimes even more precise) as probability samples. A good example of this sampling method (with non-proportional quotas) applied to ethnographic research is Jankowski’s (1991) study of criminal gangs. He observed for ten years 37 different gangs selected according to their ethnicity, size and members’ age in Los Angeles, New York and Boston.

Another example is Gouldner’s (1954) research. At that time the gypsum company employed approximately 225 people. In his methodological appendix he reported that his team did 174 interviews, that is, almost all of the population (precisely 77 per cent of it): 132 of 174 interviews were done with a ‘representative sample’ of blue-collar workers in the company. Gouldner used quota sampling stratified by age, rank and tasks. Then he did another representative sample of 92 blue-collar workers who were given a questionnaire.

The emblematic case

The emblematic case may have up to three features: average (the typical provincial hospital, the organization of a typical mountain-village town hall), excellence (a well-known car-manufacturing firm) and emerging (or avant-garde, such as recent juvenile phenomena). Middletown, the two famous studies carried out by Robert S. and Helen M. Lynd (1929, 1937), is a pioneer attempt to study a typical American community applying social anthropology methods. The households chosen were ‘typically’ American and not statistically representative ones. Other examples are two studies that Gouldner (1954) carried out between 1948 and 1951 in a small gypsum extraction and refining factory, Dalton’s research (1959) at Milo and Fruhling, two companies in a highly industrialized area in the US, and, as a covert observer, Kanter’s (1977) observations of a company with high technological density for five years. As particular elements are more easily detectable through comparison, during each study it is advisable to observe at least two cases.

Snowball sampling

Snowball sampling means picking some subjects who feature the necessary characteristics and, through their recommendations, finding other subjects with the same characteristics. The research of Whyte (1943) may be an example: thanks to Doc, a young unemployed man who attended the social service, the author gradually managed to get in touch with the network of his acquaintances, the people and the groups belonging to the ‘street corner society’. If snowball sampling respects some criteria, it may also be included in probability sampling (see TenHouten et al., 1971).

As we can see, most samples used in qualitative research are still linked to a traditional style of constructing samples, i.e., sampling individuals instead of concepts.

REPRESENTATIVENESS AND GENERALIZABILITY

P: We have arrived at the last issue: the relationship between representativeness and generalizability. Even here there are some commonplaces left to be deconstructed:

1 only findings from a representative sample are automatically generalizable to the population;
2 (therefore) representativeness leads to generalizability;
3 if a study is not carried out on a representative sample, its findings are not generalizable;
4 findings of qualitative studies are not generalizable.

In the methodological literature, generalizability is usually considered a direct and automatic consequence of representativeness.

S1: Why is it not?
P: Let us take another example. Why, sometimes, do pre-electoral polls fail? As their samples are statistically representative, generalization should be automatic and predict correctly the (probability of) results.

S4: If polls fail it means that their samples weren’t representative!
P: This is a trick. Do you mean that representativeness is evaluated only after the result? It seems to me quite tautological, a kind of statistical Darwinism whereby success reframes the past. What about two different polling agencies working with the same statistically representative sample and producing different results?

Too often we forget that between the representativeness of a sample and the generalizability of its findings there is a number of activities that depend on at least eight different domains: the trustworthiness of operational definition, the trustworthiness of operationalization, the reliability of method, the suitability of conceptualization, the researcher’s accuracy, his or her degree of success with field relations (people can lie), and data and interpretation validity (data analysis error – see Groves, 1989). These aspects and their associated mistakes (called measurement error) may jeopardize the equivalence between representativeness and generalizability. And it happens quite frequently in any research study.

S4: So we are not certain …
P: The only way is to check. However, in social research the generalizability of findings is quite problematic because it is objectively very difficult to check to see if, at the end of the study, the findings from a sample actually mirror the rest of the population. To do such a thing we would have to check the population anew by a census. This seldom occurs; an example is given by polls on voting intentions: the research is carried out on a representative sample and then, thanks to the findings of the ballot (a census of votes), we can check the generalizability of the data drawn from the sample. Nevertheless, as we saw before, success or failure of forecasts could not depend only on the representativeness of the sample. There are other factors, which have nothing to do with the method, that might have an impact on the success of the forecasts: chance, luck, artificial manipulations during data processing (weights), …

S1: Lack of new political parties.
P: Right! Indeed, pre-electoral polls usually show a difference when a new party enters the electoral competition for the first time. This failure happens because sampling is always based on the previous census (the past election), not on the present situation. So forecasts may be unsuccessful even though statistically representative samples are employed. This means that, if findings deriving from samples may be eliminated by the findings of the elections, it is always advisable to check the population: theoretically, without the confirmation given by the elections, all findings might be valid. Without this control we would never be able to know which different data published by polling agencies mirror the population. Unfortunately, there is no similar census on most of the topics sociological research is interested in.

GENERALIZABILITY IN STATISTICS

P: Let us see briefly how generalization in statistics works. This is called inference. It is an academic issue since we have already seen that in most surveys there aren’t statistically representative samples. To estimate the probability that the finding (e.g. an existing relation between variable A and variable B) drawn from your sample is also in the population, you are helped by some statistical tests. There are many statistical tests for
controlling hypotheses. The best known is chi-square. However, if you do a study of a whole population (e.g. all of the 300 students in the Research Methods class instead of just some), you do not need such a test and an inspection cross-tab by cross-tab is needed. In other words, statistical tests of significance are used only when you want to infer from a sample to the whole population. In the past there was a long controversy about the appropriateness of applying statistical significance tests in social sciences (see Morrison and Henkel, 1971). The main criticisms, well known by survey methodologists, are:

1. the arbitrariness of the probability threshold of 0.05 below which the null hypothesis is rejected;
2. the associated crudeness of this dichotomous decision, without any gradation, in order to accept or refuse the probability of the null hypothesis being valid for the population;
3. the statistical test assumes a rigorously random sample, otherwise the procedure is nonsense. However, as we have already seen, this is quite difficult to obtain in social research;
4. the chi-square is heavily affected by sample size: it increases in significance when sample size increases! With very large samples even small trends will appear significant.

So statistical tests are of modest usefulness in social research. They can be helpfully used only as orienting criteria for the researcher’s decision-making.

S3: In addition, chi-square is a significance test only …

P: Right. It tells if a relation is likely to exist in the population only. It doesn’t say anything about the strength of this relationship. You need other tests, associated to the nature of your variables (nominal, ordinal, interval and so on), such as phi, Cramer’s V, Pearson’s C, and so on. However, they are rarely used. Generally speaking, all these issues have been neglected with the consequence that many survey researchers still venerate these tests and many field researchers are still afraid.

ARE QUALITATIVE RESEARCHERS’ FINDINGS NOT GENERALIZABLE?

S1: So methodologists who say that qualitative studies are not generalizable should be more careful because survey studies also have many troubles with generalization?

P: You got the point: statistical inference in social studies is quite problematic. Unfortunately this hasn’t been pointed out by most qualitative methodologists, who (on this issue) accepted the received view. For example, Strauss and Corbin state:

in terms of making generalization to a larger population, we are not attempting to generalize as such but to specify … the condition under which our phenomena exist, the action/interaction that pertains to them, and the associated outcomes or consequences. This means that our theoretical formulation applies to these situation or circumstances but to no others. (1990: 191, bold in the original text)

S3: They refer to a kind of ‘internal generalization’ …

P: Yes. And a similar position is held by Lincoln and Guba (1985: 20–21), Hammersley (1992: 186ff.) and many others. The underlying idea is that there are two kinds of generalizations: a generalization about a specific group or population (applied in surveys and polls) and a generalization about the nature of a process (applied in field research). Consequently sampling requirements are completely different in the two cases. The latter is based on the notion of ‘theoretical sampling’. To make clearer the difference between these two kinds of generalization, some authors have called the second one ‘transferability’ (Guba, 1981; Lincoln and Guba, 1985: 77, 217), ‘naturalist generalization’ (Stake, 1983: 282), ‘analytical generalization’ (Yin, 1984: 31), ‘extrapolation’ (Alasuutari, 1995: 157) and ‘moderate generalization’ (Williams, 2002). These proposals are
politically wise because they try to give qualitative research scientific legitimation; however, they do not solve adequately the theoretical and practical problems involved in the notions of representativeness and generalizability.

We might ask: if a qualitative researcher’s findings are not generalizable, why does she or he carry out a study? What is the point, for instance, for an ethnographer spending months planning a study, weeks negotiating access to the field, months observing social actors’ behaviour, days organizing and analysing ethnographic notes and, finally, writing a research report if then she or he is told that the findings are only applicable to his or her community/organization/group (the sample) and not to many/all cases of the population? This is masochism!

Besides, this was not the way in which the studies done by Goffman, Gouldner, Whyte and so on were received. No one would remember them for the findings of one setting only.

S2: What is the trick then?
P: The logical mistake is in confusing the representativeness of the case with the representativeness of its characteristics observed by the ethnographer. Even though the case may be a firm, we should take into account the relationship among employees, the psychological and relational effects of a new work organization, the power relationships between chiefs and subordinates, and so on. So the sample is a compound of such characteristics, and is not just the firm. For this reason I consider it misleading to use the widespread expression ‘case study’ and also damaging to the image of qualitative research. As a matter of fact, to use Mason’s words ‘you may have sampled people but what you really want to compare are their experiences’ (1996: 96). It is what Gouldner (1954) did: he sampled blue-collar workers to discover their values, motivations and so on.

And Becker, writing about the cognitive process of generalization, stated: ‘in every city there is a body of social practices – forms of marriage, or work, or habitation – which don’t change much, even though the people who perform them are continually replaced through the ordinary demographic process of birth, death, immigration, and emigration’ (2000: 6).

S4: Yes, but we do not know the distribution in the population.
P: But nobody can know this; neither can statistics help you. You can quantify or correctly estimate the distribution only for characteristics already counted in a census. How many of them are there in social research? So the only method (as far as I know) is by evaluating the social pervasiveness of your concepts under study.

As Alberoni and his colleagues wrote in their Introduction about a study carried out on 108 political activists of the Italian Communist Party and the Christian Democratic Party:

if we want to know, for instance, how many activists there are of both parties in the whole country coming from families of the Catholic or Communist tradition, [this] study is useless; quite the reverse, if we want to show that family background is important in determining if a citizen will be an activist in the Communist rather than in the Christian Democratic party, this research can give the right answer. If we want to find out what are and what have been the percentages of the different ‘types’ of activists…in both parties, the study is useless, while if we want to show these types exist the study gives a certain answer. … The study does not aim at giving a quantitative objective description of Italian activism but can help in understanding some of its essential aspects, basic motivations, crucial experiences and typical situations which gave birth to Italian activism and help to keep it alive. (Alberoni et al., 1967: 13)

Alberoni was too afraid of the predictable quantitative criticism.

A NEW CONCEPT OF GENERALIZABILITY FOR QUALITATIVE RESEARCH?

In this chapter I have documented the need for a new, bottom-up, socially informed and practically driven theory of sampling, representativeness and generalization. This theory goes back
and remains faithful to the original central problem of representativeness as theorized by statisticians: the variance of the phenomenon under study. The variance is the only worry the researcher needs to take into consideration.

The concept of generalizability now introduced is based on the idea of social representativeness, which goes beyond the limits of statistical representativeness. The aim is to observe extensively the relations between variables, not only to assess (which is always a quite problematic task) the number of persons who feature one characteristic. Therefore, generalizability is mainly a practically and contingent outcome related to the variance of the research topic; in other words it is a function of the invariance (regularities) of the phenomenon, not a standard or automatic algorithm of a statistical rule.

This is a reason why the findings of the studies of Goffman on social embarrassment, deference and demeanour, Sacks and his colleagues on telephone conversation machinery, Whyte (1943) on social organization and leadership in a group, Gouldner (1954) on bureaucracy in medium enterprises, Cicourel (1968) on the description of a typical juvenile offender, van Dijk (1983) on the cognitive processes of racial prejudice, and Norman (1988) on accidents in the use of technology in everyday life, and so on, have always been considered generalizable.

In qualitative research, generalizability concerns general structures rather than single social practices, which are only an example of this structure. The ethnographer does not generalize one case or event that, as Max Weber pointed out, cannot recur but its main structural aspects that can be noticed in other cases or events of the same kind or class. For example, in the conclusions of his study of the relationship between a psychotherapist and a patient suffering from AIDS, Peräkylä says:

The results were not generalizable as descriptions of what other counselors or other professionals do with their clients; but they were generalizable as descriptions of what any counselor or other professional, with his or her clients, can do, given that he or she has the same array of interactional competencies as the participants of the AIDS counseling session have. (Peräkylä, 1997: 216, quoted in Silverman, 2000: 109)

Something similar Peräkylä, happens in film and radio production with noise sampling. The squeak of the door (which gives us the shivers when we watch a thriller or a horror) does not represent all squeaks of doors but we associate it with them. We do not think about the differences between that squeak and the one made by our front door; we notice the similarities only. They are two different ways of thinking, and most social sciences aim at finding such patterns.

While the verbal expressions of an interactive exchange may vary, exchange based on the question–answer pattern features a formal trans-institutional (even though not universal) structure. While maybe to lay a page of a newspaper on the floor and declare one’s sovereignty over it (Goffman, 1961) is a behaviour observed in one psychiatric clinic only, the need to have a private space and control of territory has been reported many times, though in different forms. However, the extension of this structure must be well organized. On the other hand, as Rositi states, we may reasonably doubt the generalizability of findings of ‘studies of 1000–2000 cases which claim to sample the whole population. We have to wonder if we should prefer such samples with such aims…. Studies with samples of 100–200 conversational interviews, structured to “describe” variables rather than a population, are definitely more suitable for a new model of studying society’ (1993: 198).

ACKNOWLEDGEMENT

I wish to thank Howard Becker, Aaron Cicourel, two anonymous referees and the editors for helpful comments and suggestions.

NOTES

1 Incidentally I emphasize that this remains a problem even for studies based on samples.
2 See http://www.census.gov/dmd/www/samfaq.htm
3 The author also writes: ‘the situation is slightly better where the unit is a group because an aggregate of individuals is usually institutionalized (schools, universities, companies, hospitals, commercial activities); however, the situation becomes more complicated for unofficial structures: private language or computer courses, sport, leisure and cultural associations, charities, amateur theatre groups, bowling clubs, pensioners’ recreation groups and so on’ (Corbetta, 1999: 333).
4 Even among those who seem to accept the legitimacy of inferential procedure, there are authors who have doubts about it, such as Isidor Chain (1963: 512), who states that we are never sure that findings on sample can be applied to population unless we also simultaneously do a census.
5 As Strauss and Corbin (1990: 176) explain: ‘open sampling is associated with open coding. Openness rather than specificity guides the sampling choices’. Open sampling can be done purposively (e.g. pp. 183–4) or systematically (e.g. p. 184) or occurs fortuitously (e.g. pp. 182–3). It includes on-site sampling.
6 Cf. Davis et al. (1941).
QUALITY AND CREDIBILITY

7 Cf. Drake and Cayton (1945).
8 Cf. Warner (1949).
9 As Barisione and Mannheimer write: 'The experience in public opinion surveys has showed that quota samples are on the average reliable; as a matter of fact their results do not differ from results obtained with random samples of the same size more than two random samples differ between them' (1999: 54). This is another case when experience and practice predominate over abstract statistical principles (1999: 54).
10 According to the authors the expression ‘transferability’ does not stand for a researcher’s inference process but a choice of the reader, who may transfer this information to other situations on the base of a solid thick description provided by the researcher (Lincoln and Guba, 1985: 362).

REFERENCES
