ANALYZING INTERVIEW DATA FOR THE HISTORY OF SCIENCE

Elihu M. Gerson
Tremont Research Institute
458 Twenty-Ninth Street
San Francisco, CA 94131
415-285-7837   gerson@ieee.org

12 November 1998

Prepared for a conference on "Interviews in Writing the History of Recent Science" held by the Immunology Project, Stanford University Program in the History of Science, Palo Alto, California, 28 - 30 April 1994. I am grateful to Anselm Strauss, Howard S. Becker, and William C. Wimsatt for instruction and discussions over many years of the ideas presented here. I am also grateful to M. Sue Gerson, William Glen, James R. Griesemer, Jane Maienschein, and Jan Sapp for comments and suggestions, and to M. Sue Gerson for continuing support.

INTRODUCTION

In the light of traditional stereotypes about historical scholarship, analyzing interview data as part of a historical research project seems to be a strange thing to do. History is about the past, and we expect the people involved to be dead and gone; so how can we interview them? However, historians of twentieth century science have frequently conducted interviews with their subjects. Historians have also begun to collect information on events as they happen, recognizing that interviewing provides an invaluable supplement to the documents. This is particularly true since many kinds of information, e.g., the kinds of personal details given in correspondence and diaries, are typically not available during respondents' lifetimes, or, sometimes, for many years after their deaths. But analyzing interviews is not an obvious or trivial task, and so we need some discussion of the methodological issues involved.

Why analyze?

A long historiographic tradition emphasizes description as the primary, if the not the sole, legitimate kind of historical activity. But there is no way to describe a situation without theoretical ideas or preconceptions about what is going on. There are always differences in perspective which give different weights to the same facts, and which select different facts as worthy of attention, even in those rare situations where the questions asked are simple, clear, unequivocal and uncontroversial. So between the first conversations with respondents and the completion of a final draft there is a great deal of analysis and theorizing. The point then, is to do this analysis as well as possible. This means, on the one hand, minimizing the inevitable distortions and omissions (or at least, their negative effects) as much as possible. On the other hand, it means noticing and exploiting opportunities for new insights, discoveries and interpretations.

Both these ends are best served if the analysis is systematic and rigorous. A study can hardly be conducted without making many decisions about how data are to be collected, recognized, and interpreted. Surely it would be better if these decisions be made in a reasoned way. Yet historians have been curiously reluctant to develop a body of codified practices for ensuring detection of error on the one hand, and facilitating new insights on the other. This is not to say that historical research lacks scholarly techniques– rather to the contrary. From paleography to multivariate statistical analysis, historical research has a full complement of effective data-handling procedures. The lack is in the development and teaching of systematic methods for using these techniques; that is, matters of study design and the logic of historical analysis. These latter issues are often dismissed as topics for the philosophy of history, a subject taught in philosophy departments by people who generally do not do historical research. Recently, some historians have begun to urge systematic consideration of their methodological problems (e.g., Karsten and Modell 1992). Here, I concentrate on one aspect of these problems, procedures for analyzing data systematically.

People often think of analysis as something which comes after the interviewing: first we get the facts, then we analyze them in order to decide what they mean, then we write up the results. But in practice, researchers analyze a great deal before and during interviews as well. The questions they ask often depend upon the responses they've gotten in earlier interviews. Decisions about whom to interview and how to interview them often depend on information which turns up in the course of talking to people or working through archives. Hence, researchers need some analysis in order to get the most out of the interviews.

There are many ways to analyze interview data. Which ones we adopt depends in part on the purposes of the study, the kinds of knowledge already available, and the kind of results wanted. It is worth noting some of important characteristics of interviews for scholarly history of science projects. Projects whose interviews do not have these characteristics may have different data collection and analysis problems.

First, history of science projects are typically unstructured (or loosely structured). That is, historians rarely have a fully developed questionnaire with a fixed series of questions. Instead, there is a general topic list which may vary from respondent to respondent. The loose structure means that historians are usually willing to follow up new leads and new topics which arise in the course of the interviews.

Second, interviewers on history of science projects usually have some technical knowledge of the field they're studying. Often, interviewers' knowledge of the history of a field far exceeds the working scientists'. Interviewers are thus capable of asking cogent questions about the technical content and implications of the research they are studying. Indeed, their historical depth often means that historians can frame questions in ways unfamiliar to working scientists. On the other hand, historians very rarely have the same command of current knowledge in a field as the working scientists in it, even when they are trained in that field. They are thus always at some disadvantage vis-a-vis the people and projects they study.

Finally, history of science projects are usually basic research projects rather than "applied" projects such as public education, policy formation, or criminal investigation. The primary audiences for these studies are historians and other scholars.

The discovery approach

Before getting into the particulars of data analysis, I want to say a few words about the approach I'm taking in this paper. It is called the discovery approach, because it is a method of systematic discovery developed by Anselm Strauss, Barney Glaser, and their associates (Glaser and Strauss 1967, Strauss 1987, Strauss and Corbin 1990, Gerson 1991). My version also relies very heavily on the work of Howard S. Becker (Becker 1967, 1970, 1986, 1997). The approach is rooted in the interactionist school of sociology, a version of the Pragmatism developed by Wiiliam James, John Dewey, and George Herbert Mead. Of course, there are many other approaches to analysis available, and they overlap with this one in varying degrees.

The discovery approach is inductive and open-ended. It aims at discovery of new phenomena and construction of new concepts and theories rather than confirmation of hypotheses. The concepts and theories which result from the discovery approach are called grounded because they come from systematic analysis of the data rather than derivation from first principles. Data collection and analysis proceed in tandem, and depend upon one another. For example, choosing the next respondents to interview routinely depends upon the results of a previous analysis, as does the formulation of the questions asked and topics covered. This is one of the chief ways in which the discovery approach differs from other methods often used in the social sciences, such as sample surveys.

The discovery approach is based on the comparative method. Every observation is accompanied by questions about how things vary. Systematic and rigorous comparison means using guiding concepts or theories as an organizing framework. Some concepts are especially useful as starting points for research because they aid in organizing observations very effectively, while remaining imprecise enough to avoid overly constricting conclusions. I propose one such guiding concept below, but first I want to consider the broader strategy of research.

GETTING THE STORY STRAIGHT AS A KIND OF RESEARCH STRATEGY

The developers of the discovery approach have concentrated on methods for discovering generalities which apply across many times and places. But historians typically focus on a different kind of research problem, which I shall call “getting the story straight.” This means developing an adequate description of a particular event, series of events, or situation. Unlike many theory construction methods, the discovery approach can also be used to develop good descriptions systematically. Before I discuss procedures for doing so, it’s useful to say a few more words about getting the story straight.

Getting the story straight concentrates on description of particulars, in contrast to other kinds of research, which focus on developing new general principles or laws of nature (Goudge 1961, Hull 1975, 1981, 1992). Ideally, a good description is one which is complete, accurate, and unbiased. Of course, no actual description can reach this ideal. The problem then, is to find ways of making descriptions as complete, accurate and unbiased as possible– or at least, as is practical.

Historical research is more than just an account of particulars. Interpretation is needed to give coherence to the description. Adequate interpretation in turn, requires using various categories and theories, even if only tacitly. Indeed, interpretation requires multiple models or conceptual schemes. In order to make sense of events, even very ordinary events, we need many kinds of background knowledge, together with theories which relate different aspects of life. At the most trivial level, we must be able to recognize, for example, that someone’s being “born in poverty” is a description of social class and not of geography. But such observations also carry with them the supposition that class-of-birth is a significant thing to know– that fighting poverty while growing up has consequences for the development of lives in general, and scientific careers in particular. Suppositions of this kind are necessary if we are to make any sense of events. Hence, describing a particular situation adequately means, not *avoiding* the use of theory, but rather using *many* theories (each specialized and limited in scope) to understand the particular situation. Carl Becker, one of the most important American historians of the 20th century, made this point long ago:

> "[T]he historian knows well that no amount of testimony is ever permitted to establish as past reality a thing that cannot be found in present reality.... Tacitus is a good witness, and when he says the Germans do not inhabit cities, we believe him, though we do not know precisely what he means by cities.... If he had said a thousand times over that the Germans had wings, we should still say that the Germans had no wings.
>
> The classic expression of this truth is of course Hume's famous argument against miracles. That argument does not really prove that miracles never occurred in history; it proves only that there is no use having a past through which the intellect cannot freely range with a certain sense of security. If we cannot be on familiar terms with our past, it is no good. We must have a past that is the product of all the present. With sources that say it is not so, we will have nothing to do; better still, we will make them say it was so. The sources say– and it is a commonplace now that they say nothing more persistently, or with greater particularization of detail– that during the Middle Ages miracles were as common as lies. The modern historian admits that there were lies, but denies that there were miracles. He not only rejects the miracle,– the explanation of the fact,– he rejects the facts as well; he says that such facts are not proved; for him, there were no such facts. And he rejects these facts, not because they are contrary to every possible law of nature, to every possible experience, but simply because they are contrary to the comparatively few laws of nature which his generation is willing to regard as established." (C. L. Becker, 1910, at pp. 12 - 14 of the 1958 reprint)

The distinction between general principles and particular circumstances is an analytical one, and no research is ever devoted exclusively either to describing circumstances or to generalizing. Rather, all studies do both. Studies do differ in which kinds of knowledge, principles or particulars, are considered "background" to the problem at hand and which are at the focus of attention. That is, the differences between general principles research and particular circumstances research lie in how different research activities are related to one another.

In experimental physical science, for example, the particular circumstances are given in the experimental "lashup" which generates the data. Lashups can fail in many ways. For example, materials can become contaminated, or instruments may be miscalibrated. Hence, in order to be sure of what they dealing with, scientists are obliged to do the "natural history" of their own experiments, using many partial and specialized models to understand and eliminate sources of error, artifact, and bias. For example, experimenters must be able to assess the likelihood that an unusual strip chart recording is the result of accidentally jostling the recorder. Sorting out the consequences of many complex technical steps requires detailed knowledge of the procedures and equipment used in the experiment, including the peculiarities of the local situation. Knowing the detailed natural history of an experiment allows experimenters to control the results by eliminating unwanted effects, phenomena permitting. Once this occurs reliably, the particular circumstances which gave rise to the experimental results become part of that study's history. Scientists no longer care about the particular circumstances which gave rise to the results, and are indifferent to the fate of used test tubes, reagents, experimental animals and other materials consumed in the course of the work. The particulars of the study are divorced from the general principles which they helped to establish and then ignored.[1]

In historical research, by contrast, particular circumstances cannot be treated as "scaffolding" and removed at the conclusion of the study. Rather, particular circumstances are typically at the focus of attention. It is precisely those aspects of the situation that cannot be summarized by general principles which are of the greatest interest. Historians typically want to know how a situation or sequence of events differs from all the others, not how and why it is similar to others.

But recognizing differences depends upon the use of similarities. One cannot say that John and Mary are unlike one another without reference (possibly tacit) either to some third person (e.g., George) or to a more abstract standard of comparison (e.g., definitions of "man" and "woman."). Similarly, to say that George is an unusual man requires some sense of what men usually are, which is a kind of similarity among them.

In this situation, general principles serve as auxiliary devices used to explain aspects of the events of interest. Biographers, for example, are interested in their subjects' development from early childhood. However, unless there is something strikingly unusual about the subject, biographers don't discuss biological theories of human development in relating their subjects' histories. Instead, they take such matters for granted just because everyone's development is explained by such theories. Mentioning them does not add anything to the story of why a particular subject turned out the way he or she did.

Hence, theories "screen out" the aspects of a situation that can be explained with current knowledge. In doing so, they highlight and reveal the unique or unexplained aspects of

[1] Usually, scientists think about the process of making experimental results independent of the contexts which produce them as establishing the *reproducibility* of a procedure. The best discussion of this process is Collins 1985.

the situation. The emphasis is on what remains when the explanatory power of the theories has been exhausted.

In order to screen out the predictable, one must have good classifications and models for recognizing it. Such models are not always available; often, they must be constructed. The detailed flow of events examined by historians often belies the conventional classifications or popular theories which serve as a starting point for analysis. Getting the story straight thus means building or borrowing multiple specialized concepts and models and using them to choose and interpret the data of a more-or-less comprehensive picture of some particular circumstances. Each partial model contributes something to understanding the specific situation. Paradoxically then, historians must often construct or modify theories in order to see the unique aspects of the situations that concern them.

The historian's task thus seems like an inverted image of the experimental scientist's. For scientists, generalizations are the primary product of the resesarch, and they are used to encompass the unexplained particulars as much as possible, and thereby eliminate them as problematical. For historians, generalizations are background devices used to bring the unexplained particulars into high relief. Moreover, the unique *configuration* of particular circumstances is often of special importance. For this reason, generalizations are often implicit in historians' writings, and their analytical context is not developed. The implicit character of these theoretical commitments makes it very difficult for readers to evaluate them.

STRATEGIES FOR GETTING THE STORY STRAIGHT.

Let us turn our attention to the specifics of getting the story straight. My discussion here does not pretend to be a complete text. Rather, it is a summary of the procedures used in the discovery approach to construct classifications and models with data drawn from interviewing and field observation. Because the discovery approach depends upon the comparative method, it cannot be used to conduct studies for which the comparative method is unsuitable.

Specific techniques of analysis such as conducting interviews or reviewing and coding raw data are covered in standard textbooks of sociological field research methods, and don't need repetition here (cf e.g., Douglas 1985, Schatzman and Strauss 1973, Whyte 1984). I want to emphasize some of the more important strategic issues which arise from concentrating on getting the story straight, as they appear in the work of collecting and analyzing interview data. Perhaps the most important point is that analysis should begin as soon as data begin to accumulate. Waiting for notes and transcripts of interviews to pile up is poor strategy. Instead, we use the intermediate results of analysis to choose future respondents and shape the content of interviews. That aside, three strategic issues are especially important for getting the story straight: charting the audiences of the focal line of work, coding, and using the resulting classifications.

The notion of audience

Let us call the research we are studying the focal research and researchers. Like every activity, the focal research has audiences which review and evaluate the work, and make use of the focal research results as part of their own activities in turn. These audiences include the focal researchers themselves and their colleagues in the same specialty. They also include the full range of actors and activities with which the focal line of research interacts.

For example, consider the heredity research of T. H. Morgan and his students in the years around World War I. This work laid the foundations of classical *Drosophila* genetics, and became one of the centerpieces of 20th century biology. The audiences for this research consisted first of Morgan and his students themselves. Other scientists interested in research on inheritance also formed an important audience; these included W. E. Castle and his students at Harvard, R. A. Emerson and his students at Cornell, E. B. Babcock and R. E. Clausen at the University of California, and many others.[2] Other audiences included agricultural geneticists and breeders (Kimmelman 1983), the Carnegie Institute of Washington, which funded the research for many years, and the administrative staffs of Columbia University and the California Institue of Technology, where the research was conducted.

Audiences constrain one another by setting limits on each other's activities. By cooperating or refusing to cooperate with a line of work, each audience exerts pressure on that line to conform to its desires. One obvious example is the relationship between researchers and the sponsors that fund them. Sponsors often seek to steer the research toward some questions and approaches, and away from others. But other audiences exert demands as well. Hobbyists (e.g., bird watchers) often exert pressure on scientists to shape their research in convenient ways. Similarly, professional "clients" of basic researchers (e.g., agronomists, physicians and engineers) routinely make demands. An audience is *significant* if its cooperation (even if only passive) is necessary for the focal research to continue. For example, sponsors and colleagues

[2] This story has been told from many points of view; cf. G. Allen 1978 for a biography of T. H. Morgan, E. A. Carlson 1981 for a biography of Morgan's student H. G. Muller, R. E. Kohler 1994 for the use of Drosophila. Detailed participant histories are provided by Dunn 1965, Sturtevant 1965 and Carlson, 1965, 1974.

are significant audiences. But other audiences also exert a powerful shaping forces on the way work is done.

Each audience has audiences in its turn. These secondary audiences constrain the conduct of the primary audiences of the focal research. For example, Federal funding agencies such as the National Science Foundation are constrained by Congress, which encourages some programmatic emphases and discourages others. The relationships among lines of work and their audiences forms a dense and highly ramified web of mutually constraining and supporting relationships. Understanding the way a primary audience reacts to the focal research requires knowing something about the demands on it from secondary audiences. This is not to suggest that we map the entire web of connections between the focal research and the world at large; that would be impractical. But we do need an overall picture of the way the focal research fits into the larger scheme of things.

The notion of audiences gives us a convenient way to to construct this overview. Listing all the audiences to a line of research provides a skeleton list of the actors and tasks which shape the work in some fashion. Of course, some audiences are more important than others. An unreasonable amount of effort may be needed to identify each and every participant in a system of audiences. But having at least a skeleton picture guides "next steps" in the historical study, and provides a crude check on completeness.

We know enough about research to expect several kinds of audiences in almost any situation. These "standard" audiences include (beside the focal researchers themselves) collaborators, technicians and other support staff, families and friends, competitors in the same line of research, other colleagues in the same and other specialties, hobbyists, "customers" outside of science, (e.g., engineers, physicians), suppliers of instruments and materials, host organizations and their administrative offices, regulators, sponsors, the press, and the "general public." Each line of research does not necessarily have all these audiences. For example, protein chemistry research has few amateur practitioners. Each audience is not equally important to every line of research. But these audiences do appear quite often, and it is worth having a list of them to use as a first classification. Some specialties have additional specialized audiences. For example, paleontologists must often deal with collectors and commercial dealers in specimens.

The full list of audiences is developed by examining journals and other publication of interest to the focal line of research, and through the inteviewing process itself. Everyone a focal researcher interacts with is (or represents) an audience. The central tactic in building a list of audiences, is to discover who is on the opposite side of every kind of interaction in which the focal researchers participate. If a respondent teaches, for example, then his or her students are an audience. If respondents buy supplies or equipment, then the suppliers are an audience. Of course, it's especially important to identify the significant audiences, i.e., those that can stop the work by refusing to cooperate.

Coding

The basic analysis task, called coding, organizes observations into useful categories. The important thing in coding is to treat everything which occurs as data. It is difficult to overemphasize the importance of this point. Every aspect of events which touches on the problems or subject-matter of the study, no matter how trivial or mundane, becomes an observation to be interpreted. We can make discoveries reliably just to the extent that we become proficient at rendering observations problematical in order to analyze them.

Nothing about the situations we observe dictates the codes we use; a given event can always be coded in many ways. How we code depends on the research problem and the way we've conceptualized our study. Of course, this may change during the course of a project; we often find ourselves recoding the original data of a project in response to novel concepts which have grown out of the analysis.

We code by treating each observation as exemplifying some value of some variable, or some attribute of some property. Specifying *what* attribute of *which* property is the critical coding step. For example, in the early 1920's, Roy Chapman Andrews was planning an expedition to the Gobi dessert for the American Museum of Natural History (Andrews 1922). Andrews approached J. P. Morgan (a museum trustee) and asked him to help pay for the expedition. Morgan agreed to provide a substantial part of the budget. For coding purposes, the property might be "kind of sponsor" and the attribute is "private benefactor." Of course, this view of it is from Andrews' perspective; from Morgan's point of view, the matter is one of spending money on science rather than, say, art; or on philanthropy rather than something else.

Once we have a property and one of its attributes, the next step is to specify the other attributes which the same property has. For example, in what other ways (besides donations from private benefactors) do scientists raise funds to support their research? Such questions start a series of comparisons, in which we look at fund-raising activities conducted by different scientists under different circumstances. When an observation doesn't fit into the classification, we recognize a new attribute or a new property. In the case of the funding example, we soon find "private foundations" and "government agencies" as additional sources of funds.

A problem arises when an observation seems to exemplify more than one attribute of the same property. When this occurs, the attributes have to be specified more carefully in order to distinguish the overlapping characteristics which make them ambiguous. For example, Alexander Agassiz (1835 - 1910) was an invertebrate zoologist who directed the Museum of Comparative Zoology at Harvard (founded by his father, Louis). He was a wealthy man and a major benefactor of the Museum. After stepping down as director, he financed his own research at a private laboratory in Newport, R.I. (Winsor, 1991). Is Alexander Agassiz to be considered a private benefactor of his own research? Or is it better to have a new kind of sponsorship, self-sponsored research? All scientists sponsor their own research in some degree, so the category is ambiguous. What is the value of recognizing a separate attribute of "self-sponsored" apart from "No sponsor"? It is often possible to code a property in different ways. Choosing among codes depends on the analytic problems of the research. If our concern with Agassiz is exclusively with his role as a leading student of echinoderms at the end of the 19th century, then we can probably afford to ignore subtleties in the classification of sponsorship. On the other hand, if our concern is with Agassiz as one of many people who were searching for effective ways to match the demands of research and the larger society at the end of the 19th century, then varieties of sponsorship will take on considerable importance.

Another difficulty arises when we make use of a pre-existing classification, and find no observations for one of the categories in the classification. When this occurs, we must consider the possibility that the category is empty; i.e., that it is merely a logical possibility. Such categories can be eliminated by simply deleting them from the classification, or by collapsing them with another category. For example, suppose we start with a classification of sponsors which includes both "private benefactor" and "foundation," but find no instances of foundation support in our data. We might be tempted to collapse "benefactor" and "foundation" together as "private support." But this must be done cautiously, for it may be that we simply haven't run across any occasions of foundation support as yet. Or perhaps the category is anachronistic; foundations in the form we know them today did not appear until the turn of the 20th century (Jonas, 1989). In general then, we shouldn't collapse two categories together without a good reason for doing so. Collapsing two categories is tantamount to saying that they are "really" the same sort of thing, and that we should not make the distinction between them. But in order to do this, we must have some rationale for making the claim.

Properties should be as specific as possible. Analysis depends, after all, upon the capacity to distinguish among circumstances. Using very vague or general categories (e.g., "things needed to get the work done") can leave us with classifications into which everything fits, but from which nothing of interest emerges. Categories should thus be made as specific as possible as long as the data will support their recognition.

Using classifications

Coding builds classifications. We build or borrow many of these classifications during the course of a research project. Many of them are taken-for-granted routine arrangements, while others are developed for the first time in the course of the research. We use classifications to organize data, and to identify and remind ourselves of gaps which must be filled. In coding kinds of sponsor, for example, we want to know if every scientist has a sponsor, and if every kind of sponsor is represented in our classification scheme. This kind of question leads us to collect new or additional information in our interviews. This is the first analytical use of classifications, but there are more powerful uses as well.

Cross-classification of properties is a technique which leads to many useful results. For example, suppose we have the classification of sponsors and a classification of project types as “Field expeditions” and “Laboratory experiments.” Then we can cross-classify as in Figure 1.

	Type of Project	
Type of sponsor	Field expedition	Lab experiment
None		
Self		
Benefactor		
Foundation		
Gov’t. agency		

Figure 1. Cross-classification of “Type of sponsor” and “Type of project” properties.

Figure 1 is a cross-*classification*, not a cross-*tabulation*. We are considering the usefulness of the classifications rather than the statistical distribution of instances. For example, we aren’t concerned here with how many field expeditions (as opposed to laboratory experiments) have no funding at all. Instead, we are concerned to know if it makes sense to think about field expeditions and laboratory experiments with no funding. Clearly, if such projects exist, then it makes sense to think about them. It may not make much sense to think about combinations of properties we haven’t seen. The combination may be logically impossible given the way we’ve constructed our codes, or perhaps we simply haven’t happened across one yet. Or, the absence of observations for a combination where we expect to find them may suggest that one or both classifications bneed revision. Empty cells then, are open questions.

A cross-classification is useful if it sorts observations in a patterned way. If observations fall randomly in the cross-classification, there is no explanatory or descriptive value to the cross-classification. A pattern in the distribution of observations suggests a hypothesis which can be refined with further observations. This way of looking at it does not necessarily imply the use of quantitative comparisons among cells. Ragin (1987) for example, has described techniques for rigorous comparative analysis which do not depend on quantitative techniques.

Of course, it would be interesting to know that, for example, proportionately ten times as many of one kind of study as the other have government funding. But these quantitative comparisons pose another kind of data compilation problem, and require another kind of analysis, so I will not consider them here. The classical texts for this kind of analysis are Lazarsfeld and Rosenberg 1955, and Lazarsfeld et al. 1972. There are many current treatments; an outstanding one is Lieberson 1985.

A system of effective classifications and cross-classifications acts as the framework of a systematic description. Some of the classifications (such as the one of sponsorship) will be useful in many studies. Some will be unique to the scope of a particular study. Every study has a group of puzzles, that is, observations which don't fit well into the available classifications, or which fall into the "wrong" cells. These puzzles are not merely disconfirmations of a theory or category scheme. They are that, in a narrow sense. But much more importantly, they point the way toward a revision of the classifications we're using, and hence toward a more refined and effective analysis. Indeed, such puzzles are the most useful and valuable data, because analyzing them leds directly to refinements of our knowledge and to new discoveries. There is nothing new to this idea; it is a version of the notion that studying the way something breaks is a good way to learn how it is built (e.g., Garfinkel 1967). Puzzling observations thus provide the means for an important test whenever a new property or attribute is erected. When a new property is proposed, the puzzles can be assessed to see if they fit. If they do, then the new category gains immediate strength. Of course, this is useful only if the proposed new category also works well on the observations which have already been coded.

Analyzing the debating positions adopted by scientists provides an especially interesting example. Often, scientists' positions are not completely consistent. Scientists may emphasize one aspect of a situation at one time, and another aspect at another time. Or they may simply change their minds. Some sociologists (e.g., Gilbert and Mulkay 1984) have argued that this lability in scientists' accounts obviates the possibility of analysis. By contrast, the discovery approach treats this situation as additional data to be analyzed. The apparent inconsistencies among scientists' positions are a means to recognizing additional causal factors operating in a situation. Additional field work and analysis will reveal and classify those factors. In short: inconsistencies pose open questions, and hence are means to discovery.

KNOWING WHEN THE STORY IS STRAIGHT: ROBUSTNESS

How do we know when we've got the story straight? Perfection isn't possible. No analytical scheme or single research project can account for all the contingencies and details in a particular situation. Hence, there can be no complete explanations or descriptions. There are always some possibilities left open, some questions left unanswered, in any study. How then, do we know when to declare a project finished, whether successfully or not? How do we know when our descriptions are adequate? We need evaluation criteria in order to decide when our work is good enough. Confirmation of theory by experiment is often held to be the most important evaluation criterion in scientific research. But confirmation is about testing theories. Hence, confirmation is inappropriate as a criterion of success when we're engaged in getting the story right. What then, is a reasonable criterion of success? Obviously, there are many desiderata. I propose that the *robustness* of the description should be the primary virtue of getting the story right, just as successful prediction is the primary virtue of theory construction.

By robustness I mean the notion that many different things come together to support a single conclusion or interpretation (Wimsatt 1972, 1981). In the words of biologist Richard Levins: "[T]ruth is the intersection of independent lies." (Levins 1966: 426). The terms consilience, colligation, convergence and triangulation have also been used with similar meanings. Here, I shall use "triangulation" to refer to the data collection procedure, and "robustness" to refer to the evaluation criterion or outcome of the procedure.

There are two broad kinds of robustness: robustness of audiences and robustness of tasks. Robustness of audiences means agreement among different participants on the course or outcome of the work. Most concretely and immediately, we have much more confidence in a respondent's claims when we find other respondents agreeing with them. This is the simplest form of audience robustness, and the one most closely tied to interviewing. We are more confident yet when these other respondents come from different audiences. That is, a single summary observation is more robust when multiple different participants support it, and still more robust when different *kinds* of participants support it.

Robustness of tasks means agreement among different research processes or their outcomes. For example, several different kinds of evidence may lead to the same conclusion, what William Whewell called "consilience of inductions." (Whewell 1840). A classic example of this is Darwin's hypothesis that species are mutable and related by descent. Darwin supported this argument with biogeographic, anatomical, developmental, and fossil evidence.

A second kind of task robustness occurs when different procedures lead to the same conclusion. For example, we may collect data by examining primary documents such as notebooks, letters, and diaries; by examining the content of published articles and books; by conducting sample surveys; by interviewing scientists; and by participant observation in laboratories. If each of these different techniques leads us to the same conclusion about the focal research, then we may be confident of the results.

These different kinds of robustness can occur simultaneously of course, and that is just what we hope to find in the course of our study– a kind of *second-order* robustness. When we achieve this kind of robustness, then we can be quite confident of our results.

CONCLUSION

As historians of science increasingly concern themselves with contemporary science, they begin to encounter many of the methodological problems traditionally faced by social scientists. But historians, concerns are not those of social scientists. Historians are concerned with adequate description. This is not simply a matter of collecting facts and making sure they are accurate. The process also requires extensive use of many partial theories and concepts to interpret and organize the facts. The methods of the discovery approach can be used in systematically developing descriptions of complex situations and events in a rigorous way. Doing so means analyzing the data generated by interviews. Analysis begins with the first interviews, and proceeds in tandem with the interviewing process.

There are many kinds of analysis, and each requires many steps. Rather than go into detail, this paper focuses on a few major analysis strategies for getting the story straight. The first major strategy is to discover the audiences of the focal research and build a broad picture of the focal research's context. The second strategy is coding. Coding builds the classifications which are a major tool of analysis. The central point of coding is to treat everything as data. In this view, each observation represents some attribute of some property. Coding is the process of deciding which attribute of which property are represented by particular observations.

The classifications and cross-classifications built in the coding process are the basis of systematic description. Observations which do not fall into the system of classifications are the basis for new observations, additional coding, and new concepts. Analysis of inconsistencies is thus a means to recognizing additional causal factors operating in a situation.

Robustness is the most important evaluation criterion systematic description. Firm reliance cannot be placed on a single interview, single respondent, a single subject, or a single audience. Rather, conclusions must be cross-checked and reinforced by an increasingly sturdy system of independent observations.

REFERENCES

Allen, G. 1978. *Thomas Hunt Morgan: The Man and His Science*. Princeton, NJ: Princeton University Press.

Andrews, R.C. 1922. *On the Trail of Ancient Man*. New York: Putnam's.

Becker, C.L. 1910. "Detachment and the writing of history." *Atlantic Monthly* 106: 524 - 536. Reprinted as pp. 3 - 28 of P.L. Snyder (Ed.) *Detachment and the Writing of History: Essays and Letters of Carl L. Becker*. Ithaca, NY: Cornell University Press, 1958

Becker, H.S. 1967. "Whose side are we on?" *Social Problems* 14: 239 - 247.

Becker, H.S. 1970. *Sociological Work: Method and Substance*. Chicago: Aldine.

Becker, H.S. 1986. *Doing Things Together: Selected Papers*. Evanston, IL: Northwestern University Press.

Becker, H.S. 1997. *Tricks of the Trade: How to Think About Your Research While You're Doing It*. Chicago: University of Chicago Press.

Carlson, E.A. 1974. "The *Drosophila* group: the transition from the Mendelian unit to the individual gene." *Journal of the History of Biology* 7: 31 - 48.

Carlson, E.A. 1981. *Genes, Radiation, and Society: The Life and Work of H.J. Muller*. Ithaca, NY: Cornell University Press.

Cartwright, N. 1983. *How the Laws of Physics Lie*. New York: Oxford University Press.

Collins, H.M. 1985. *Changing Order: Replication and Induction in Scientific Practice*. Beverly Hills, CA: Sage Publications.

Douglas, J.D. 1985. *Creative Interviewing*. Beverly Hills, CA: Sage Publications.

Dunn, L.C. 1965b. *A Short History of Genetics*. New York: McGraw-Hill.

Forman, P. 1991. "Independence, not transcendence, for the historian of science" *Isis* 82: 71 - 86.

Garfinkel, H. 1967. *Studies in Ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall.

Gerson, E.M. 1991. "Supplementing grounded theory." Pp. 285 - 302 in D. Maines (Ed.), *Social Organization and Social Process: Essays in Honor of Anselm Strauss*. Chicago: Aldine de Gruyter.

Gilbert, G.N. and M. Mulkay. 1984. *Opening Pandora's Box: A Sociological Analysis of Scientists' Discourse*. New York: Cambridge University Press.

Glaser, B.G. and A.L. Strauss. 1967. *The Discovery of Grounded Theory*. Chicago: Aldine.

Goudge, T.A. 1961. *The Ascent of Life: A Philosophical Study of the Theory of Evolution*. Toronto: University of Toronto Press.

Hull, D.L. 1975. "Central subjects and historical narratives" *History and Theory* 14: 253 - 274.

Hull, D.L. 1981. "Historical narratives and integrating explanations." Pp. 172 - 188 in L. Sumner, J. Slater & F. Wilson (Eds.), *Pragmatism and Purpose: Essays Presented to Thomas A. Goudge*. Toronto: University of Toronto Press.

Hull, D.L. 1992. "The particular-circumstance model of scientific explanation." Pp. 69 - 80 in M. Nitecki & D. Nitecki (Eds.), *History and Evolution*. Albany, NY: State University of New York Press.

Jonas, G. 1989. *The Circuit Riders: Rockefeller Money and the Rise of Modern Science*. New York: Norton.

Karsten, P. and J. Modell. 1992. (Eds.) *Theory, Method, and Practice in Social and Cultural History*. New York: New York University Press.

Kimmelman, B. 1983. "The American Breeder's Association: Genetics and Eugenics in an Agricultural context, 1903 - 1913." *Social Studies of Science* 13: 163 - 204.

Kohler, R.E. 1994. *Lords of the Fly:* Drosophila *Genetics and the Experimental Life*. Chicago: University of Chicago Press.

Lazarsfeld, P.F., A.K. Pasanella and M. Rosenberg. 1972. (Eds.) *Continuities in the Language of Social Research*.New York: Free Press.

Lazarsfeld, P.F. and M. Rosenberg. 1955. (Eds.) *The Language of Social Research: A Reader in the Methodology of Social Research*.New York: Free Press.

Levins, R. 1966. "The strategy of model-building in population biology." *American Scientist* 54: 421 - 431.

Lieberson, S. 1985. *Making It Count: The Improvement of Social Research and Theory*. Berkeley: University of California Press.

Lynch, M. 1985. *Art and Artifact in Laboratory Science: A Study of Shop Work and Shop Talk in a Research Laboratory*. London: Routledge.

Ragin, C.C. 1987. *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. Berkeley: University of California Press.

Schatzman, L. and A.L. Strauss. 1973. *Field Research: Strategies for a Natural Sociology*. Englewood Cliffs, NJ: Prentice-Hall.

Strauss, A.L. 1987. *Qualitative Analysis for Social Scientists*. New York: Cambridge University Press.

Strauss, A.L. and J. Corbin. 1990. *Basics of Qualitative Research*. Beverly Hills: Sage Publications.

Sturtevant, A.H. 1965. *A History of Genetics*. New York: Harper and Row.

Whewell, W. 1840. *The Philosophy of the Inductive Sciences, Founded upon Their History.* London: John W. Parker.

Whyte, W.F. 1984. *Learning from The Field: A Guide from Experience*. Beverly Hills, CA: Sage Publications.

Wimsatt, W.C. 1972. "Complexity and organization." Pp. 67 - 86 in K. Schaffner and R. Cohen (Eds.), *PSA 1972*.. Boston: D. Reidel.

Wimsatt, W.C. 1981. "Robustness, reliability, and overdetermination." Pp. 124 - 162 in M. Brewer & B. Collins (Eds.), *Scientific Inquiry and the Social Sciences*. San Francisco: Jossey Bass.

Winsor, M.P. 1991. *Reading the Shape of Nature: Comparative Zoology at the Agassiz Museum*. Chicago: University of Chicago Press.