



Generalization in Qualitative Research

■ **Geoff Payne**

University of Plymouth

■ **Malcolm Williams**

University of Plymouth

ABSTRACT

Earlier treatments of *moderatum* generalization (e.g. Williams, 2000a) explicitly addressed interpretivist sociology. This article extends that earlier argument by examining some of its implications for a wider range of qualitative research methods. It first adopts an empirical approach, providing concrete illustrations from the most recent volume of *Sociology* of what sociologists actually do when describing the meaning of their findings. In the light of this, we reconsider the significance of *moderatum* generalization for research practice and the status of sociological knowledge, in particular making the case that research design should *plan* for anticipated generalizations, and that generalization should be more *explicitly formulated* within a context of supporting evidence.

KEY WORDS

moderatum generalization / qualitative research methods / research design

In experimental or survey research, generalizing claims are explicit and constitute the explanation/generalization schema that is the basis of scientific reasoning. In qualitative research, generalizing claims are less explicit. Indeed, some interpretivist sociologists (e.g. Denzin, 1983; Denzin and Lincoln, 1995; Marshall and Rossman, 1989) minimize the relevance of generalization or even deny any intention toward generalization in qualitative research.

A belief that one must choose between an ‘interpretive sociology’, which rejects all generalization, and a sociology dependent on total or axiomatic generalizations (represented by statistical generalizations or physical laws) is too

simplistic (Williams, 2000a, 2000b, 2001). Qualitative research methods can produce an intermediate type of limited generalization, '*moderatum* generalizations'. These resemble the modest, pragmatic generalizations drawn from personal experience which, by bringing a semblance of order and consistency to social interaction, make everyday life possible. Indeed, a strong claim can be made that in qualitative research (even in the interpretivist sociology loudest in its rejection of generalization) such *moderatum* generalizations are unavoidable.

Williams' initial advocacy of the idea of *moderatum* generalization explicitly addressed interpretivist sociology. This article extends that earlier argument by examining some of its implications for a wider range of qualitative research methods. After discussing how generalization has been treated in the methodological literature, this article explores the issue in two rather different ways. It first adopts an empirical approach, by looking at what sociologists actually do. This provides both concrete illustrations and a small amount of evidence to support our contentions. The second part then reconsiders some of the significance of *moderatum* generalization for research practice and the status of sociological knowledge, in particular making the case that research design should *plan* for anticipated generalizations, and that generalization should be more explicitly formulated within a context of supporting evidence.

The Logic of Generalization

To generalize is to claim that what is the case in one place or time, will be so elsewhere or in another time. Everyday social life depends on the success of actors doing just this. The reliability of inductive generalization in everyday life (Schutz's lifeworld) is adequate for the purpose, permitting a great deal of incorrect inductive assumption without serious consequences. In the natural sciences (and later the social sciences) the desire for greater reliability led to the refinement of generalization toward inductive reasoning based on probability, or toward deductive (or axiomatic) reasoning (Hacking, 2001: 247–54). Deductive generalizations are possible in large-scale physical systems where invariant laws operate (albeit with local violations), whereas in the social world and much of the physical world (particularly at the biological level) systems are inherently complex and characterized by feedback mechanisms. This is especially the case in the social world where consciousness produces meaningful behaviour and individuals may attach different meanings to the same actions or circumstances, or conversely, act differently on the basis of similarly expressed meanings. Critics of generalization in sociology have alighted on this latter characteristic to maintain that generalization is impossible (Denzin, 1983: 133).

However, two classes of generalization are available to the sociologist. First, under certain circumstances, the more familiar inductive probabilistic reasoning may be used. This usually takes the form of generalizing from sample to universe in survey research.¹ Ideally this depends upon some form of prob-

ability sampling, whereby within each stratum every case has the same known probability of selection. Where a probability sample is not possible, other strategies aimed at emulating a probability sample are adopted (e.g. quota sampling, multiple capture). The issue of variance is not avoided in survey research, but arguably there is enough invariance in the social world to make statistical generalization possible. We realize that this last statement is contested, but the argument for the second class of generalization, *moderatum* generalization, is independent of (though complementary to) statistical generalization.

Moderatum generalizations are first and foremost moderate. They most resemble the everyday generalizations of the lifeworld in their nature and scope, though it is possible to express them formally.² They are moderate in two senses. First, the scope of what is claimed is moderate. Thus they are not attempts to produce sweeping sociological statements that hold good over long periods of time, or across ranges of cultures. Second, they are moderately held, in the sense of a political or aesthetic view that is open to change. This latter characteristic is important because it leads such generalizations to have a hypothetical character. They are testable propositions that might be confirmed or refuted through further evidence. For the reasons indicated above they can never lead to axiomatic generalizations, though they may be tested statistically. Statistical testing is not, however, a requirement: further research of other kinds is just as likely to sustain or modify them.

Although *moderatum* generalization is inherently modest, it cannot be taken to occur naturally and automatically in the process of doing research, or be left to the readers' discretion, or indeed, to the kinds of generalization that current qualitative researchers often seem to produce unconsciously. Research should be designed and undertaken in such a way that *moderatum* generalizations are consciously produced. Nor do such generalizations offer a 'soft option' to solving the problem of generalization: whereas everyday lifeworld generalizations need not be wholly accurate or carefully set in an evidential context, sociological research should aim at constructing externally valid and unambiguous generalizations, even when these take a moderated form.

Alternative Approaches to Qualitative Generalization

One of the two main approaches to generalization in qualitative sociology has been to emphasize internal validity and proceed as though what really matters is the *quality* of the *original* piece of research providing the platform on which generalization can be built. If the quality of the base study is satisfactory, then generalization is assumed to be potentiated: little attention need then be explicitly paid to considering whether generalization is sustainable. This perspective initially developed primarily among American quantitative sociologists in the 1980s. Geertz (1973), LeCompte and Goetz (1982), Mitchell (1983), and Lincoln and Guba (1985) offered a set of terms to handle reliability and

validity issues so that the findings of one study might be taken as establishing, inter alia, more general truths. 'External validity' depended on 'thick description' of the fieldwork; the richness of the data collected and full reportage of the care used in its collection serving two purposes. It demonstrated that the researcher's direct account was reliable and internally valid, and it provided the *reader* with the information necessary to decide whether the findings might be transferable to other settings.

In the proliferation of terminology, LeCompte and Goetz emphasize 'external reliability' as the form of generalization where a repeated study would produce the same results (in practice, replications are rare, and even rarer in print). Their focus is on re-study to validate the original work, rather than on generalization per se. They argue that reports should give extensive details of who the informants were; the social setting of the data collection; how the data were collected; the effect of the researcher's status on that process; and the researcher's original theoretical orientation. However, this reflexive audit addresses only a first stage in generalization, establishing the credibility of the initial study, and setting out the characteristics of the study which would be crucial for comparing the first with any subsequent study. In other words it addresses what Seale (1999: 108) usefully calls the 'sending' or actually researched site, but not the 'receiving' site(s) or universe to which the generalization might be applied.

Lincoln and Guba (1985) reformulate the problem, treating external reliability and replication under the headings of 'consistency' and 'dependency', and the question of how widely findings might be applied as 'transferability' (i.e. external validity). Their solution is for researchers to provide enough thick description 'to enable someone interested in making a transfer to reach a conclusion about whether the transfer can be contemplated as a possibility' (1985: 316). This still deals only with the sending end of the connection, nor can it always be left to the reader to decide whether to generalize. As we show, authors themselves are the ones who generalize.

Thick description, valuable though it is, places excessive emphasis on reliability. It has led to an implicit perception that generalization will be acceptable, provided only that the truth claims about sending sites (of the original accounts) are credible. In responding to other criticisms from quantitative methods about bias and neutrality, qualitative research has been distracted from generalization per se into discussions about the quality of the original research. Obviously, if an original study cannot claim certain standards of competence, there is no point in going on to generalize – but the same point applies to quantitative work.

A current illustration of this shortcoming is Spencer et al.'s (2003) *Quality in Qualitative Evaluation: A Framework for Assessing Research Evidence*. In producing their valuable guide to assessing research for use in 'the development and implementation of social policy' (2003: 2), its authors drew on an extensive literature review and a wide audience of specialists. However, among its 18 key topics only one directly addresses cross-applicability of findings, and among

86 sub-topics, there are (at the most generous) only nine mentions, including 'contribution to theory'. If a study is good enough on its own terms, generalization is largely assumed to be legitimate. What is missing is an integrated consideration of both the sending and the receiving sites at the same time.

For completeness, we should include 'theoretical generalization' and grounded theory in this overview. The plausibility of these theoretical accounts depends on their theoretical rigour rather than the typicality of the case, or any empirical comparison with other sites (e.g. Mitchell, 1983; Payne and Payne, 2004; Yin, 1984). Grounded theory (if used properly) calls for a rigorous iterative retesting of progressive interpretations against a full 'theoretical sample'. In this limited sense, interpretations have a stronger evidential base, but the focus is still on the initial study rather than on generalization.

More recently in Britain, some of these limitations have been discussed by Hammersley (1992), Silverman (1993) and Seale (1999), who in slightly different ways have explored the idea that the internal logic of studies might initially be assumed to have wider application, so drawing attention to the status of evidence in conclusions based on empirical findings. Hammersley has noted that the implied extension of any specific claim about systematic social patterns is problematic. Particularly in the context of the theoretical generalization debate, Seale is sceptical about the way 'unwarranted assumptions are made about the characteristics of the population of cases not yet studied' (1999: 112). It is in this latter tradition that the present article belongs, although these earlier contributions have for the most part addressed generalization in a broad sense, rather than moderated generalization.

The main alternative approach, at least in published articles, has been to completely ignore the problem. While this is methodologically inadequate, it has not prevented the publication of more than a dozen articles with this deficiency in *Sociology*, i.e. a leading journal, during 2003. On these grounds, avoiding the question is apparently a legitimate practice under contemporary canons of academic publishing in sociology.

Generalization in Current Qualitative Sociology

An examination of the articles, rejoinders and notes published in Volume 37 of *Sociology* (2003) will illustrate the centrality of generalization to sociological research, and offer some evidence for our stance. This does not provide a representative sample of all sociological activity but as the official, general sociology journal of the British Sociological Association, and one of the leading English language journals in the discipline, it is hard to argue that *Sociology* does not reflect mainstream tendencies in British sociology (Payne et al., 2004). Our intention is to show concrete examples of what has already been introduced in more general terms, i.e. our exemplifications are a heuristic, expository device, an *illustration*. While they are also a *small* unit of evidence in support of our argument they are not on their own a sufficient evidential base

for our argument, because they are not a probability sample selected to reflect all other published sociology. Our own generalizations must take moderatum form and are testable in other sites.

The 38 entries in Volume 37 include 14 which involve no empirical data. Although here we are not concerned with non-empirical contributions, we note in passing that theoretical articles usually entail under-evidenced generalizing statements about sociological practice. Of the other 24 empirically based contributions, every one makes generalizing statements, though seven are excluded from the discussion because they use quantitative methods and the logic of statistical sampling for their generalizations. It would be tedious to document every generalization from all 24 entries by quotation. We therefore list examples of pages where we have identified generalizing statements as a look-up table in the Appendix. Most of the examples come from 'conclusions' and abstracts.

There are no articles based on strictly interpretive methods in Volume 37. Two claim case study status but as they do not observe its conventions (presenting data that are regarded as specific to the setting and to be treated in isolation from other studies [Yin, 1984]), they seem to mean simply that they have single site data. The 17 contributions using qualitative methods almost all list more than one method of data collection, a total of 34 qualitative research techniques comprising 11 different types of qualitative method. Non-systematic interviewing (usually described as 'in-depth' or 'semi-structured') (eight cases) and documentary/content analysis (seven cases) were the most common methods, followed by focus groups and participant observation (five each) and key informants (three examples). Audiotape transcripts were used twice, while oral history, life histories, expressive tasking and grounded theory were all used once. Our discussion here can legitimately claim to cover most 'qualitative methods' (although we would not wish to insist on an absolute dichotomy between qualitative and quantitative methods).

The numbers of informants/sources varied (and was not always clear) but with two exceptions, data were collected from relatively few people: between eight (Kjølørød, 2003) and about 60 (e.g. Korac, 2003 and others). In almost all cases, the reader was given very little methodological information about why or how the specific informants had been recruited, or whether the selection of these particular informants or sources might influence the data that were collected. Two exceptions, where very brief reference to this was made, were Edwards and Imrie (2003: 245) and Pearson and Soothill (2003: 781).

There was almost no explicit discussion of the grounds on which findings might be generalized beyond the research setting. Despite this, *all the 17 articles made generalizations*, albeit of different kinds. Kelly (2003: 37) defended her generalizations on the grounds of later feedback from a conference with a wider range of key informants, while Hislop and Arber (2003: 710) cautiously linked their conclusions with a call for more studies. Punch (2003: 289–90) made generalizing claims but also denied making them. The most explicit comment on generalization came from Gladney et al. (2003: 311), who claimed

moderatum status for their position. However, all four exceptions consisted of only very brief comments.

The generalizations in some of the remaining qualitative articles could be captured as *moderatum* generalizations represented in the taxonomy which follows, but others were vague, sweeping and essentially immoderate. The vague generalizations were often juxtaposed with theoretical statements by other sociologists, so that it was hard to tell who was making the generalization. Words like 'suggest', 'tend', 'illustrate', and 'some of' were also used in a way that rendered claims unclear. This could be interpreted as further evidence of sociology's collective failure to achieve clarity in generalization.

Using Evidence in Qualitative Sociology

The limitations of current generalizing practices can be further illuminated by examining two *Sociology* articles – Korac (2003) and Wray (2003) – in greater detail. In the first, the introductory paragraph sets out its goal: to '*examine the situation of*' those '*experiencing*' a widespread and undesirable social phenomenon of sociological interest. It '*aims to stimulate a debate on interventions useful for*' dealing with the social phenomenon in question (Korac, 2003: 51). Interviewing collected information from people in two different settings, but explicitly did not attempt to cover the whole range of difference believed to exist:

[T]he research strategies to reveal the subjective world of the actor's experience are considered more appropriate for gaining knowledge about problems [of this kind] than the social mapping of numerical data and statistical methods. (2003: 53)

In the quotation above, the author chooses to juxtapose 'social mapping' and 'the subjective world of the actor's experience' as means of exploring the research topic. It follows that the competing capacities of the two rival methods become important. The article sets itself up to do *better than* a social mapping approach (the latter presumably based on a census or probability sample), not least in demonstrating 'the variety of situations that [actors] encounter in their everyday life' (2003: 53).

However, the data from the interviews:

... do not claim to be representative of the situation of all [persons experiencing the social phenomenon]. They are, however, *demonstrative* of the complexity of the process of [ameliorating the social phenomenon] and of the problems of how to facilitate it. (2003: 54, original emphasis)

We understand this to be an argument by exceptionalism. The finding of variety among a small group serves the logical function of demonstrating the exceptional existence of circumstances that do not conform to general (governmental policy) assumptions. The general applicability of the policy is therefore impugned by its failure to include the exception within its ambit.

This is an inherently problematic exercise in two ways. First, while qualitative methods give a different and usually more detailed and complex characterization of any given situation, social mapping is likely to cover a larger number and wider range of these situations and therefore can potentially tap into variation as a whole. The choice of method should in fact turn on what *kinds* of variation one is seeking, and the *rate* at which these are anticipated to occur. Second, while the approach taken by the author may be logically valid, it is both methodologically and politically inadequate. As the interviewees are *not* representative of the categories from which they have been drawn (which is what the article explicitly says) how can we tell whether their circumstances apply only to those interviewed, or also to the thousands of other people *supposedly* in similar situations? Is the discovered variety of situation a tiny and therefore insignificant part of 'the problem', or much more important because it is widespread? Is it realistic to expect policies designed to handle tens of thousands of people to cope with every minor variation that may occur? In a harder form of this argument, those who are 'not representative' are 'unrepresentative', and hence a poor basis on which to claim variety of circumstances or to legislate. In common with most of the articles in Volume 37, generalization is both explicit and implicit in this article, though at no point is the methodological basis for these generalizations discussed. The article does directly criticize generalization in survey research (Korac, 2003: 53), although this seems to be introduced only as a specious justification for the use of qualitative methods.

As the research focuses on the 'complexity' of situations encountered in two (out of many possible) settings, it is a little surprising to discover that two-thirds of the interviews have been conducted in one setting, and one-third in the other. Assuming that twice as much effort went into collecting information from the bigger group, yielding twice as much raw data, it is possible that the attention of the researcher was diverted towards that group's situation. Indeed, this suspicion is partly substantiated by the structure of the article: apart from those sections dealing with both groups together, about one-third more space is allocated to specific coverage of the larger group. The article does not reflect on the possibility that having twice as many interviews with one group might plausibly have some connection with the overall *emphasis and interpretation* in the analysis of what the informants said in their interviews and thus on the generalizations drawn.

It is acknowledged that while the article does not comment on this, it does report some inconsistencies between the two groups. Members of the larger group are only half as likely to be over 40 years old, and four times less likely to have children. They are also more likely to have completed secondary and/or higher education, nearly twice as likely to be in paid employment, and 10 times more likely to have experienced the target social phenomenon over eight years ago, rather than more recently. While these differences are listed, there is no discussion of the possibility that age, education, parenthood (household type), employment, or timing of experience might plausibly have some connection

with what informants said in their interviews and indeed at least partly *be constitutive of the variety of situations* which so concern the author.

We can summarize the generalization issues exemplified by this article as follows:

- 1 If the intention is to demonstrate a general variety in circumstances, this can be done in a number of ways: examining two examples in detail is not axiomatically a superior method to examining more cases in less detail.
- 2 Where the intention is to show the limits of government policy, the use of examples is rhetorically powerful, but suffers the defect of not demonstrating the frequency and spread of the circumstances identified by the exemplification. Questions of *demonstrating* differences ('complexity of processes'; 'variety of situations' [Korac, 2003: 53]) are linked to how *representative* the examples of difference are. Why do we need 'to stimulate a debate' (2003: 51) unless there is an *extensive and widely shared* problem?
- 3 When emphasis is placed on the 'actors' point of view', then conclusions drawn from in-depth interviews are likely to be related to *who* (how many people, of what types, in which groups) has been interviewed. Without a broad range of evidence, generalizations cannot be sustained.
- 4 Where qualitative research is framed as a critique of national policy, it is the research itself which raises the question of generalization (and should address it), not those who subsequently become readers of the research.

The Structure of Argument in Qualitative Sociology

The second example of generalizing practice (Wray, 2003) is a persuasive and stimulating study of gender, age and health which also addresses national policy. It is based on 'in-depth semi-structured interviews and focus groups' with 170 women in several minority ethnic groups. Its strengths, however, have nothing to do with the larger number of interviewees, but rather with the comparatively greater caution in the conclusions it draws. The final section in the article combines summary statements about what respondents have *said*, with more general evaluative sentences about older women as a whole. The author uses the format of examples from fieldwork – such as 'most women noted how despite changes to the agility of their bodies ... Being in good health was linked to ...' (Wray, 2003: 525) – as a basis for making immediately juxtaposed general claims:

... the needs of older women should be recognized and prioritized by local councils and other funding agencies. If older women are to have a good quality of life they need to feel in control of their present and future. *This meant making their own decisions, being included in the wider social and political arena and continuing to have a role in society.* (p. 525, emphasis added)

This latter quotation shows first how the specificities of the study are re-interpreted as policy imperatives of general applicability, and then in the last italicized sentence, pulled back by use of the past tense to how the author understands the empirical statements that have been elicited from informants. The juxtaposition makes it clear from which part of the evidence the generalization is coming. On the other hand, the quotation also demonstrates how closely empirical specificities are linked to generalities, and how easy it is to move from one to the other without explicit discussion of whether generalization is justified on the basis of the data actually collected.

To be more precise, the author has collected data by interviews and focus groups, from nine different self-defined ethnic groups (p. 513). If we assume (because we are not told) that all of the nine ethnic groups were included in some numbers, it seems a reasonable *moderatum* position to take that the experiences of aging that the author discusses are *not confined to only a few* ethnic groups, but are widely encountered. To the extent that the experiences of physical activity and sense of self highlighted by the article were systematically reported by very many of the women (and this again is not reported), it is a reasonable *moderatum* position to suggest that *many other (un-interviewed) older women (elsewhere in Britain)* might also share the experience. The kinds of empowering physical activities reported *may* plausibly offer the benefits suggested, but the article would be stronger if the grounds for claiming plausibility were spelled out.

However, to claim that these *particular* activities have the *same* benefits for *all* older women in *every part* of Britain would be a riskier generalization. We might also ask whether the activities might benefit *younger* women as well, or indeed, older *men*. How much counter-evidence (older women exercising but not reporting benefits) would it take to impugn the researcher's conclusions?

This second example is by no means atypical in not addressing these points. Sociological conventions in research papers do not currently require such justifications. If the intention is to generalize from the situation of a small number of people (in this case 170) to local authority policy throughout Britain, then we would suggest that it is incumbent on researchers to confront the need to establish the grounds on which generalizing claims are being made, and the limits to the generalization. Equally, even when the research goal is less concerned with policy and simply addresses enhanced understanding, if researchers are to make extensive generalizing statements like 'A code of ethics exists among collectors' or 'an instrumental society ... tends to empty life of its richness' on the basis of eight interviews (Kjølørød, 2003: 469, 474), it is not unreasonable to ask on what grounds such statements can be justified.

We can summarize the generalization issues exemplified by this article as follows:

- I Generalization is more likely to be plausible if it is approached with caution, moderating the range of the generalizing conclusions. Too ambitious

a set of conclusions undermines the credibility of otherwise competent research.

- 2 While the distinction between evidence and conclusion needs to be clearly demarcated, so that the reader is enabled to distinguish between the two, generalizations are more credible if the exposition connects the generalization to the specifics of data that provide its foundation. This can be partially achieved by simple juxtaposition, but explicit linkages with the relevant sub-set of data are an even better way of sustaining the generalization. This is not simply an aspect of final reportage. Good research design helps to identify the data necessary for the kinds of generalizing conclusions that may be anticipated at the outset.
- 3 Where several elements are involved in the study, these need to be fully reported. Thus if there are several sub-groups among the informants, if there are a variety of activities that operationalize the focal concepts, or if outcomes may differ for wider population groups, the reader needs to have access to published information about similarities and difference in those sub-groups, activities and wider populations. This enables readers to evaluate the status of the researcher's generalizations.
- 4 Similarly, researchers should indicate their own assumptions about similarities and differences between the research site and the other ('receiving') sites, and between the research's informants and other actors. This process would not only assist the reader, but clarify in the researchers' own minds what it is that they are claiming.
- 5 Even when the quality of the data is good, and the evidence is consistent with the conclusions drawn, research design may not have eliminated alternative explanations or constraints on generalization. The constraints on generalization need to become a standard part of the analytical discussion.

How Moderatum is Moderatum?

The two illustrations of current qualitative research point towards three related elements that form the basis for potential *moderatum* generalization. Generalization needs *explicit* discussion in every study. Valid generalization cannot be assumed to be a natural outcome of fieldwork, but rather should be considered from the outset. That entails treating it as part of *research design*, deciding on the range (or more precisely the moderation) of the generalization to be attempted, and setting up the data collection exercise appropriately – not least taking care over the *selection of the sample* of people or sites to be researched. The extent of the grounds for generalization depends both on the characteristics of what is being studied and, crucially, on the similarities of the research site to the sites to which generalization is to be attempted.

Thus, to follow Mason's typology of sampling (1996: 92–3), the qualitative papers in Volume 37 of *Sociology* are best described as being based on small

selections of units which are acknowledged to be part of wider universes but not chosen primarily to represent them directly. Like the narrower case of interpretive methods discussed by Williams, there is no means of knowing, let alone mathematically calculating, the probability that what is found in these samples is repeated in their wider universes. Although as we noted earlier, their authors do draw such inferences, there are no formal grounds for so doing (Williams, 2000a: 216–7).

Since we cannot make statistical inferences and therefore calculate probabilistic estimates for a universe, we are constrained to reduce the ambition of our generalization. We can moderate our generalizations in five main ways:

- 1 The most obvious and important case is the question of to which/how many other settings might the findings of a research study apply? Researchers may wish to claim that their findings are widely applicable, or, more circumspectly, applicable only to certain limited types of sites or categories of person. 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.
- 2 A second kind of moderating recognizes the limitations of *time periods*. A current claim is likely to be more valid than a claim about future conditions: social change will ultimately invalidate most sorts of the claims made by sociologists. Sociology has a marked tendency to cling on to generalizations based on originally well-received studies, and to lose sight of how specific they are to particular points in time (Payne, 1996).
- 3 A third area of moderation is how accurately the research has characterized the study topic. An account, whether qualitative or in terms of estimates of proportions or rates or occurrences outside of the immediate research setting, may be very precise or much more approximate. Statistical generalization follows a logic of estimates falling between a higher and a lower figure, in a confidence interval (although publishing and professional conventions in British sociology mean that these are rarely reported). The generalization may claim high or lower levels of *precision of estimates*.
- 4 Alternatively, generalization may be moderated by limiting claims to basic patterns, or tendencies, so that other studies are likely to find something similar but not identical. This fits more comfortably with the logic of methods using non-statistical generalization. It differs from precision of estimates because the details of findings are seen as contingent on the social processes in which they are embedded, and therefore the question for generalization is the limits of this *contingency*.
- 5 Finally, the nature of the generalization 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 relationships
- 7 Psychological dispositions / behaviour

The ordering and content of the list is not intended to be definitive, but hopefully it gives a sense of the variety of social characteristics that may be generalized. Moreover some characteristics are more likely to hold through time and space than others.

A simple hypothetical example serves to illustrate this point. An observation study of an international supermarket chain could make generalizing statements beyond the study site to supermarkets in other countries, concerning many (though not all) products and brands (type 1 in the taxonomy). Likewise, management structures would have common features across several countries, though there may be national variations that could prove exceptional (type 2). The physical layout of supermarkets in general, but the one studied in particular, would have general features that will wholly or partly determine how customers use them. One European chain, for example, accepts no credit cards (3). However, the individual location of a particular supermarket, or the advertising strategies in one country, may produce local variation (1, 2, 3). Certain symbols (primarily the supermarket logo) will be meaningful across cultures, though certain products will have meanings limited to certain locations (4): for example, 'Marmite' or 'Vegemite' are particular British/ Australian affectations! Group relationships, or relationships between two people will be less generalizable and will be the outcome of interactions between the wider culture of the supermarket, their social milieu outside the supermarket and psychological dispositions/behaviour (5, 6, 7). Often the critique of generalization in qualitative research rests on the variance argument in these latter phenomena, but it is easy to see that many of the social contexts (and of course some group/dyadic relationships) can be generalized beyond the specific context, at the very least as *testable statements*.

Access, Location and Logistics

This outline taxonomy demonstrates that generalization depends on what levels of social phenomena are being studied, and therefore that generalization is normally not a unitary process, because few studies operate at a single level. However, within each sub-type of the taxonomy, the same logic of comparison is at work: what are the similarities between the research setting and the wider universe which might justify extending findings specific to one place to others

which have not (yet) been researched? It follows that selection of research sites and informants, however difficult a choice, is crucial to the status of generalizing conclusions.

It is here that the two main traditions of methods most sharply differ. In quantitative methods, sample designs usually involve an explicit intention to scale up the findings, from a sub-set to the larger universe of people or institutions from which the sub-set has been taken (although we observe in passing that in practice, sampling/generalizing in quantitative sociology is not an entirely straightforward exercise [Payne and Payne, 2004]). In qualitative methods, the selection of 'samples' is usually not accompanied by such explicit intentions. More typically, the research site is selected for its convenience of access. The informants from whom data are therefore collected are chosen in an opportunistic or ad hoc way, even if care is taken with selection *within* the site.

Nonetheless, qualitative research does involve 'sampling'. *All* research entails sampling, because it is not humanly possible to study all cases at one time. The choice of 'sample' in qualitative research is often given little discussion in the methods sections of published papers, but common sense suggests that researchers do not deliberately seek out a sample which is atypical of the general type of social processes that they wish to study. Where there is discussion of the choice of site, it usually explains the practical constraints on the choice, implying that there are no grounds for believing that the choice led to the study of a seriously atypical setting. For example, Edwards and Imrie do not see their contribution to 'the development of disability theory' as constrained by the collection of data in Weymouth and Gateshead, chosen 'because of the willingness of local authority access officers to take part in the research' (2003: 245).

In other words, sampling is closely linked to 'access' to data. The term 'access' in discussions of research methods has normally been applied to the practical difficulties that researchers encounter with 'gatekeepers' and negotiating entry into a research setting. It has also been used in discussions about finding sites. Widely distributed social phenomena offer greater choice of where they may be investigated. A researcher located in a single place sensibly considers the practicalities of conducting investigations within travelling range of that base. For example, Crawford's (2003) study of *Manchester* ice hockey supporters is presumably not simply a personal foible, but also reflects their convenient location for his period at the University of *Salford*. Research reports sometimes explain their site selection in these terms, even down to the fortuitous availability of a potential site. Postgraduates are often directed to a research site because of their supervisors' contacts and knowledge of potentially cooperative organizations.

Thus 'location' implies not just simple physical location, but also social location in terms of networks and processes, which provide awareness of research issues, where they are most clearly manifested, and the means of establishing contacts with potential informants. Because of access and location,

when most sociologists carry out qualitative research, generalization comes low down the research process agenda. What effectively moderates generalization (whatever the intention of the researchers) is not the sampling per se, but the accidental outcome of access and associated logistical and resource decisions.

Practical constraints of location, funding and staffing drive the research process as much as the philosophical foundations of social science methodology. Qualitative methods have the advantage of offering compact, manageable units of research. In calling for more considered generalization, we are not asking for larger, quasi-statistical samples. From a logistical stance, qualitative research is unlikely to be able to handle this, even if it wanted to, because it could not cope with the scale of fieldwork needed. It is one thing to carry out four focus groups with 30 people in two locations (Edwards and Imrie, 2003: 245): it would be a much more daunting task to carry out, say 40 focus groups with 300 people (or 400 focus groups with 3000 people). Equally, eight life history interviews about stamp collecting and its like may be interesting (Kjølørød, 2003: 475): 800 detailed interviews about stamp collecting would, for most of us as researchers or as readers, be less attractive. The solution does not lie in that direction. This does not mean, however, that some key characteristics of a site might not be compared with other sites to check for potential generalization.

Nor does it mean that a random sample of people should never be studied using qualitative methods, nor a small group asked to complete a questionnaire. For instance, focus groups can still cover a range of representative settings and handle quite large numbers (e.g. 454 informants: Reay and Lucey, 2003: 126–7). Audiotaped semi-structured interviewing can be used to collect data from a set of respondents randomly sampled from the electoral roll (e.g. Savage et al., 2001: 878). These are, however, relatively rare cases: questions of sampling and generalization have been conflated with questions of data collection. If we wish to generalize – as most sociologists, we would argue, implicitly do – then there is no *logical* reason why we should only use quantitative methods, provided we take on board the idea of statistical sampling. On the other hand, the practical *logistics* of research restrict both the scale of data collection in qualitative methods, and the moderatum generalization that can flow from it.

Some Concluding Observations

Despite previous claims that generalization can be left to consumers, provided producers engage in ‘thick description’, generalization in qualitative research remains a major problem. Reviewing of the 2003 volume of *Sociology* has shown that sociologists, using a variety of qualitative methods, have in fact actively engaged in generalization, but have done so without discussing the reach of their conclusions. Generalization in qualitative research addresses two audiences. The audience may be the academy, whose members evaluate research

studies before admitting them to the canon, as asserted by Lincoln and Guba (1985): generalization decisions can be left to the reader. Our argument is that sociologists are currently not sufficiently sensitized to generalization and therefore, while authors continue to make sweeping claims, the academy is failing to grasp the limitations of its literature. The present article is an exercise in raising awareness of this problem.

The second audience is policy-makers, so that findings to some extent serve a rhetorical function of dramatizing an issue, rather than establishing true generality. We are unhappy about this characterization of sociological research, but would nevertheless argue that more considered, and therefore more justified, generalization actually strengthens such rhetoric, and so may more effectively challenge policy. Generalization remains a problematic issue.

Our claim for this constitutes a moderatum generalization, primarily based on the 2003 volume of *Sociology*. It is in the nature of moderatum generalizations that they are testable against data and subject to alternative reasoning. Analysis of another year, or another journal, *might* produce different results (for example, not every other qualitative paper might make generalizations) but we can identify no reason to believe this would be true *to the extent that it would impugn our conclusions*. Our data are particularly striking because they systematically show *every single article* conforms to our strictures. We can further support our moderatum generalization on the basis of our reading of the general literature. As we noted above about non-empirical articles, it is conventionally accepted that a complete documentation of tendencies or lacunae in a field of study is not required before presenting a case for innovation. Although we are taking advantage of this convention, it is clearly a less than satisfactory one.

In using a moderated generalization to advocate moderatum generalization, we focus on whether 'sending' and 'receiving' sites have at least some basic similarities. This involves identifying constraining features of one's planned site which differentiate it from other receiving sites, and moderating one's generalizations accordingly. The identification of varying degrees of similarity and difference provides a mental map of the sites to which generalization can, and cannot, be extended. This identification of similarities will often be hypothetical in nature, rather than empirically based. We accept that this is complicated by logistical limitations: opportunistic site selection will normally be incompatible with even moderatum generalization.

In other cases, moderating generalization does not reduce the importance of internal validity, but rather adds an emphasis to external validity. Generalization should always entail ensuring that alternative explanations for associations and patterns can be discredited by the study. Unless there is a rigorous logical structure to the analysis, combined with careful operationalization, interpretations will be open to challenge. This is a feature of both qualitative and quantitative sociologies, but in the latter the emphasis on pre-fieldwork development of research measurements and theorization has meant that the problem has been less than in qualitative approaches.

By starting with more reflection on the end point of the research process, researchers should be better placed to adapt their research designs, so avoiding excessive generalizing claims, and engaging actively with expressing their more modest claims in clearer terms. This means moving away from dependence on vague terminology or appeals to other scholars' (non-moderated) prior statements. In future, we would welcome authors including in their published articles short accounts of what they see as the extent of their moderatum generalizations, and the grounds for their claims. We believe this would provide the framework for greater clarity of position – are we being told about the social actions of 60 people, or about the social actions of a wider range of people? A case can be made for the former, although it is not one that we believe offers the best way forward for sociology, let alone those studies which wish to address governmental policy. Indeed, our reading of current practice leads us to believe that most qualitative sociologists engage, consciously or unconsciously, in generalization, and so are ultimately more interested in the wider range of people. The solution to the problems this lies in integrating moderatum generalization into mainstream research practice.

Notes

- 1 The logic of generalization in experimental research (specifically in social research Randomized Control Trials [RCTs]) is a little different depending on random assignment to experimental and control groups. If, after treatment, differences are observed then generalizations to a wider population can be claimed in respect of the treatment, though such claims are subject to certain assumptions holding (Orr, 1999: 46–50).
- 2 Any generalizing statement will take the same logical form of $P \rightarrow Q$, though strength of association and characterization of the conditionals will vary in different kinds of generalization. See for example Boudon (1974).

References

- Boudon, R. (1974) *The Logic of Sociological Explanation*. Harmondsworth: Penguin.
- Crawford, G. (2003) 'The Career of the Sport Supporter: The Case of the Manchester Storm', *Sociology* 37(2): 219–37.
- Denzin, N. (1983) 'Interpretive Interactionism', in G. Morgan (ed.) *Beyond Method*, pp. 129–46. Beverley Hills, CA: Sage.
- Denzin, N. and Y. Lincoln (1995) 'Transforming Qualitative Research Methods. Is it a Revolution?', *Journal of Contemporary Ethnography* 23(3): 349–58.
- Edwards, C. and R. Imrie (2003) 'Disability and Bodies as Bearers of Value', *Sociology* 37(2): 239–56.
- Geertz, C. (1973) *The Interpretation of Cultures*. New York: Basic Books
- Gladney, A.P., C. Ayars, W.C. Taylor, P. Liehr and J.C. Meininger (2003) 'Consistency of Findings Produced by Two Multidisciplinary Research Teams', *Sociology* 37(2): 297–313.

- Hacking, I. (2001) *An Introduction to Probability and Inductive Logic*. Cambridge: Cambridge University Press.
- Hammersley, M. (1992) *What's Wrong with Ethnography?* London: Routledge.
- Hislop, J. and S. Arber (2003) 'Sleepers Wake! The Gendered Nature of Sleep Disruption among Mid-life Women', *Sociology* 37(4): 695–711.
- Kelly, L. (2003) 'Bosnian Refugees in Britain: Questioning Community', *Sociology* 37(1): 35–49.
- Kjølsrød, L. (2003) 'Adventure Revisited', *Sociology* 37(3): 459–76.
- Korac, M. (2003) 'Integration and How We Facilitate It', *Sociology* 37(1): 51–68.
- LeCompte, M. and J. Goetz (1982) 'Problems of Reliability and Validity in Ethnographic Research', *Review of Educational Research* 54(1): 31–60.
- Lincoln, Y. and E. Guba (1985) *Naturalistic Inquiry*. London: Sage.
- Marshall, C. and G. Rossman (1989) *Designing Qualitative Research*, 2nd edn. London: Sage.
- Mason, J. (1996) *Qualitative Researching*. Thousand Oaks, CA: Sage.
- Mitchell, J. (1983) 'Case and Situational Analysis', *Sociological Review* 31(2): 187–211.
- Orr, L. (1999) *Social Experiments*. Thousand Oaks, CA: Sage.
- Payne, G. (1996) 'Imagining the Community', in S. Lyon and J. Busfield (eds) *Methodological Imaginations*, pp. 17–33. Basingstoke: Macmillan.
- Payne, G. and J. Payne (2004) *Key Concepts in Social Research*. London: Sage.
- Payne, G., M. Williams and S. Chamberlain (2004) 'Methodological Pluralism in British Sociology', *Sociology* 38(1): 153–63.
- Pearson, J. and K. Soothill (2003) 'Using an Old Search Engine: The Value of *The Times Index*', *Sociology* 37(4): 781–90.
- Punch, S. (2003) 'Childhood in the Majority World: Miniature Adults or Tribal Children?', *Sociology* 37(2): 277–95.
- Reay, D. and H. Lucey (2003) 'The Limits of "Choice": Children and Inner City Schooling', *Sociology* 37(1): 121–42.
- Savage, M., G. Bagnall and B. Longhurst (2001) 'Ordinary, Ambivalent and Defensive: Class Identities in the Northwest of England', *Sociology* 35(4): 875–92.
- Seale, C. (1999) *The Quality of Qualitative Research*. London: Sage.
- Silverman, D. (1993) *Interpreting Qualitative Data*. London: Sage.
- Spencer, L., J. Ritchie, J. Lewis and L. Dillon (2003) *Quality in Qualitative Evaluation: A Framework for Assessing Research Evidence*. London: Government Chief Social Researcher's Office, Cabinet Office.
- Williams, M. (2000a) 'Interpretivism and Generalization', *Sociology* 34(2): 209–24.
- Williams, M. (2000b) *Science and Social Science: An Introduction*. London: Routledge.
- Williams, M. (2001) 'The Status of Generalization in Interpretive Research', in T. May (ed.) *Qualitative Research: An International Guide to Issues in Practice*, pp. 125–43. London: Sage.
- Wray, S. (2003) 'Women Growing Older', *Sociology* 37(3): 511–28.
- Yin, R. (1984) *Case Study Research: Design and Methods*. Beverly Hills, CA: Sage.

Appendix

Examples of generalization in qualitative research, *Sociology* Volume 37, 2003.

<i>Authors' names</i>	<i>Main title of article</i>	<i>Page nos.</i>
<i>Issue No. 1</i>		
Kelly, L.	Bosnian refugees in Britain	37, 46
Korac, M.	Integration and How We Facilitate It	61–63
Reay, D. & Lucy, H.	The Limits of 'Choice'	122, 138–9
<i>Issue No. 2</i>		
Crawford, G.	The Career of the Sport Supporter	220, 233, 234
Edwards, C. & Imrie, R.	Disability and Bodies as Bearers of Value	239, 250
Gottfried, H.	Temp(t)ing Bodies	257, 271
Punch, S.	Childhoods in the Majority Worlds	287, 289, 290
Gladney, A. et al.	Consistency of Findings	311
Speer, S. & Hutchby, I.	From Ethics to Analytics	315, 333–4
<i>Issue No. 3</i>		
Kjølsrød, L.	Adventure Revisited	467, 474
Butt, T. & Langridge, D.	The Construction of Self	488, 489
Wray, S.	Women Growing Older	524–525
<i>Issue No. 4</i>		
Waites, M.	Equality at Last?	650, 651
Hislop, J. & Arber, S.	Sleepers Wake!	695, 709
Mallinson, S. et al.	Historical Data in Health Inequalities Research	771, 773–4
Pearson, J. & Soothill, K.	Using an Old Search Engine	781, 788

Geoff Payne

Is Emeritus Professor of Sociology, University of Plymouth, and Visiting Professor of Sociology at Glasgow Caledonian University. His research on class, mobility and social inequalities has used both qualitative and quantitative methods. Recent books include *Key Problems in Social Research* (Sage, 2004, with Judy Payne) and *Social Divisions* (Palgrave, 2005, 2nd edn).

Address: School of Sociology, Politics and Law, University of Plymouth, Plymouth PL4 8AA, UK.

Malcolm Williams

Is Reader in Sociology in the School of Sociology, Politics and Law, University of Plymouth. His most recent books are *Making Sense of Social Research* (Sage, 2003) and

Science and Social Science (Routledge, 2000). His methodological interests include representation, objectivity and probability, and empirical research interests include urban–rural migration and housing need. He has published in each of these areas.

Address: School of Sociology, Politics and Law, University of Plymouth, Plymouth PL4 8AA, UK.

E-mail: M.Williams@plymouth.ac.uk