



Re-conceptualizing Generalization: Old Issues in a New Frame

Giampietro Gobo

INTRODUCTION

Even though qualitative methods are now recognized in the methodological literature, they are still regarded with skepticism by some methodologists, mainly those with statistical training. One reason for this skepticism concerns whether qualitative research results can be generalized, which is doubted not only because they are derived from only a few cases, but also because even where a larger number is studied these are generally selected without observing the rigorous criteria of statistical sampling theory. In this regard, the methodology textbooks still distinguish samples into two types: *probability* samples (simple random, systematic, proportional stratified, non-proportional stratified, multistage, cluster, area, and their various combinations), and *non-probability* ones (haphazard or convenience, quota, purposive, of the emblematic

case, snowball, telephone)¹. With regards the latter is stated:

the obvious disadvantage of nonprobability sampling is that, since the probability that a person will be chosen is not known, the investigator generally cannot claim that his or her sample is representative of the larger population. This greatly limits the investigator's ability to generalize his or her findings beyond the specific sample studied (...) A nonprobability sample may prove perfectly adequate if the researcher has no desire to generalize his or her findings beyond the sample (Bailey, 1978: 92).

This position again tends to relegate qualitative research to the marginal role of furnishing ancillary support for surveys, which is precisely as it was conceived by Barton and Lazarsfeld (1955) and the methodologists of their time.

The aim of this study is to show that this methodological denigration of qualitative research is overly severe and unjustified,

for three reasons. First, because the use of probability samples and statistical inference in social research often proves problematic. Second, because there are numerous disciplines, in both the social and human sciences, whose theories are based exclusively on research conducted on only a few cases. Third, because, *pace* the methodological orthodoxy, a significant part of sociological knowledge, is idiographic. My intention is therefore not to criticize sampling theory or its applications; rather, it is to remedy a situation where statistical inference is deemed the only acceptable method, and idiographic generalization as scientifically ill-founded. Finally qualitative researchers do not need to throw away the baby generalization with the bathwater of probability sampling, because we can have generalizations without probability.

THE PROBLEMATIC USE OF PROBABILITY SAMPLES IN SOCIAL RESEARCH

Several authors (among them Goode and Hatt, 1952; Chain, 1963; Galtung, 1967; Capecchi, 1972) have stressed that the application of statistical sampling theory in sociological contexts gives rise to various difficulties. This theory, in fact, requires the researcher to construct a probability sample (one, that is, where each subject's likelihood of being selected is known and also every item has an equal chance of being selected), and the cases must be selected in rigorously random manner. But these two requirements are not easy to satisfy in social research, because their fulfillment encounters a series of obstacles, not all of which can be overcome.

There is no space to describe in depth the problems and limits of statistical sampling theory (see Gobo, 2004). I will briefly examine three limits only:

- 1 *The difficulty of finding sampling frames* (lists of population) for certain population sub-sets, because these frames are often not available. How, for example, can a random sample of the unemployed be extracted if the whole list of unemployed people is not available beforehand? It is true that many unemployed people are enrolled at job placement offices, but it is equally true that not *all* unemployed people are so enrolled. Consequently, the majority of studies on particular segments of the population cannot make use of population lists: consider studies on blue-collar workers, the unemployed, home-workers, artists, immigrants, housewives, pensioners, football supporters, members of political movements, charity workers, elderly people living alone, and so on.
- 2 *The phenomenon of nonresponse*. The concept of random selection is theoretically very simple and, thanks to the ideal-typical image of the box, quite clear to the general public. This clarity is misleading, however, because human beings differ from balls in a ballot box in two respects: they are not immediately accessible to the researcher, and they are free to decide not to answer. In fact, account must be taken of 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 about which we have been able to obtain information); the two sets may correspond, but usually some of the objects in the first sample are not surveyed. As Groves and Lyberg (1988: 191) pointed out, nonresponse error threatens the characteristic which makes the survey unique among research methods: its statistical inference from sample to population. If the sample is at odds with the probability model, nothing can be said about its general representativeness; that is, about whether it truly reproduces all the characteristics of the population.
- 3 *Representativeness and generalizability: two sides of the same coin?* The social science textbooks usually describe generalizability as the natural outcome of a prior probabilistic procedure. In other words, the necessary condition for carrying out a statistical inference is previous use of a probability sample. It is forgotten, however, that probability/representativeness and generalizability are not two sides of the same coin. The former is a property of the *sample*, whilst the latter concerns the *findings* of research. Put otherwise: between construction of a sample and confirmation of a hypothesis there intervene a complex set of activities which pertain to at least seven different domains: (1) the trustworthiness of operational definitions and operational acts; (2) the reliability of the data collection instrument;

(3) the appropriateness of conceptualizations; (4) the accuracy of the researcher's descriptions, categorizations, and/or measurements; (5) to be successful with observational (or field) relations; (6) the validity of the data; and (7) the validity of the interpretation. These activities, and their relative errors (called 'measurement errors' in the literature), may impair the connection between probability/representativeness and generalizability – a not infrequent occurrence in a complex activity like social research.

These drawbacks do not signify that probability sampling and statistical inference are instruments by their nature unsuited to social research. Rather, according to the research setting, they are instruments with certain practical disadvantages that can sometimes be remedied and sometimes cannot.

In light of these difficulties, probability sampling cannot be propounded as the only model suited to the generalization of findings. As Geertz (1973: 21) points out, it is not only statistical inference that enables the move from 'local truths to general visions.' Moreover, as we have seen, not all sociological phenomena can be studied with rigorous application of the principles of sampling theory, the consequence being that the adoption of other forms of generalization has been vital for social research: otherwise, an important part of sociological theory (that based on research conducted on a few cases or even on haphazard or convenience samples as in the cases of, for example, Gouldner, Dalton, Becker, Goffman, Garfinkel, Cicourel) would never have been produced.

GENERALIZATION AS SEEN BY QUALITATIVE METHODOLOGISTS

Qualitative researchers have taken up a variety of positions in reaction to the pronouncement that those who do not use probability samples cannot generalize. The most extreme of them have (paradoxically) on the one hand accepted the verdict but on the other dismissed sampling as 'a mere positivist worry' (Lincoln and Guba, 1979; Denzin, 1983).

Some have aptly pointed out that 'most social anthropological and a good deal of sociological theorizing has been founded upon case studies' (Mitchell, 1983: 188) or has been the product of exclusively theoretical inquiry (without, that is, being grounded on systematic research). The more moderate have complied with the injunction of the statisticians but reconceptualized the problem by claiming that there are two types of generalization (which they have termed in various ways): *enumerative* (statistical) vs. *analytic* induction (Znaniecki, 1934: 236; Mitchell, 1983: 191); *formalistic/scientific* vs. *naturalistic* generalization (Stake, 1978: 6); *distributive* vs. *theoretical* generalization (Hammersley, 1992: 186ff; Williams, 2000: 215; Payne and Williams, 2005: 296–7). The first type of generalization involves estimating the distribution of particular features within a finite population; the second, eminently theoretical, is concerned with the relations among the variables in any sample of the relevant kind (moreover, the population of relevant cases is potentially infinite). The latter is usually based on identifying causal or essential relations among particular categories, whose character is defined by those relations, so that it is inferred that all instances of those categories are involved in the specified type of relation.

Even though some qualitative researchers may privately agree with Znaniecki (1934: 236–7) that analytical induction is the true method of science and it is the superior method (because it discovers the causal relations of a phenomenon rather than only the probabilistic ones of co-occurrence), the idea that there exist two types of generalization represents acceptance of the statisticians' diktat. It also represents acceptance of a 'political' division into areas of competence: a compromise already envisaged by some members of the Chicago School, like Burgess (1927), who maintained that statistics and case studies were mutually complementary² with their own criteria of excellence.

The distinction between the two types of generalization has been drawn with exemplary clarity by Alberoni and colleagues, who

wrote in their *Introduction* to a study on 108 political activists of the Italian Communist Party and the Christian Democrat Party as follows:

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

The two generalizations are therefore made in completely different ways³.

This moderate stance has been adopted by the majority of qualitative methodologists, some of whom have sought to underscore the difference between statistical and 'qualitative' generalization by coining specific terms for the latter. This endeavor has given rise to a welter of terms: 'naturalistic generalization' (Stake, 1978: 6), 'transferability' (Lincoln and Guba, 1979), 'translatability' (Goetz and LeCompte, 1984), 'analytic generalization' (Yin, 1984), 'extrapolation' (Mitchell, 1983: 191; Alasuutari, 1995: 196–7), 'moderatum generalization' (Williams, 2000; Payne and Williams, 2005), and others.

Five concepts of generalization

At least five different positions on the generalizability of research results can be identified within the qualitative methodological tradition (see Ragin and Becker, 1992; Gomm, Hammersley and Foster, 2000).

The first position, the most radical of them, has been assumed by Lincoln and Guba (1979), Guba (1981), Guba and Lincoln (1982), and Denzin (1983). It adheres to the traditional position that qualitative research is

an idiographic account which lays no claim to generalization (see Burrell and Morgan, 1979). Norman K. Denzin is very explicit on the matter:

The interpretivist rejects generalization as a goal and never aims to draw randomly selected samples of human experience. For the interpretivist every instance of social interaction, if thickly described (Geertz, 1973), represents a slice from the life world that is the proper subject matter for interpretative inquiry (...) Every topic (...) must be seen as carrying its own logic, sense or order, structure, and meaning. (Denzin, 1983: 133–4)⁴

Guba and Lincoln (1981: 62) likewise claim that 'it is virtually impossible to imagine any human behavior that is not heavily mediated by the context in which it occurs. One can easily conclude that generalizations that are intended to be context free will have little that is useful to say about human behavior.' However, Guba and Lincoln moderate their position by introducing two novel elements: a new formulation of the concept of *working hypothesis* proposed by Cronbach (1975), and the new concept of *transferability*.

According to Cronbach, 'when we give proper weight to local conditions, any generalization is a working hypothesis, not a conclusion' (1975: 125). Hence, Lincoln and Guba maintain,

local conditions (...) make it impossible to generalize. If there is a 'true' generalization, it is that there can be no generalization. And note that 'working hypotheses' are tentative both for the situation in which they first uncovered and for other situations; there are always differences in context from situation to situation, and even the single situation differs over time. (1979, reprinted 2000: 39)

They now make their own proposal, which has become well-known in qualitative research:

How can one tell whether a working hypothesis developed in Context A might be applicable in Context B? We suggest that the answer to that question must be empirical: the degree of *transferability* is a direct function of the *similarity* between the two contexts, what we shall call '*fittingness*'. Fittingness is defined as the degree of congruence between sending and receiving contexts. If Context A and Context B are 'sufficiently' congruent, then working

hypotheses from the sending originating context *may* be applicable in the receiving context. (Lincoln and Guba, 1979, reprinted 2000: 40)

However, transferability is not an inferential process performed by the researcher (who cannot know all the other contexts of research). Rather, it is a choice made by the reader, who on the basis of argumentative logic and a thick description (of the case study) produced by the researcher, may decide (on his/her own responsibility – see Gomm, Hammersley and Foster, 2000: 102) to *transfer* this knowledge to other situations that she/he deems similar (Lincoln and Guba, 1979, reprinted 2000: 40). The reader, basing this on the persuasive power of the arguments used by the researcher, decides on the similarity between the (sending) context of the case studied and the (receiving) contexts to which the reader him/herself intends to apply the results (Guba and Lincoln, 1982: 246).

To conclude, these authors are convinced that ‘generalizations are impossible since phenomena are neither time- nor context-free’; however, ‘some transferability of these hypotheses may be possible from situation to situation, depending on the degree of temporal and contextual similarity’ (Guba and Lincoln, 1982: 238).

A second, more moderate, approach has been proposed by Stake (1978: 1994), who argues that the purpose of case studies is not so much to produce general conclusions as to describe and analyze the principal features of the phenomenon studied. If these features concern an emblematic case of political, social, or economic importance (for example, the decision-making procedures of a large institution like the US Department of Defense), the ‘intrinsic case study’ will *per se* produce results of indubitable intrinsic relevance⁵, even though they cannot be generalized in accordance with the canons of scientific induction:

naturalistic generalization, arrived at by recognizing the similarities of objects and issues in and out of context and by sensing the natural covariations of happenings. To generalize this way is to be both intuitive and empirical, and not idiotic.

Naturalistic generalizations develop within a person as a product of experience. They derive from the tacit knowledge of how things are, why they are, how people feel about them, and how these things are likely to be later or in other places with which this person is familiar. They seldom take the form of predictions but they lead regularly to expectations (...) These generalizations may become verbalized, passing of course from tacit knowledge to propositional; but they have not yet passed the empirical and logical tests that characterize formal (scholarly, scientific) generalizations. (Stake, 1978: 6)

A third position, which is contiguous to the intrinsic case study, has been put forward by Connolly (1998). It starts from the distinction between extensive vs. intensive studies. The aim of the former (like case studies) is to identify statistically significant and therefore generalizable causal relations; the aim of the latter is to reconstruct in detail the mechanisms that connect cause and effect. Like Stake, Connolly relieves the case study of responsibility for formal generalization, but he gives it a task complementary to such generalization, explaining (via the mechanisms) correlations whose statistical significance has already been documented by other studies.

These three positions have a common basis consisting in the concept of ‘theoretical sampling’ proposed by Glaser and Strauss (1967), Schatzman and Strauss (1973) and Strauss (1987): when we do not possess complete information about the population, cases are selected according to their status on one or more properties identified as the subject matter for research. 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). And Strauss and Corbin are very explicit on the concept of generalization:

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)

In other words the aim is not to generalize to some finite population but to develop theoretical ideas that will have general validity.

More practical are authors who engage in 'evaluation research' (Cronbach, 1982; Pawson and Tilley, 1997). These ground their reasoning on the notion of the cumulability of knowledge: case study after case study, in the course of time in a particular sector of research, there accumulates a repertoire or inventory of the possible forms that a particular object of study may assume. As Pawson and Tilley (1997: 119–20) put it, in polemic with Guba and Lincoln, what can be transferred between studies are not 'lumps of cases' but 'sets of ideas' which enable understanding of general mechanisms. In other words, cumulability is the prelude for qualitative generalizability.

The final position, and perhaps the oldest of them, is represented by Znaniecki's method of analytic induction. The purpose of analytic induction is to uncover causal relations through identification of the essential characteristics of the phenomenon studied. To this end, the method starts not with a hypothesis but with a limited set of cases from which an initial explanatory hypothesis is then derived. If the initial hypothesis fails to be confirmed by one case, it is revised. Additional cases of the same class of phenomena are then selected. If the hypothesis is not confirmed by these further cases, the conceptual definition of the phenomenon is revised. The process continues until the hypothesis is no longer refuted and further study tells the researcher nothing new (Znaniecki, 1934: 236ff). The inner logic of analytic induction derives from Mill's 'method of agreement' and 'method of difference.'

There are several variants of Znaniecki's method of analytic induction. One of them is Mitchell's (1983) critical case study approach.

Analytic induction revisited has been also widely used in comparative studies based on a small numbers of cases 'when little more

than a handful of nations or organizations – sometimes even fewer – are compared with respect to the forces driving a societal outcome such a political development or an organizational characteristic' (Lieberson, 1992, reprinted in 2000: 208).

The unavailability of generalization

Sampling and generalizing are unavoidable practices because, even before being scientific, they are everyday life activities deeply rooted in thought, language, and practice (Gobo, 2004). With regard to thought, cognitive psychologists have demonstrated the tendency of people to generalize on the basis of a few observed characteristics or events, a process called the *heuristic of representativeness* by Kahneman and Tversky (1972) and Tversky and Kahneman (1974). With regard to the world of language, the same function is performed, as Becker 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 seller shows a sample of cloth to the customer; in a paint shop the buyer skims through the catalogue of color shades in order to select a paint; the buyer tastes in order to choose a wine or a cheese; the teacher asks a student questions to assess his or her knowledge about the syllabus. In everyday life, social actors constantly sample and generalize. As Gomm, Hammersley and Foster point out, 'we all engage in naturalistic generalizations routinely in the course of our life, and this may take the form of empirical generalization as well as of informal theoretical inference. Given this, there is no reason in principle why case study research should not provide the basis for empirical generalization' (2000: 104). This is also because the unavailability of generalization is epistemologically and reflexively founded. As Gomm, Hammersley and Foster acutely observe:

the very meaning of the word 'case' implies that what it refers to is a case [instance or example]

of something. In other words, we necessarily identify cases in terms of general categories (...) the idea that somehow cases can be identified independently of our orientation to them is false. It is misleading to talk of the uniqueness of cases (...) we can only identify their distinctiveness on the basis of a notion of what is typical or representative of some categorial group or population. (Gomm, Hammersley and Foster, 2000: 104)

The unavoidableness of generalizing is such that 'in practice, much case study research has in fact put forward empirical generalizations' (Gomm, Hammersley and Foster, 2000: 98) and 'current qualitative researchers often seem to produce [generalization] unconsciously' (Payne and Williams, 2005: 297).

FOR AN IDIOGRAPHIC SAMPLING THEORY

The thesis (which I have called 'moderate') that there are two types of generalization has had the indubitable merit of cooling the dispute with quantitative methodologists and of legitimating two ways to conduct research. However, this political compromise has also had a number of harmful consequences.

First, it has not stimulated reflection on how to emancipate 'qualitative' generalization from its subordination to statistical inference. Traditional methodologists continue to attribute inferior status to qualitative research, on the grounds that although it can produce interesting results, they have a limited extension only. This long-standing positivist prejudice has been recently reinforced by the extreme positions taken up by Lincoln and Guba (1979) and Denzin (1983). Their insistence that generalization in interpretative research is impossible, and that their work is not intended to produce scientific generalizations, paradoxically fits perfectly with the equally intransigent position of quantitative methodologists. As Gomm, Hammersley and Foster, observe, 'to deny the possibility of case studies providing the basis for empirical generalizations is to accept the views of their critics too readily' (2000: 98). Even though

it may be a coincidence, the fact that Egon Guba was a well-known statistician before he became a celebrated qualitative methodologist may have heightened the inflexibility of the debate. Consequently, an unexpected consequence of this paradox is that interpretivism has been just as positivist on qualitative generalization as quantitative methods have.

Second, the concept of theoretical sampling has failed to address the problem of sample representativeness which Denzin himself (1971) considered so important. Likewise, the concept of transferability provides 'no guidance for researchers about *which* case to study – in effect, it implies that any case may be as good as any other in this respect' (Gomm, Hammersley and Foster, 2000: 101). This omission has been pedagogically harmful because it has permitted several generations of qualitative researchers entirely to neglect – in the belief that 'anything goes' – this aspect of the investigative process.

Third, an opportunity has been missed to rediscuss the entire issue, addressing it in more practical (and not solely theoretical) terms with a view to developing a new sampling theory: an idiographic theory, joint and equal with statistical theory, and which remedies a series of ancestral misunderstandings:

denial of the capacity of case study research to support empirical [distributive] generalization often seems to rest on the mistaken assumption that this form of generalization requires statistical sampling. This restricts the idea of representation to its statistical version; it confuses the task of empirical generalization with the use of statistical techniques to achieve that goal. While those techniques are a very effective basis for generalization, they are not essential. (Gomm, Hammersley and Foster, 2000: 104)

Sampling in some contemporary sciences

The first step in this endeavor is to survey certain disciplines – paleontology, archaeology, geology, ethology, biology, astronomy, anthropology, cognitive science, linguistics (which for some scientists is more reputable than sociology) – and see how they have

tackled the problems of representativeness and generalizability. In certain respects, these are disciplines akin to qualitative research, for they work exclusively on few cases and have learnt to make a virtue out of necessity. As Becker writes:

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. (1998: 70–1)

For reasons of space, it is not possible here to provide an exhaustive account of how these disciplines have dealt with the above issues. But by way of example, consider the following study, which is one of the dozens published on the subject. It appeared in the journal *Nature* on January 23, 2003.

The scientist Xing Xu and colleagues (2003) of the Institute of Vertebrate Palaeontology, Beijing, had found six fossils in the province of Liaoning, North China. The impression left in the rock was of two pairs of wings and a long feathered tail of what appeared to be a *Microraptor gui*: a dinosaur less than one meter in length which lived in that region of China around 130 million years ago. According to its discoverers, the fossil was the missing link between terricolous dinosaurs and modern birds, the intermediate evolutionary stage for which scientists had long been searching. The discovery has fuelled the debate among paleontologists on the origin of flight. Whilst the close kinship between birds and dinosaurs is accepted by almost all scientists, there is much disagreement on the evolutionary stages that led to winged flight. The predominant theory is that wings began to develop, not to enable flight but to help the ancestors of birds to run faster. The small dinosaur discovered in China instead appeared to support the opposite hypothesis, namely that the direct ascendants of birds were animals

which climbed trees and used wings to glide back to earth. This was the theory propounded, for example, by the American naturalist, William Beebe, who as early as 1915 had predicted the existence of feathered dinosaurs exactly like *Microraptor gui*. However, the British journal urged caution when evaluating the importance of the discovery: the *Microraptor* could also be an evolutionary blind ally which had not left descendants.

There are therefore numerous disciplines which work on a limited number of cases, and do so consciously; in fact, there is animated discussion within them on sampling and generalizability. Moreover, this procedure is adopted by other disciplines as well: for instance, biology, astrophysics, history, genetics, anthropology, linguistics, cognitive science, psychology (whose theories are largely based on experiments, and therefore on research conducted on non-probabilistic samples consisting of psychology students). Why, we may ask, is this procedure acceptable for monkeys, rocks, and cells but not for human beings? Why do the majority of disciplines work with/on non-probability samples (regarded as being just as representative of their relative populations and therefore as producing generalizable results) while in sociology this is not possible? Why can a geneticist like Luca Cavalli Sforza of Stanford University argue that the evolution of language has had a direct impact on our genetic heritage, while in sociology a similar claim would require very different methodological support? The majority of these disciplines start from the assumption that their objects of study possess quasi-invariant states on the properties observed: that is, their states with respect to a property (e.g. size of the brain or the physique of a hominid) vary little and slowly among members of the class. Consequently, these disciplines are unconcerned about their use of only a handful of cases to draw inferences and generalizations about thousands of people, animals, plants, and other objects. Moreover, science studies the individual object/phenomenon not

in itself but as a member of a broader class of objects/phenomena with particular characteristics/properties.

FOUR PROPOSALS FOR AN IDIOGRAPHIC SAMPLING THEORY

The above survey of disciplines midway between the natural sciences and the social science yields a number of suggestions for formulation of an idiographic sampling theory. They can be summarized in the following four steps:

- (a) abandon the (statistical) principle of probability;
- (b) recover the (statistical) principle of variance;
- (c) pay renewed attention to the units of analysis;
- (d) identify social regularities.

Representativeness without probability

The use of probability samples does not automatically signify the use of representative samples. Random and representative are terms neither synonymous nor necessarily inter-related. 'Randomness' concerns a particular procedure used to select the cases to include in a sample, while 'representativeness' concerns the outcome of the selection. One may question whether the former is the obligatory path for the latter. Nor do representativeness and probability form a natural pair, since it may be possible to construct a representative sample using other procedures. Qualitative research (or at least a part of it) does not relinquish the aim of working with representative samples; it only rejects the obligatory nexus between probabilistic and representative (on the one hand), or between randomness and representativeness (on the other).

It is therefore not necessarily the case that a researcher must choose between an (approximately) random sample or an entirely subjective one – or between a sample which is (even only) partially probabilistic and one about whose representativeness absolutely

nothing can be said. Between the rationalism and the postmodern nihilism underlying these two positions, one may attempt to address the problem in practical terms, doing so by examining the nature of the units of analysis considered, rather than adhering to standard procedural rules. As stressed by Rositi (1993: 198), we may reasonably doubt the generalizability of findings from

studies of 1,000–2,000 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)

Variance: From (general) principle to (local) practice

The second step is to recover the (statistical) principle of variance, which has received less attention than the probability principle. Contrary to the latter's standardizing intent and automatist inclination (which are among the reasons for its success), variance is a criterion which requires the researcher to reason, to conduct contextual analysis, and to take local decisions. Under the variance principle,

in order to determine the sample size, the statistics must first know the range of variance that the researcher intends to measure (at least in sufficiently close terms) because it is likely that, if the range of variance of variable X is high, n [the number of individuals to interview] will be high, whereas if the range of variance is restricted (for example to only two modalities), n may be very restricted as well. (Capecchi, 1972: 50)

Hence, it is more likely that a sample will be a miniature of the population if that population is tendentially homogeneous; and it is less likely to be so if the reference population is tendentially heterogeneous. Consequently, if the variance is high, the researcher will require a large number of cases (in order to include every dimension of the phenomenon studied in his/her sample). If, instead, the variance is low, the researcher will presumably

need only a few cases, and in some instances only one. In other words,

it is important to recognize that the greater the heterogeneity of a population the more problematic are empirical generalizations based on a single case, or a handful of cases. If we could reasonably assume that the population were composed of more or less identical units, then there would be no problem. (Gomm, Hammersley and Foster, 2000: 104)

As also Payne and Williams (2005: 306–7) point out:

the *breadth* of generalization can be extensive or narrow, depending on the nature of the phenomenon under study and our assumptions about the wider social world (...) [hence] the generalization may claim high or lower levels of *precision of estimates* (...) [and it] will be conditional upon the ontological status of the phenomena in question. We can say more, or make stronger claims about some things than others. A taxonomy of phenomena might look like this: 1° physical objects and their social properties; 2° social structures; 3° cultural features and artefacts; 4° symbols; 5° group relationships; 6° dyadic relationship; 7° psychological dispositions/behaviour (...) This outline taxonomy demonstrates that generalizations depend on what levels of social phenomena are being studied.

The conversation analyst Harvey Sacks (1992, vol. 1: 485, quoted in Silverman, 2000: 109) reminds us of the anthropologist and linguist Benjamin Lee Whorf, who was able to reconstruct Navajo grammar by extensively interviewing only one native Indian speaker. Grammars usually have low variance. However, had Whorf wanted to study how the Navajo educated their children, entertained themselves, etc., he would (perhaps) have found greater variance in the phenomenon and would have needed more cases. On this logic, the formal criteria that guide sampling are more informed by and embedded in sociological (rather than statistical) reasoning based on contingent reflection about the dimensions specific to the phenomenon investigated and the knowledge objectives of the research.

Moreover, as said, an authoritative part of sociological theory and a large part of anthropological theory are based on

the case study: the quintessence of non-probability sampling. The research studies by Alvin G. Gouldner and Melvin Dalton belong to this category. For example, Gouldner (1954) studied a gypsum mine situated close to the university where he taught (a convenience sample, therefore⁶). In his methodological appendix, Gouldner reported that his team conducted 174 interviews – and therefore on almost all the population (precisely 77%). One hundred and thirty-two of these 174 interviews were conducted with a ‘representative sample’ of the blue-collar workers at the company, for which purpose Gouldner used quota sampling stratified by age, rank, and tasks. He then constructed another representative sample of 92 blue-collar workers, to whom a questionnaire was administered.

Dalton (1959), who was a company manager at that time, conducted covert observation at Milo and Fruhuling, the fictitious names of two American companies for which he worked as a consultant (again a convenience sample, therefore). The ethnologist De Martino (1961) observed 21 people suffering from tarantism disease; Goffman (1961) stayed for several months at a psychiatric hospital; the anthropologist Geertz (1972) attended 57 cock fights; Sacks and colleagues described the mechanics of conversational interaction by analyzing a few telephone calls; the anthropologist Crapanzano (1980) studied Moroccan social relations through the experience of Tuhami, a tilemaker. The anthropologist Griaule (1948) reconstructed the cosmology of the Dogon, a tribe in Mali, by questioning only a small group of informants; Bourdieu’s book (1993) on professions was based on 50 interviews with policewomen, temporary workers, attorneys, blue-collar workers, civil servants, and unemployed workers.

Why, one may ask, have such circumscribed studies given rise to such wide-ranging theories? In other words, why have they been generalized to other contexts? I shall answer these questions later. For the moment I would stress (and avoid) the danger of the nihilistic or postmodern drift implied

by this approach, where any sample may serve and it is not worth bothering too much about it. Instead, at a certain point of the inquiry, giving clear definition to the *units of analysis* (an operation performed before the cases are selected, and therefore before the sample is constructed) is of extreme importance if the research is not to be botched and empirically inconsistent. On analyzing a series of Finnish studies on 'artists,' Mitchell and Karttunen (1991) found that the results differed according to the definition given to 'artist' by the researchers, a definition which then guided construction of the sample. In some studies, the category 'artist' included (i) subjects who defined themselves as artists; (ii) those permanently engaged in the production of works of art; (iii) those recognized as artists by society at large; and (iv) those recognized as such by associations of artists. The obvious consequence was that it was subsequently impossible to compare the results of these studies.

Units of analysis

The standard practice in sociology and political science is to choose clearly defined and easily detectable individual or collective units: persons, households, groups, associations, movements, parties, institutions, organizations, regions, or states. The consistency of these collective subjects is vague. In practice, members of these groups are interviewed individually: the head of the family, the human resources manager, the statistics department manager, and so on.

This means that the *sampling* unit (e.g. the family) is different from the *observational* unit (i.e. the single respondent as a member of the family). Only a focus group can (at least to some extent) preserve the integrity of the collective subject. Instead, choosing individuals implies an atomistic rather than organic conception of society (Burgess, 1927), whose structural elements are taken for granted or reckoned to be mirrored in the individual (Galtung, 1967: 37), while the sociological tradition that gives priority to relations over individuals is neglected. As a consequence, the following more dynamic units are neglected as well:

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

Hence, 'a reliable sampling model that recognizes interaction must be adopted [so that sampling is conducted on] interactive units (such as social relationships, encounters, organizations)' (Denzin, 1971: 269).

The researcher should focus his/her investigation on these kinds of units, not only because social processes are more easily detectable and observable, but also because these units allow more direct and deeper analysis of the characteristics observed.

Consider the following illustrative example. Assume that we want to study work practices at call centers, which are technology-intensive workplaces. In Italy, it has been calculated that there were 1350 call centers in 2002. In order to construct a probability and representative sample, we may proceed in two ways: randomly extract a certain number of cases from the population list (which is possible because a complete list can be obtained from the Chambers of Commerce), or construct a proportional stratified sample. In this latter case, we must first classify call centers according to the properties that interest us:

- the ownership of the organization, so that we have private call centers (e.g. Vodafone), public ones (e.g. the 911 emergency helpline), and non-profit ones;

- the 'vocation,' so that we have call centers that are 'generalist' (in the sense that they provide a variety of services) or 'vertical' (i.e. dedicated to only one service, e.g. credit recovery);
- membership or otherwise of the organization for which the service is provided, so that we have call centers 'internal' to the company, or ones to which the work is outsourced;
- the classic variables such as size of the organization (small, medium, large), geographical location (north-west, north-east, centre, south, islands), etc.;
- the type of service furnished.

Note that many of these properties are mutually exclusive, so that the sampling decision must be carefully pondered. In these cases, the usual practice is for the researcher to base the probability sampling on the first property. However, this may be sociologically inadequate if the researcher's interest is in work practices, because these cannot be accessed via the variable 'ownership.' For some authors (e.g. Capecchi, 1972), representativeness does not seem to transfer from one property to another. Put otherwise: it is not the variance of the ownership of call centers that interests us here, but the variance of work practices. It might be more satisfactory to choose property (e). Experience of this sector of inquiry (but also the literature, previous research, interviews with experts or operators in the sector, etc.) shows that call centers mainly provide the following services: counseling, credit recovery, marketing, interviewing, and advertising. Constructing a probability sample on this classification is practically impossible because a population list for each of these activities does not exist. The only alternative is to use the method outlined in the previous section. Again on the basis of experience, we note that only the first of these five activities has substantial variance, while the four latter seem to have low variance. In fact, the counseling provided by call centers is multiform: it consists of information, technical assistance, psychological help or support, medical advice, or therapy. Consequently, in order to preserve the representativeness of the sample, we must sample several cases for the specific work practice of counseling. If we have insufficient resources to collect the necessary number of cases, we can restrict our research to only some activities. Other studies in the future will account for the rest.

It is evident that representativeness is not always possessed by the sample when research begins. It is a resource also acquired *ex post*, progressively and iteratively, research project after research project, with the gradual accumulation of expertise. This definition of representativeness seems somehow to tie this property to the relation between the results obtained by an individual research project and the experience of the researcher who conducts it.

In search of social regularities

I now turn to the final aspect of the entire question. There are three broad criteria which serve to orient the construction of a non-probability sample; and to each of them corresponds a particular form of reasoning alternative to inductive or statistical inference: deductive inference, comparative inference, and emblematic case.

The three criteria impose different cognitive objectives, and they are used according to the type of generalization that the researcher

wants to make. The first two criteria are in some way opposed to each other: comparative inference maximizes the probability of extracting odd cases; deductive inference selects only odd (deviant) cases. Theoretical inference instead concentrates on emblematic cases, focusing on social similarities.

Deductive inference

The first criterion consists of the choice of a critical or deviant case which can be used (*à la* Popper) to prove the refutability of an accredited or standard theory. An outstanding example of its application is provided by Goldthorpe et al.'s study (1968) of workers in the town of Luton. The distinctive feature of this inferential process is that it starts from a theory of which it intends to prove the implausibility: in this case the embourgeoisement of the working class. The theory is tested against a case comprising the largest number (and the greatest intensity) of its founding properties or requirements of this theory. If, in these optimal conditions, the consequences foreseen by the theory do not ensue, it is

extremely unlikely that the theory will work in all those empirical cases where those requirements are more weakly present. Hence the theory is falsified, and its inadequacy can be legitimately generalized. When the critical case study procedure is used, the cases are selected according to their explanatory power, rather than according to the criteria of probability theory or their typicality (Mitchell, 1983: 207, 209). Moreover, the legitimacy of the generalization (of the scant explanatory capacity of the theory just falsified) depends not only on the cogency of the rhetorical argument but also on the strength of the connections established between theory and observations.

There are many other important studies (which follow in a very broad sense the Popperian approach) which have focused on deviant cases in order to understand standard behavior: Goffman (1961) on ceremonies and rituals in a psychiatric clinic; Cicourel and Boise (1972) on the interpersonal communications of deaf children; Garfinkel (1967) on achievement of sex status in an 'intersexed' person; Pollner and Winkler (1985) on interactions in a family with a mentally retarded child; and many others.

This criterion can also be used to explore subcultures or emergent or avant-garde phenomena which may become dominant or significant in the future, although at present they are still marginal: see Festinger et al. (1956) on millennial groups after their predicted date for the end of the world had passed; Becker (1953) on marijuana smokers; Hebdige (1979) on style groups like mods, punks, skinheads; Fielding (1981) on right-wing political movements.

The deviant case can also be used to prove the refutability and falsifiability of a well-known and received theory, as in Rosenhan's (1973) study on the medical-organizational origin of psychiatric illness, or the already-cited study by Goldthorpe et al. (1968) on blue-collar workers in the town of Luton. This criterion (which is widely applied in biology, astrophysics, history, genetics, anthropology, linguistics, paleontology, archaeology, ethnology, geology) does not determine the extent

to which a phenomenon is widespread in the population. It only directs the scientific community's attention to the phenomenon's existence and the need to revise the dominant theory. The generalization to the population comes about by default: that is by virtue of the non-occurrence of the event foreseen by the theory under examination.

Obviously, the generalization must be carefully thought through. Otherwise, the danger arises of lapsing into the determinism to which Popper's falsificationism is susceptible. As Lieberman (1992: 212) emphasizes:

it is very difficult to reject a major theory because it appears not to operate in some specific setting. One is wary of concluding that Max Weber was wrong because of a single deviation in some inadequately understood time or place. In the same fashion, we would view an accident caused by a sober driver as failing to disprove the notion that drinking causes automobile accidents.

Comparative inference

The second criterion is used to make generalizations similar to statistical inferences, but without employing probability criteria. This can be done by identifying cases within extreme situations as well as certain characteristics, or cases within a wide range of situations in order to maximize variation, that is, to have all the possible situations in order to capture the heterogeneity of a population. We can choose two elementary schools where, from press reports, previous studies, interviews or personal experiences, we know we can find two extreme situations: in the first school there are severe difficulties of integration between natives and immigrants, while in the second there are virtually none. We can also pick three schools: the first with severe integration difficulties; the second with average difficulties; and the third with rare ones. In the 1930s and 1940s, the American sociologist W. Lloyd Warner (1898–1970) and his team of colleagues and students carried out studies on various communities in the United States. When Warner set about choosing the samples, he decided to select communities whose social structures mirrored important features of American society. He chose four

communities (given assumed names): a city in Massachusetts (Yankee City) ruled by traditions on which he wrote five volumes; a lonely county of Mississippi (Deep South, 1941); a Chicago black district (Bronzestown, 1945); and a city in the Midwest (Jonesville, 1949).

In comparative inferences, the cases are selected by making careful comparisons: first by seeking to find cases which represent all the forms of heterogeneity in a target population, and then by controlling whether they are sufficiently homogeneous with the type that one wants to represent. In this difficult but important analysis,

it is necessary to compare the characteristics of the case(s) being studied with available information about the population to which generalization is intended (...) we are suggesting that where information about the larger population (or about overlapping populations) is available, it should be used. If it is not available, then the potential risks involved in generalization still need to be noted, preferably via specification of likely types of heterogeneity that could render the findings unrepresentative. (Gomm, Hammersley and Foster, 2000: 105–106)

We are therefore very distant from the concepts of naturalistic generalization and transferability, which are unsatisfactory in various respects, for they ‘do not provide a sound basis for the design, or justification, of case study research’ (Gomm, Hammersley and Foster, 2000: 102). They assign the reader a function which should also be performed by the researcher (assuming responsibility for affirming the generalizability of the study’s findings). They therefore relieve the researcher of responsibility for the careful selection of cases on the basis of the variance principle, and not solely on the basis of the theoretical significance of theoretical sampling and of all research on variables (rather than cases). As Schofield (1990) notes, all too often cases seem to be chosen for reasons of convenience and are therefore atypical in various respects.

The emblematic case

If we bear the variance principle in mind, there emerges a third major criterion for

the construction of a sample: the typical or emblematic case.

Gouldner’s case studies (1954) on bureaucratization in medium-sized firms, or that by Cicourel (1968) on the relational construction of the figure of the juvenile delinquent, have been considered amply generalizable (by both researchers and readers), probably because they were typical cases and consequently grasped structural aspects of the social action in the organizations studied. Nor should we forget that the question of generalizability is closely tied to the phenomenon being researched, according to the degree of variance in its states.

This means that it is possible to find cases which on their own can represent a significant feature of a phenomenon. Generalizability thus conceived concerns more general structures and is detached from individual social practices, of which they are only an instance. In other words, the scholar does not generalize the individual case or event, which as Weber stressed is unrepeatable, but the key structural features of which it is made up, and which are to be found in other cases or events belonging to the same species or class. As Becker has recently pointed out:

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)

On this view, the question of generalizability assumes a different significance: for example in the conclusions to his study on the relationship between a psychotherapist and a patient suffering from AIDS, Peräkylä writes:

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. (1997: 216, quoted in Silverman, 2000: 109)

Something similar happens in film and radio productions with noise sampling. The squeak of the door (which gives us the shivers when we watch a thriller or a horror film) 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. These are two different ways of thinking, and most social sciences seek to find patterns of this kind.

While the verbal expressions of an interactive exchange may vary, exchange based on the question-answer pattern features a formal trans-institutional (though not universal) structure. While laying a page of a newspaper on the floor and declaring one's sovereignty over it (Goffman, 1961) is a behavior observed in one psychiatric clinic only, the need to have a private space and control over a territory has been reported many times, albeit in different forms.

INTERACTIVE, PROGRESSIVE, AND ITERATIVE SAMPLING: SOME TIPS

Having outlined the theoretical premises of an idiographic sampling theory, I shall now describe its procedural aspects. However, there is no precise logical itinerary to set out, because methodological principles and rules do not have to stand on their own – as they are instead required to do in statistical sampling theory – in that they have only a weak relation to practice. It is instead necessary to approach the entire question of sampling sequentially, and it would be misleading to plan the whole strategy beforehand. In order to achieve representativeness, the sampling plan must be set in *dialogue* with field incidents, contingencies, and discoveries. This is what I mean by 'interactive, progressive, and iterative sampling.' An excellent instance of this procedure 'is given in Glaser and Strauss's (1964, 1968) studies on dying in the hospital, where hypotheses were developed hand in hand with data collection' (Denzin, 1971: 269). Another example of changing or adding to the sampling plan on the basis of something

the researcher has learnt in the field is provided by Becker:

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! (Howard Becker, 13/7/2002, personal communication)

Consistency must be given to the sampling reasoning, but not by mere application of procedural steps. The reasoning could be as follows.

- 1 The researcher usually starts from his/her research questions. Melvin Dalton's were:

Why did grievors 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? (1959: 274)

Research questions comprise the concepts and categories (behaviors, attitudes, and so on) that the researcher intends to study.

- 2 The researcher conducts *primary* (or 'provisional' and 'open'⁷: Strauss and Corbin, 1990: 193) sampling in order to collect cases in accordance

- with the concepts. As Payne and Williams (2005: 295) suggest, 'research design should *plan* for anticipated generalizations, and that generalization should be more *explicitly formulated* within a context of supporting evidence.'
- 3 Because not every concept can be directly studied, when the researcher constructs the provisional sample, s/he considers the following aspects:
 - a specificity (focusing on specific social activities with distinctive features, like rituals or ceremonies);
 - b the field's degree of openness (open or closed places);
 - c intrusiveness (the endeavor to reduce the researcher's visibility);
 - d institutional accessibility (free-entry *versus* limited-entry situations *within* the organization);
 - e significance (frequent and high organizational significance of social activities).
 - 4 It is advisable to sample type of actions or events: 'not, then, men and their moments. Rather moments and their men' (Goffman, 1967: 3), 'not only people but moments of lived life' (Converse and Schuman, 1974: 1), '**incidents** and not persons per se!' (Strauss and Corbin, 1990: 177), in contrast with the common practice of sampling bodies, and of seeking information from these bodies about behaviors and events that are never observed directly (Cicourel, 1996). There are two reasons for this important recommendation: first, it serves to prevent the survey sampling mistake concerning the transferability of ideas about representativeness; second, the same person may be engaged in overlapping activities. For example, Dalton (1959), when studying power struggles in companies, found five 'types of cliques:' vertical (symbiotic and parasitic), horizontal (defensive and aggressive), and random. If we sample individuals, we find that they belong to more than one clique according to the situation, intention, and so on. If we consider activities, everything becomes simpler.
 - 5 To date, four main types of sampling have been developed in social research: purposive, quota, emblematic, and snowball. When cases are selected, attention should be paid to the variance of concept, so that different voices or cases can be included in the sample.
 - 6 As the research proceeds, the researcher will refine his/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 though not every hypothesis is testable (indeed the most interesting ones often are not), if the reader is to be persuaded, they must be formulated in a testable way.
 - 7 When the researcher has formulated hypotheses, s/he restarts sampling in order to collect cases systematically relating to each hypothesis, and seeking to make his/her analysis consistent. Strauss and Corbin call this second sampling 'relational and variational: is associated with axial coding. It aims to maximize the finding of differences at the dimensional level' (1990: 176). They depict the research process as funnel-shaped: through three increasingly focused steps (open, axial, and selective) the researcher clarifies his/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/he must always check whether it occurs in other samples.
 - 8 Generalization must be ensured 'across and within cases (...) [because] the danger of error in drawing general conclusions from a small number of cases must not be underestimated' (Gomm, Hammersley and Foster, 2000: 98). This concept has been sometimes rubricated as 'internal generalization,' and it implies different strategies which take account of diverse dimensions: time, sites, days, and people. The researcher should collect cases of behavior recurring at different moments of time. Because the researcher cannot observe the case-study population twenty-four hours a day, s/he must take a decision on when and where s/he will observe the population (Schatzman and Strauss, 1973: 39–41; Corsaro, 1985: 28–32). Unfortunately, case study researchers rarely make clear what they take to be the temporal boundaries of the cases they have studied (...) it is not unusual for case studies of schools to focus on one year-group or cohort of students and to assume that the experience of these students is representative of other cohorts, past and future. (Gomm, Hammersley and Foster, 2000: 109)
- Social practices always occur in certain places and at certain times of the day. Only if the researcher knows all the rituals of the organization observed can s/he draw a representative sample.
- A classic illustration is provided by Berlak et al.'s study of progressive primary school practice in

Britain in the 1970s (Berlak and Berlak, 1981; Berlak et al., 1975). They argued that previous American accounts had been inaccurate because observation had been brief and had tended to take place in the middle of the week, not on Monday or Friday. On the basis of these observations, the inference had been drawn that in progressive classrooms children simply chose what they wanted to do and got on with it. As Berlak et al. document, however, what typically happened was that the teachers set out the week's work on Mondays, and on Fridays they checked that it had been completed satisfactorily. Thus, earlier studies were based on false temporal generalizations within cases they investigated. (Gomm, Hammersley and Foster, 2000: 109–110)

Qualitative researchers do not seek to know the distribution of such behaviors (how many times); they only seek to know whether they are recurrent and significant in the organization under study. In addition, 'our concern is with representativeness of concepts' (Strauss and Corbin, 1990: 190). And finally, in regard to people and sites,

there is also likely to be variation in the behavior of both teachers and pupils across different contexts within a school. While most contact between members of the two groups probably occur in classrooms, they also meet one another in other places as well: in assembly halls, dining rooms, corridors, on game fields, and so on (...) Teacher-pupil relationships are likely to vary across mathematics classrooms, drama studios and science laboratories, for example. (Gomm, Hammersley and Foster, 2000: 111)

- 9 The researcher can sample new incidents or s/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 hypothesis testing 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 the literature on the topic. In other words, the researcher will come to know the

variance of a phenomenon cumulatively, study by study. As Gomm, Hammersley and Foster (2000: 107) acknowledge:

it is possible for subsequent investigations to build on earlier ones by providing additional cases, so as to construct a sample over time that *would* allow effective generalization. At the present, this kind of cumulation is unusual (...) the cases are not usually selected in such a way as to complement previous work;

- c representative samples are used to justify the researcher's statements.

It is therefore apparent that, although on the one hand 'generalization is not an issue that can be dismissed as irrelevant by case study researchers' (Gomm, Hammersley and Foster, 2000: 111), on the other it is not the impossible undertaking that survey researchers have always mocked. Finally, whilst probability sampling has a substantive aim – to construct a sample in order to extend the findings to the population – interactive sampling has a further task: to reflect, through its recursiveness, on the plausibility of generalizations.

CONCLUSION

Statistical inference (survey) and theoretical inference (experiment), as the two legitimate ways to draw general conclusions, continue to be used even though their application is fraught with difficulties; and they in fact end up by deviating from their theoretical principles and assumptions. Hence one fails to understand why it is not possible to resort to other forms of generalization which, though unsatisfactory, are no more unsatisfactory than those deemed superior to them. For that matter, contemporary social scientists do not have to choose between perfect and imperfect forms of generalization, but between forms of inference whose strengths and weaknesses depend on the researcher's cognitive aims, the research situation, and the nature of the phenomenon under study.

The central idea of this essay lies midway between two highly authoritative

and well-known methodological proposals: Durkheim's (1912) *cas pur* (the 'pure case'), with positivist overtones, and Max Weber's (1904) theory of ideal types. Durkheim believed that the simplest society of all for study of the elementary forms of religious life was the Australian tribe of the Arunta. The Flemish statistician and sociologist Adolphe Quételet (1796–1874) looked to the crowd for his *homme moyen* (the average man), who represented the 'normality' of the species. He was prompted to do so by the discovery that certain characteristics (physical and biological) of individuals were distributed in the populations which he studied according to the 'normal' curve constructed by the mathematician Gauss.

Conversely, Weber maintained that 'feudal society,' 'bureaucracy,' 'charisma' were genetic concepts (developed with a view to a causal explanation) and limiting concepts. They consequently could not be evaluated in terms of their reality-describing adequacy, only in terms of their instrumental efficacy. For Weber (1904), an ideal type was *not a representation* of the real; rather, it was formed by a one-sided *accentuation of one or more* points of view and by the connection of a quantity of diffuse, discrete, more or less present and occasionally absent, *particular* phenomena. Given the conceptual purity of an ideal type, it could never be empirically detected in reality; it was a utopian entity.

The typical or emblematic case suggested as a criterion for the construction of sample stands midway between the claim to have discovered the pure case (the quintessence of the phenomenon studied) and renunciation of the empirical search for cases of interest because of their typicality.

At the end of the 1980s, in a study on the interview, I documented the rituals and rhetorical strategies used by an interviewer as he made telephone calls to 10 adolescents in order to arrange subsequent face-to-face interviews (Gobo, 1990, 2001). The research involved the recording of the telephone calls and subsequent discourse analysis. Some years later, Maynard and Schaeffer (1999) conducted very similar research in the United

States. Comparison between the results of the two research studies showed that the three researchers had discovered almost identical patterns of behavior. The reason for this similarity was probably that the survey interviewers had been trained with textbooks widely used on both sides of the Atlantic, and that they had used artifacts – technological (telephone, keyboard), cognitive (questionnaires), and organizational (scripts or interview formats) – which made the social activities very similar. There are consequently numerous social research settings in which a few cases *may* suffice to make a generalization. Provided they are chosen carefully.

NOTES

1 To be stressed is that the distinction between probability and non-probability does not mark the boundary between qualitative and quantitative research: in fact, non-probability samples are also used for surveys (quota, telephone, and so on) and for experiments.

2 This compromise centered on the idea of complementarity is still accepted by numerous methodologists: see for instance Payne and Williams (2005: 297).

3 Indeed, there are some who maintain that *generalizability* is perhaps the wrong word for what qualitative researchers seek to achieve: 'Generalization is (...) [a] word (...) that should be reserved for surveys only' (Alasuutari, 1995: 156–7).

4 However, Denzin's (1971) position was very different at the end of the 1960s: he expressed himself in favor of operationalization ('this does not mean that operationalization is avoided – it merely suggests that the point of operationalization is delayed until the situated meaning of concepts is discovered,' p. 268); he believed that the use of indicators was important ('a series of empirical indicators relevant to each data base and hypothesis must be constructed, and, last, research must progress in a formative manner in which hypotheses and data continually interrelate,' p. 269), and he argued that 'it is necessary for researchers to demonstrate the representativeness of those units in the total population of similar events' (p. 269).

5 Gomm et al. (2000: 112, endnote 2) acutely point out: 'there is some ambiguity in Stake's position. He also recognizes that case studies can be instrumental rather than intrinsic, and in an outline of the 'major conceptual responsibilities' of case study inquiry he lists the final one as 'developing assertions or generalizations about the case (Stake, 1994, 244).'

6 For this reason, apparently too severe and without empirical justification is Payne and Williams' statement that: 'opportunistic site selection will normally be incompatible with even moderatum generalization' (2005: 310).

7 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 performed purposively (e.g. pp. 183–4) or systematically (e.g. p. 184), or it occurs fortuitously (e.g. pp. 182–3). It includes on-site sampling.

REFERENCES

- Alasuutari, Pertti 1995 *Researching Culture*, London: Sage.
- Alberoni, Francesco et al. 1967 *L'attivista di partito*, Bologna: Il Mulino.
- Bailey, Kenneth D. 1978 *Methods in Social Research*, New York: Free Press.
- Barton Allen H. and Lazarsfeld Paul F. 1955 Some functions of qualitative analysis in social research, *Frankfurter Beiträge zu Soziologie*, 1: 321–361.
- Becker, Howard. 1953 Becoming a Marijuana Smoker. *American Journal of Sociology*, 59: 235–242.
- Becker, Howard 1998 *Trick of the Trade*, Chicago and London: University of Chicago Press.
- Becker, Howard 2000 *Italo Calvino as Urbanologist*, paper.
- Bourdieu, Pierre. et al. 1993 *La Misere du monde*, Paris: Editions du Seuil, transl. *The Weight of the World: Social Suffering in Contemporary Society*, Cambridge: Polity, 1999.
- Burgess, Ernest W. 1927 Statistics and case studies as methods of sociological research, *Sociology and Social Research*, 12: 103–120.
- Burrell, Gibsen and Morgan, Gareth 1979 *Sociological Paradigms and Organizational Analysis*, London: Heinemann.
- Capecchi, Vittorio 1972 Struttura e tecniche della ricerca, in Pietro Rossi (ed.), *Ricerca sociologica e ruolo del sociologo*, Bologna: Il Mulino.
- Chain, Isidor 1963 An introduction to sampling, in C. Seltiz and M. Jahoda (eds.), *Research Methods in Social Relations*, New York: Holt & Rinehart, pp. 509–45.
- Cicourel, Aaron V. 1968 *The Social Organization of Juvenile Justice*, New York: Wiley.
- Cicourel, Aaron V. 1996 Ecological Validity and White Room Effects, *Pragmatic and Cognition*, 4(2): 221–263.
- Cicourel, Aaron V. and Boese, R. 1972 Sign language acquisition and the teaching of deaf children, in D. Hymes et al. (eds.), *Functions of Language in the Classroom*, New York: Teacher College Press.
- Connolly, Paul 1998 'Dancing to the wrong tune': Ethnography, generalization, and research on racism in schools, in P. Connolly and B. Troyna (eds.), *Researching Racism in Education*, Buckingham: Open University Press, pp. 122–39.
- Converse, Jean M. and Schuman, Howard 1974 *Conversations at Random: Survey Research as Interviewers See it*, New York: Wiley.
- Corsaro, William A. 1985 *Friendship and Peer Culture in the Early Years*, Norwood, N.J: Ablex Publishing Corporation.
- Crapanzano, Vincent 1980 *Tuhami. Portrait of a Moroccan*, Chicago: University of Chicago Press.
- Cronbach, Lee L. 1975 Beyond the two disciplines of scientific psychology, *American Psychologist*, 30: 116–27.
- Cronbach, L. 1982 *Designing Evaluations of Educational and Social Programs*, San Francisco: Jossey-Bass.
- Dalton, Melvin 1959 *Man Who Manage*, New York: Wiley.
- De Martino, Ernesto 1961 *La terra del rimorso*, Milano: Il Saggiatore, transl. *The Land of Remorse: A Study of Southern Italian Tarantism*, London: Free Association Books, 2005.
- Denzin, Norman K. 1971 Symbolic interactionism and ethnomethodology, in J.D. Douglas (ed.), *Understanding Everyday Life*, London: Routledge and Kegan Paul, pp. 259–284.
- Denzin, Norman K. 1983 Interpretive interactionism, in G. Morgan (ed.), *Beyond Method: Strategy for Social Research*, Beverly Hills, CA: Sage, pp. 129–46.
- Durkheim, Emile 1912 Les formes élémentaires de la vie religieuse, Paris: Alcan, transl. *The Elementary Forms of the Religious Life*, London: G. Allen & Unwin, 1915.
- Festingers, Leon, Riecken, Henry W. and Schachter, Stanley 1956 *When Prophecy Fails*, New York: Harper Torchbooks.
- Fielding, Nigel 1981 *The National Front*, London: Routledge.
- Galtung, John 1967 *Theory and Methods of Social Research*, Oslo: Universitets Forlaget.
- Garfinkel, Harold 1967 *Studies in Ethnomethodology*, Englewood Cliffs, NJ: Prentice Hall.
- Geertz, Clifford 1972 Deep play: notes on the Balinese Cockfight, *Dedalus*, 101: 1–37.
- Geertz, Clifford 1973 *The Interpretation of Culture*, New York: Basic Books.
- Glaser, Barney G. and Strauss, Anselm L. 1967 *The Discovery of Grounded Theory*, Chicago: Aldine.
- Gobo, Giampietro 1990 *The First Call: Rituals and Rhetorical Strategies in the First Telephone Call with Italian Respondents*, paper, Annual Meeting of the A.S.A., Washington D.C. August, 11–15.

- Gobo, Giampietro 2001 Best practices: rituals and rhetorical strategies in the 'initial telephone contact,' *Forum Qualitative Social Research*, 2(1), <http://www.qualitative-research.net/fqs-texte/1-01/1-01gobo-e.htm>.
- Gobo, Giampietro 2004 Sampling, representativeness and generalizability, in Seale C., Gobo G., Gubrium J.F., Silverman D. (eds.), *Qualitative Research Practice*, London: Sage, pp. 435–56.
- Goetz, J.P. and LeCompte, Margaret D. 1984 *Ethnography and Qualitative Design in Education Research*, Orlando, FL, Academic Press.
- Goffman, Erving 1961 *Asylums*, New York: Doubleday.
- Goffman, Erving 1967 *Interaction Ritual*, New York: Doubleday Anchor.
- Goldthorpe, John H., Lockwood, David, Bechhofer, Frank and Platt, Jennifer 1968 *The Affluent Worker: Industrial Attitudes and Behaviour*, Cambridge: Cambridge University Press.
- Gomm, Roger, Hammersley, Martyn and Foster, Peter (eds.) (2000) *Case Study Method*, London: Sage.
- Goode, William and Hatt, Paul, K. 1952 *Methods in Social Research*, New York: McGraw-Hill.
- Gouldner, Alvin G. 1954 *Patterns of Industrial Bureaucracy*, New York: The Free Press.
- Griaule, Marcel 1948 *Dieu d'eau: entretiens avec Ogotemméli*, Paris: Éditions du Chêne.
- Groves, Robert M. and Lyberg, Lars E. 1988 An overview of nonresponse issues in telephone surveys, in R.M. Groves, P.P. Biemer, L.E. Lyberg, J.T. Massey, W.L. Nicholls II and J. Waksberg (eds.), *Telephone Survey Methodology*, New York: Wiley.
- Guba, Egon G. 1981 Criteria for assessing the trustworthiness of naturalistic enquiries, *Educational Communication and Technology Journal*, 2(29): 75–92.
- Guba, Egon G. and Lincoln, Yvonna S. 1981 *Effective Evaluation: Improving the Usefulness of Evaluation Results Through Responsive and Naturalistic Approaches*, San Francisco: Jossey-Bass.
- Guba, Egon G. and Lincoln, Yvonna S. 1982 Epistemological and methodological bases of naturalistic inquiry, *Educational Communication and Technology Journal*, 30: 233–252.
- Hammersley, Martyn 1992 *What's Wrong with Ethnography?*, London: Routledge.
- Hebdige, Dick 1979 *Subculture: The Meaning of Style*, London and New York: Routledge.
- Kahneman, D. and Tversky, A. 1972 Subjective probability: A judgment of representativeness, *Cognitive Psychology*, 3: 430–454.
- Lieberson, Stanley 1992 *Small N's and Big Conclusions: An examination of the Reasoning in Comparative Studies Based on Small Number of Cases*, reprinted in R. Gomm, Hammersley, M. and Foster P. (eds.) (2000), *op. cit.*
- Lincoln, Yvonna, S. and Guba, Egon, G. 1979 *Naturalist Inquiry*, Beverly Hills, CA: Sage. (Reprinted partially in Gomm Roger, Hammersley, Martyn and Foster, Peter (eds.) 2000 *Case Study Method*, London: Sage, pp. 27–42.
- Mason, Jennifer 1996 *Qualitative Researching*, Newbury Park: Sage.
- Maynard, Douglas W. and Schaeffer, Nora Cate 1999 Keeping the gate, *Sociological Methods & Research*, 1: 34–79.
- Mitchell, Clyde J. 1983 Case and situation analysis, *Sociological Review*, 31: 187–211.
- Mitchell, R. and Karttunen, S. 1991 Perché e come definire un artista?, *Rassegna Italiana di Sociologia*, XXXII(3): 349–64.
- Pawson Ray and Tilley Nick 1997 *Realistic Evaluation*, Sage: London.
- Payne, Geoff and Williams, Malcolm 2005 Generalization in qualitative research, *Sociology*, 39(2): 295–314.
- Peräkylä, Anssi 1997 Reliability and validity in research based upon transcripts, in David Silverman (ed.), *Qualitative Research*, London: Sage, pp. 201–19.
- Pollner, Melvin and McDonald, Wikler Lynn 1985 The social construction of unreality: a case study of a family's attribution of competence to a severely retarded child, *Family Process*, 24: 241–254.
- Ragin Charles C. and Becker Howard S. (eds.) 1992 *What is a Case?* Cambridge: Cambridge University Press.
- Rosenhan, David L. 1973 On being sane in insane places, *Science*, 179: 250–8.
- Rositi, Franco 1993 Strutture di senso e strutture di dati, *Rassegna Italiana di Sociologia*, 2: 177–200.
- Sacks, Harvey 1992 *Lectures on Conversation*, Oxford: Blackwell.
- Schatzman, Leonard and Strauss, Anselm L. 1973 *Field Research*, Englewood Cliffs, NJ: Prentice Hall.
- Schofield Janet Ward 1990 Increasing the generalizability of qualitative research, in E.W. Eisner and A. Peshkin (eds.), *Qualitative Inquiry in Education: The Continuing Debate*, New York: Teachers College Press, pp. 201–232.
- Silverman, David 2000 *Doing Qualitative Research*, London: Sage.
- Stake, Robert 1978 The case study method in social enquiry, *Educational Researcher*, 7: 5–8 (Reprinted in Gomm Roger, Hammersley, Martyn and Foster, Peter (eds.) 2000 *Case Study Method*, London: Sage, pp. 19–26).
- Strauss, Anselm 1987 *Qualitative Analysis for Social Scientists*, Cambridge: Cambridge University Press.

- Strauss, Anselm and Corbin, Juliet 1990 *Basics of Qualitative Research*, London: Sage.
- Tversky, Amos and Kahneman, Daniel 1974 Judgment under uncertainty: Heuristics and biases, *Science*, 185: 1123–1131.
- Weber, Max 1904 Die 'Objektivität' sozialwissenschaftlicher und sozialpolitischer Erkenntnis, *Archiv für sozialwissenschaft und Sozialpolitik*, XIX: 22–87, transl. On the methodology of the social sciences, Illinois: The Free Press of Glencoe, 1949.
- Williams, Malcolm 2000 Interpretativism and generalization, *Sociology*, 34(2): 209–24.
- Xing Xu, Zhonghe Zhou, Xiaolin Wang, Xuewen Kuang, Fucheng Zhang and Xiangke Du 2003 Four winged dinosaurs from China, *Nature*, 421: 335–339.
- Yin, Robert K. 1984 *Case Study Research*, Thousand Oaks: Sage.
- Znaniecki, Florian 1934 *The Method of Sociology*, New York: Farrar & Rinehart.