J. Macaulay
January, 1979

Some Barriers to Drawing Conclusions from Social Science Research[1]

1. The Problem of Snapshot Perspectives. Some of the most enlightening (and certainly most interesting) descriptions of poverty are descriptions of "soft" or "ethnographic" data. These are data that may have been gathered systematically, but which are reported in an anecdotal or narrative manner. This kind of research report is invaluable in that "hard" (survey, demographic, experimental, etc. data often cannot supply the whole picture of a situation or even a readily comprehensible piece of it. Kriesberg (1970) notes that in the study of social networks, for example, ethnographic and survey data lead to different conclusions. On one hand, one might surmise that survey data are misleading because so many important aspects of a situation go unreported when using a prearranged set of questions. On the other hand, survey data sometimes show data reported in a literary fashion to be highly unrepresentative. It is the latter danger that faces the reviewer or synthesizer of social science research. The problem may not be intrinsic to the reports themselves but rather may stem from the natural tendency to generalize from a few cases or a specific situation to a whole population. An ethnographic report may be accurate, subtle and insightful but the cases covered may represent atypical rather than typical situations. See for example, Oscar Lewis, 1966; Liebow, 1967. The most highly visible, vivid, poignant or arresting examples of life in poverty are probably exceptions rather than the rule. The use of poverty vignettes and snapshots properly lies in exemplification, not proof. They are valid examples of what can come about but not reliable bases for generalizations about an entire group.

[1]This list is taken from a critical paper on the literature on poverty and related problems.

2. Sample Bias and the Loss of Qualifying Tags in Secondary Reporting. The problem of "soft" data just described is a subtype of the larger problem of sample bias. Much research is necessarily done with a specific group which has a unique set of characteristics. Sample limitations don't necessarily mean that the research is useless — the results may, in fact, be highly significant for a whole range of groups and reliable, valid generalizations. The main problem arises when, in the course of synthesis and summarizing, the qualifying tags which identified the sample get lost or the limits to generalizability are forgotten. We then tend to find such developments as noncomparable samples being compared to prove a point, or subgroup characteristics being taken as belonging to a whole population, and so on. For example, the very significant results from the first negative income tax experiment in New Jersey and Pennsylvania concern male heads of mostly-intact, poor, urban families. Such families form a rather small subgroup among those at risk of joining the welfare rolls. The experimental results yield probably valid and reliable conclusions about the effect of a negative income tax on this group and some good guesses as to what the effects would be on the entire population of poor people who might someday be covered by a universal negative tax scheme. However, inevitably it seems, it has been widely reported that this experiment clearly and definitely showed much more — e.g., that a guaranteed income would not cause "the poor" to stop working or to stay on welfare and not seek gainful employment-without incentives to do so.

3. Bias in the Definition of the Population from Which a Sample is Drawn. A more pervasive and subtle problem is not bias in sampling and subsequent lapses of memory for sample limitations but bias in definition of what the sample is a sample of. The limitations of some samples may escape the notice of a research consumer, if not spelled out carefully, and even of the researchers themselves. For example, "the poor" consist of people whose economic problems are caused by factors that differ widely across individuals, groups, nations, cultures and time. Some are poor because they're disabled and some are poor because they're

not white; some are poor because times are hard and some are poor because they work at a job that doesn't pay much; and some are undoubtedly poor because they just have had bad luck while others are poor because they're basically lazy. Besides making it hard on the reviewer who is looking for information on only one type of poor person, this identity of all types of poor people — and particularly the merging of various types into an amorphous "lower class" or the categorization of, say, all minority persons as poor — tends to obscure some very relevant causal factors in certain problem situations.

A study by Polansky, Borgmann, and DeSaux (1972) will serve here not only as an example of this unannounced kind of sample bias, but also as an example of the way careless conclusion-drawing could warp policy. This study concerns poor rural women who were neglectful mothers. The report began with a warm appreciation of the many poverty-stricken mothers in the original sample who were good mothers in spite of hard times. Only a "handful" of neglectful mothers were found in the original sample. However, the authors ignore this detail and on the same page go on to ask, "What causes marginal child caring and neglect? [It] is doubtless a phenomenon of poverty" (p. 6), and never discuss the good mothers again. Actually, they were studying a sample of neglectful mothers who were in rural poverty. They have no evidence that the neglectful mothers would have been better if they had not been poor or did not live in a rural area, and on the whole their evidence suggests that child neglect is not a common phenomenon of poverty.

It may not always be possible to fully specify the universe to which a sample belongs. It is certainly more difficult to identify important dimensions of poverty than it is to identify related socioeconomic or racial factors. However, when research results are such as to have the sort of effect that these might have (e.g., promoting the idea that poverty is likely to lead to bad mothering and thus mothers in poverty need therapy or, if resistant to therapy, should have their children removed), the effort to careful fully define samples seems essential.

4. Incomplete definition of the situation in which research is done. A problem parallel to the preceding one is that of unjustified generalization from circumscribed research situations to large classes of situations. Much of the research on the "Psychology of poverty," particularly that with children, is carried out in "experimental" situations that are said to be "controlled. These situations may be relatively invariant from one participant to the next (but see Barber, *1976,* on this point); the problem is whether or not they resemble situations to which generalizations are made and whether participants' behavior is validly labeled. For example, "ability to defer gratification" and future versus present orientation are "traits" much studied by developmental psychologists in laboratory situations. It is not yet known, however, whether the behavior studied or the results observed in fact are relevant to the probability that the children involved will become adults with a drive to escape poverty, achieve, etc. The research is perhaps better characterized *as involving* strange behavior of children in strange Situations with strange adults," as Bronfenbrenner *(1974)* put it. It is research, Bronfenbrenner now believes, which is "ecologically invalid," research which gives only a partial picture of the children being studied and perhaps no insight whatsoever into the environment in which the behavior of interest takes place. This problem is a far reaching one, not only in developmental psychology but in many major areas of interests to social psychologists. This and related problems, in fact, *constitute one of* the major foci of the widespread self-analysis the field of social psychology is undergoing. (See, for example, any issue of Personality and Social Psychology Bulletin.)

5. The dataless cell. A recurrent problem encountered in the literature on poverty has been that of discovering whether the data show something distinctive about the poor as compared to the monomer. This is the other face of the problem raised with the study by Polansky et al. above. They did not have a sample *of* mothers within which they could make class comparisons. (The point made above was that they ended up studying a sample that was

not of poor mothers but of poor mothers with particular characteristics. But they assert that "child neglect is doubtless a phenomenon of poverty. It is not widespread among those who are better off . . ." (p. 6). They cite one study as evidence for this proposition which, on examination, turns out to be conjecture by a participant-observer of poverty. Later in the book they present evidence that AFDC mothers score as high on a childcare scale as self-supporting mothers. It appears in the end that they could have written the same book if they had used a sample of middle class mothers.

Nonexistent comparison figures are cited so often in the poverty literature and believed in so strongly that when researchers find disconfirming data they then conclude that they have aberrant results. It is probably safest to assume that comparable figures for poor and nonpoor groups do not exist unless specifically cited and that if data were collected, they would not show poor - nonpoor differences (See Blum and Rossi, 1968, and Ziegler, 1971, for similar conclusions.)

6. Invalid inferences of causation. A related pair of points concerns the value of prospective studies over retrospective studies and the confusion of correlation with causation. If one takes a group of individuals with a problem and finds factors common to the whole group, one has then only a viable hypothesis concerning causation. The problem is to discover if the group under study actually is a random sample of all individuals characterized by the supposed causative factor — e.g., a prospective study should follow the retrospective study. Polansky et al., for example, should study a representative sample of mothers of all classes to see if the factors they think cause child neglect precede and are predictive of neglectful mothering. Generally, it is my observation that if prospective studies of the supposed pathologies of poverty are undertaken, they tend to end up as debunking studies. I suspect that this may be due to a tendency among social scientists to underestimate human potential for self healing and adaptability.

A parallel problem is that of drawing inferences about problem causes (and cures) from correlation statistics. For example, economists construct elaborate models based on regression statistics which produce fine-tuned and probably valuable results, but they untenable assumptions and conclusions about causation. For example, Honig's (1974) study showed that the size of the average AFDC stipend, the proportion of female-headed families, and AFDC rates were positively related across a sample of urban areas. Honig assumed that this showed "welfare-induced" family splitting, but presented no evidence to show that stipend size was a factor antecedent to family splitting and going on welfare. It is very possible that the observed relationships are "spurious"-e.g., simultaneously influenced by social and economic factors which were not entered into the analysis. (Much of the evidence that suggests the latter possibility is probably unknown to economist Honig since it. was not published in economic journals, a problem that will be described more fully below.)

7. Bias in the outside world. Researchers are sometimes forced to work with social statistics that carry within them the biases of society. For example, possibilities of class, race and sex bias in statistics on juvenile delinquency include (1) bias in the tendency to notice and label nonapproved behavior as delinquency, (2) bias in the reporting of delinquency to authorities, (3) bias in the disposal of cases-that are reported, and (4) bias in judgment rendered should a case go to court. Given all these factors, it is rare that one can safely draw firm conclusions from social statistics about the varying incidence of some phenomenon in different places, periods, and groups. Factors of bias which also vary across places, periods, and groups will often be equally valid candidates for explanation of observed differences. (See, for example, Campbell and Ross, 1968.)

8. Similarities, differences, and what constitutes "significance." In addition to making sure alleged differences really exist, there is the problem of deciding how much difference

-constitutes an important difference. For example, in the New Jersey-Pennsylvania negative income tax experiment a small minority of wives quit their jobs. There were significantly more of these in the guaranteed income group than in the control group. Is this significant in more than the statistical sense? Probably not. In the first year of the study 14% of the experimental group wives were labor force participants versus 16.3% of the control group. Differences in the second and third years are 1.3 and 0.5 percentage points, respectively. Furthermore, the sample is not a good one of working wives. Intact families with a working wife are usually too far above the poverty line to have fallen into the study sample. Sample families have a far lower -rate of working wives than -the population at large. Thus. we have a sample of an atypical group of working wives whose incomes do not lift their families out of the poverty class. Even if the observed differences were much larger, they would be of questionable value to economists who sought firm conclusions about what wives generally would do if family income were guaranteed at a specific level. (Nonetheless, economists have drawn what seem to be firm conclusions from this result; see Barth,, Carcagno and Palmer, 1974, p. 214; Mahoney and Mahoney, 1975, pp. 189-190; Watts and Rees, 1975, pp. 83-84.)

A related problem is that of stretching small differences to dichotomously characterize whole classes. For example, perhaps it is significant in terms of service and protection needs that Black neighborhoods are more victimized by criminals, delinquents, and drug addicts than white neighborhoods, but this should not overshadow the equally significant fact that the great majority of Blacks are not characterized by "deviance." However, such distributional differences in important factors often lead to unjustified labeling in the poverty literature. The best known example is to be found in the Moynihan view of the Black family published in 1965 as an official government document. As Justice Thurgood Marshall has pointed out (in his dissent in N. Y. State Department of Social Services vs. Dublino, 93 S. Ct. 2507, 1973), such labeling habits and the myths they support tend to have effects on policy, with regrettable results.

Another version of this problem is that of using proportions that obscure the absolute size of a problem or of a blessing. Proportionately, for example, the upper class produces more geniuses than the working class. But given the difference in absolute size between the upper and working classes, it is also true that the greatest number of geniuses come from the working class. That fact might be very relevant to, say, a policy maker interested in seeing that the country increases its number of active geniuses.

9. Problems of inadequate theories, concepts and methods. It goes without saying that theory and method are always inadequate and probably always will be, given the human and cultural limitations of social scientists. However, there are some specific problems which could be solved without touching the eternal need for consideration of more variables, more complex models, and better methodology. Among these problems are those of defining and shaping the variables one is going to work with, such as the following:

False dichotomies. Dichotomization of variables is sometimes unavoidable in research. but it is often the case that dichotomies are poor mirrors of the real world. The poor/nonpoor dichotomy stands as a prime example of false dichotomies, as the continuing debate on where to draw the poverty line demonstrates

Spurious symmetry. Herzog (1970) uses an old Yiddish proverb that goes, "Money is not so good as lack of money is bad" to illustrate a problem she terms "spurious symmetry This often involves the coalescence of value continua with behavioral trait continua. For example, if "dependency" among the poor is bad, independence must be good. But this creates some problems. For example, take the Appalachian mountain man who accepts welfare but otherwise seems independent of the larger society; is he admirable or not? The illogical

solution in the poverty literature is to redefine his independence (as "familism," or excessive loyalty to his family) rather than to rethink the concept.

Dubious continua. Dependency/independence can also be used as an example of a dubious continuum. These terms have commonly been used as end points for a continuum in the poverty literature, but some psychologists who perhaps have been working with the concept of dependency longer than poverty researchers — have come to the conclusion that independence is not a fixed trait that readily stays put at one end, with dependency at the other. Rather,, these are constructions put upon complex person-to-person or person-to-environment relationships, constructions which change as one's vantage point changes.

Absent alternatives. Another problem continuum is to be found in questions like, "When did you stop beating your wife?" The parallel in the poverty literature is "What's wrong with the poor that they're poor?" Although the Bible has insisted all along that poverty and goodness go hand in hand, poverty researchers have seldom considered this possibility, at least until minority sociologists came along. (See for example, Hill, The Strengths of the Black Family, 1971 Such continua tend to limit our understanding For example, child neglect has been measured by a scale that ends at zero and that does not extend into the realms of love and nurturance. Thus we might learn how much child neglect is seen among the poor but not how much good parenting there is. The absence of an alternative to neglect implies that good parents are rarely found among the poor which is not the case. Attitude scales also sometimes reflect a misguided conventional wisdom; for example some scales make it appear that one could never love the mentally ill, junkies and other "deviants, nor scorn material goods and worldly achievements, and so on.

Pejorative labels and deficit hypotheses. Reiss *(1975)* has pointed out that some social scientists show the same inappropriate labeling habits that they accuse other professionals of. Reiss felt that the problem was compounded by sociologists' odd preference for psychological over sociological explanations. Ryan *(1971)* has given a book-length description of these