December
1953

American SOCIOLOGICAL Review

Volume 18
Number 6

Official Journal of the American Sociological Society

MEASUREMENT IN SOCIOLOGY *

SAMUEL A. STOUFFER
Harvard University

We now possess considerable knowledge about the conditions favoring inventions in Western culture. With particular reference to technology, sociologists like Ogburn and Gilfillan years ago cleared paths for better understanding. In Gilfillan's *Sociology of Invention,* for example, there are listed and documented some 39 propositions, or principles, relating to the appearance and acceptance of new ideas in technology. The last chapter of Conant's little book *On Understanding Science* is called "Certain Principles of the Tactics and Strategy of Science" and summarizes 21 propositions which he has illustrated from case studies of crucial developments in the history of physics and chemistry.

Why should not we, as sociologists, take an explicit look at the process of invention in the discipline of sociology itself, as a special case of the general working of invention in technology and science? This might not be easy, because all of us have vested interests in sociology which can bias us. But if students of culture do not examine their own discipline as a specimen of culture, who else will do it better?

My observations will be confined to only a limited segment of such undertaking, namely to some aspects of the place of measurement in the process of invention in sociology. I shall take the word measurement broadly to include the use not only of a metric, but also of ordinal position and even of mere enumeration. Loose and inclusive as this is, my observations will be illustrative only, will omit important aspects of measurement in sociology, and will leave out entirely broad realms of description and analysis which are richly productive for sociology though not involving measurement. Such material would need the most careful consideration in any comprehensive treatment. As Cooley liked to point out, the phenomena of life are often better distinguished by pattern than by quantity. Hence, he suggested that a motion picture of the nesting habits of a mallard duck might tell far more about the duck than measurements of the tail feathers.[1] Even if I must leave to others the consideration of areas of investigation represented by many kinds of non-quantitative description and analysis which are staples of sociological, ethnological, and psychological literature, I wish no misunderstandings about my profound respect for their importance, especially in the exploratory phases of research, although I would not accord quite the same order of finality to the conclusions of, say, a DeTocqueville or a Sumner or a Freud, as Redfield seems to do in his challenging discussion of "The Art of Social Science."[2]

We have come quite a way in the last generation or two in the development of quantitative methods. We are even able to measure interactions, and to some extent behavior patterns, not just pin feathers. In-

* The Presidential Address read before the annual meeting of the American Sociological Society, held in Berkeley, California, August 30–31 and September 1, 1953.

[1] C. H. Cooley, *Sociology and Social Research,* p. 315.

[2] *American Journal of Sociology* (November, 1948), pp. 181–190.

deed, the advances in techniques have seemed so rapid as compared with advances in sociological knowledge that some scholars, in their less thoughtful moments, may yearn wistfully for a moratorium on technical progress to give our substantive knowledge a chance to catch up. The phrase "He's a mere technician" is a not uncommon epithet.

A central proposition in the theory of inventions is the postulate that an invention in technology or science ordinarily is not a discovery like an uncharted island emerging from the Pacific mist before the eyes of a Captain Cook, but rather is a long process of juxtaposing, in new combinations, complexes of elements all or most of which are already well known. Among the obvious conditions for such a new juxtaposition are a readiness to see it if it happens and the technical possibilities of seeing it.

The readiness to see it may be due to expectation engendered and disciplined by prior theory, or may be due simply to some combination of habits of curiosity, habits of sharp observation, and luck. The technical possibility of seeing it may depend on the prior existence of an entire technology or a combination of science and technology such as lies behind an electron miscroscope or a modern super-calculating machine.

Let us get to work by examining, for illustrative purposes, a sociological topic in which there is a current interest—the subject of role, and role expectations or obligations. Suppose we would like to study some such proposition as: If a person has obligations in roles X and Y which conflict when situation S occurs, the probability is high that he will sacrifice X and continue in Y. Being scientists, at least by aspiration and self-designation, we want to be the authors not only of an idea but of an idea which can be shown to be wrong if it is wrong or right if it is right.

In the first place, we may need to show what a given role obligation is. If we consider a range of behavior extending from what would never be tolerated to what would be applauded as behavior beyond the line of duty, we will probably visualize a range within that range which would not be subject to severe disapproval at the one end and would not transcend the bounds of favorable expectation at the other. Within that narrower range may lie a still narrower range of behavior which might represent the range of normal expectation, though this might not coincide with or sometimes might not even overlap a range of ideal expected behavior. Moreover, it is likely that such a nest of ranges will not be perceived the same by all members of our social group. There will be variability in individuals' perceptions. We have a measurement problem on our hands.

We look into our carpenter's chest of measuring techniques and find a good many tools. These tools have a history. Without the hundreds of man-years which have gone into them we could hardly get started. We are dealing with attitudes toward certain kinds of behavior. The direct measurement of attitudes is rather new, but indirect measurement based usually on inferences from official collections of data—such as Durkheim used in *Suicide*—are somewhat older. Behind the invention of attitude measuring devices are many complexes of inventions. There is the cumulated experience of testing intelligence and various aptitudes. There are two or three generations of laboratory work in psychophysical measurement. There is a variegated experience in fields like market research and public opinion research—illustrating vividly Conant's observation that science often owes as much to applied and commercialized technologies as they in turn owe to science; lessons learned by practitioners in question-wording, interviewing, and sampling are a case in point. There is a large and growing body of mathematical theory and practical computing methods to provide statistical tests of measurement adequacy. And there are several competing techniques, some of which seem to be mutually contradictory. We have seen in the past generation numerous examples in our own field of what Ogburn and Dorothy Thomas called parallel inventions, only a small fraction of which are likely to survive in anything like present form. Thus associated with such names as Thurstone and Guttman are quite different models for approaching ostensibly the same goals. Out of efforts to reconcile seemingly disparate ideas often come new and better ideas. And we must not overlook the contribution of hardware. Some kinds of measurements or analysis—factor analysis is an example—

would be prohibitively costly if not impossible except for modern computing machines. The latest electrical computers are opening up new regions in statistical theory, stimulating lines of inquiry which otherwise might never have been started.

Frustrations growing out of inadequacies in direct measurement stimulate search for alternatives. Most of the conventional techniques for measuring attitudes assume that a respondent can or will answer a direct question about his attitude. But Freud and those who preceded and followed him produced evidence to the contrary. It is easy to demonstrate discrepancies. Getzels, for example, was interested in studying the norms in a Northern women's college with respect to association with Negroes.[3] The official tradition being liberal, almost all the girls in a dormitory said that they personally would not mind having a Negro as a roommate, but when asked how their friends in the same dormitory felt, they tended to report that their friends would dislike the idea. Either the girls misperceived or misrepresented their friends' attitudes or misperceived or misrepresented their own. Such a result is obviously relevant to our desire to measure role expectations properly. A possible way out may be the use of what are called projective techniques. Here again is a complex, or several competing complexes, of inventions. Techniques like the Rorschach ink-blot test or the TAT picture interpretation test, or various forms of the Sentence Completion Test, which have a history of indebtedness to a number of streams of influence, are still controversial. Most of the inventors in this field have been clinical psychologists or psychiatrists. The consequence is illustrative of propositions in the theory of inventions. Being clinicians, the inventors often had little interest in psychometrics or concern with whether their tests met acceptable statistical requirements of reliability or validity. Hence enthusiastic hopes for the efficacy of such tests tend to be cooled when objective standards are used to check clinical intuitions concerning their efficacy. If, on the one hand, the imaginative inventors were limited by lack of contact with modern statistics, on the other hand the conventional psychometricians, who were in contact neither with psychiatric theory nor with clinical cases, were in no position to dream up such tests in the first place. Only if a combination of the two can be achieved are inventions likely to result which will have both the needed rigor and the needed imagination.

In addition to measurements which might derive from respondents' reports, there is the possibility of measurement which might be derived from an investigator's own observations of verbal or non-verbal behavior of the persons concerned. The Hawthorne studies of role behavior in an electrical plant were able to measure the range in individual output tolerated by members of a small work group which set its own restrictions on production even though members might at least in the short run profit more by higher productivity.[4] Such observations often are difficult to make and costly if an observer has to be on the spot more or less continually for a long time. Moreover, actual behavior as observed may bear little or no relation at times to ideal behavior, which still may have to be ascertained from questioning respondents. Techniques for systematic recording of on-going behavior are themselves a complex of many ideas. Time-sampling procedures for observing children in play groups were introduced at Minnesota a generation ago by Olson, Goodenough, Anderson, and others, and developed further by Dorothy Thomas and associates at Yale. Gadgets like Chappel's interaction recorder were elaborated by Bales and others into devices for classifying interaction in small groups into any of several categories at the same time as the activity is observed. (It may be noted parenthetically that the practical use of such a device requires much skill if two observers are to record the same thing simultaneously and thus guarantee reliability. Whether it is Ogburn or Gilfillan speaking, or Conant on the strategy and tactics of science, we hear that the longest and hardest part of the job in producing an innovation which will

---

[3] J. W. Getzels, *The Assessment of Personality and Prejudice by the Method of Paired Direct and Projective Questionnaire,* Harvard Ph.D. dissertation, 1951.

[4] Roethlisberger and Dixon, *Management and the Worker,* Cambridge: Harvard University Press, 1939.

take hold is usually the job of making it practical and economical. Much of the effort of Bales and his associates has gone into the study and development of techniques for economical training of accurate observers. He even has constructed a special machine with vari-colored lights whose sole purpose is to cut down training time). The new Minnesota laboratory for small-group studies has built-in electrical gadgets for obtaining subjective reports via push buttons which the subjects operate at the same time as their interactions are also recorded by outside observers. Such a device is, like most inventions, only a slight alteration of older devices, many of which in this case also involved the transfer of button pushes to a record on a continuous sheet of paper.

Since in the sociological illustration we are carrying along with us we are interested in measuring role expectations, we must not overlook another technique which, in combination with others previously mentioned, might be useful to us. I refer to role playing. This technique, which Moreno and those trained by him did so much to develop and elaborate, has significance for us far beyond eliciting projective-type information from an individual subject who is taking an unaccustomed role. For it opens a new range of possibilities in the experimental study of group behavior where our experimental subjects are, we hope, "behaving naturally" and where others in the group are trained role players serving as stooges who can shift the course of action according to a prearranged design. The technical problems involved in the selection, training, and utilization of stooges are many and perhaps still only dimly appreciated. Much the largest part of the time, for example, involved in an experiment recently reported by Mills [5] went into learning how to make the role players behave. Some day this will probably become a sub-discipline of its own, with canonical treatises on *The Care and Feeding of the Stooge.* Role playing is, of course, not a technique of measurement, but it is a method which provides ways of checking the possible validity of devices like pencil and paper tests, as well as opening the way to otherwise impossible experimental designs where measurements can be made.

Well, it is evident that if we wish to measure role obligations we have quite a body of technical knowledge available. We are heirs to the efforts of countless people—of whom, incidentally, only a handful were sociologists and of whom many were not even academicians. Yet with all this wealth of know-how, we will find the measurement process an anything but easy task, because none of our tools have the precision or the fool-proof character that we want. This does not mean, of course, that we need to succumb to counsels of despair.[6]

Let us now suppose, as a prelude to our remaining observations, that we can measure what we are after.

Even if we can measure the competing obligations in roles X and Y respectively, we would fall short of our initial objective unless we show that, given conflict in situation S, there is a high probability that a person will sacrifice X and continue in Y. This implies the design of an experiment which contains the possibility of some kind of quantitative test, even if crude. Or, if not an experiment, a plausible facsimile which of course would mean depressing our initial sights.

The demand for experimental proof is rather young in Western culture, as Conant emphasizes in his exposition of physics and chemistry. It is still younger in fields which are closer to sociology or social psychology and are better examples for us in many respects, like biology or medicine. Except for a few dramatic developments in anatomy and physiology, the habit of experimentation in medicine is only a few decades old and hardly has begun in psychiatry. Conant repeatedly emphasizes the tremendous difficulties involved in pioneering experiments where the problem now looks simple and obvious, yet where failure to identify significant variables and control them vitiated results and even retarded progress for decades. So much more complicated was the problem of controlling variables in biology that decisive results in some areas

---

[5] Theodore M. Mills, "Power Relations in Three-Person Groups," *American Sociological Review,* 18 (August, 1953), pp. 351–357.

[6] Such as sometimes seem to be preached by writers like Herbert Blumer. See Blumer's incisive "What is Wrong with Social Theory," unpublished paper presented at the current session of the society.

had to wait on new techniques of experimental design based on mathematical statistics. The statistical innovations of Fisher, which in turn stood on the shoulders of many predecessors—including, importantly, an applied statistician in the Guinness brewery—and have been elaborated by hundreds of followers, are possibly among the necessary conditions for successful experimentation in fields like ours, where the cost of a single experiment in time and money is so great that many of the less relevant but still disturbing variables may have to float around loose and be gathered up in a statistical net rather than be separately controlled. The literature on the mathematics of experimental design and on its applications is now increasing at such an exponential rate that non-specialists cannot keep up with it. Although the mathematics becomes more esoteric and difficult, some of the results are in the direction of facilitation of simpler designs at lower costs with minimal loss of information. Ideas of sequential analysis advanced by Wald and as yet little developed in practice may cut experimental costs in two or better. Use of order statistics instead of statistics based on a metric may give surprisingly good approximations to far more exacting and expensive procedures and permit the use of simple measurements not previously admissible. The making of such ideas practical is, as might be expected from invention theory, proving to be much more time consuming than the initial formulation of the ideas. And there are areas important to sociological experimentation where we have as yet no technical help. For example, there is yet no appropriate probability model with which to formulate a null hypothesis for testing the significance of measures based on a time sequence of interactions within an experimental social group.

The frequent unreadiness of mathematical statistics to facilitate our experimental designs is only one of the handicaps which is holding back experimentation in sociology. Perhaps even greater is the lack of accumulated and transmissible experience in practical arts of handling people in the design and execution of an experiment. Possibly the major proportion of the time spent in training a chemistry Ph.D. is used not in expounding known formulas but in teaching him the arts of the laboratory—transmitting myriads of small cues and skills, including such humble ones as when to suspect that a test tube is about to blow up in his face. Only a few sociologists are now trying to do experiments—and these in not more than a half dozen or so institutions. If, however, as few as ten new Ph.D.'s in sociology a year get training in experimentation and continue to practice and train others who will in turn train others, the curve of transmissible experience will soar, just as it has in the past thirty years in the practical use of statistics. There will still be barriers of cost and it remains to be seen how much we really mean what we say when we salaam before the ideal of verification. For proof comes high. And it is by no means always rewarding to the experimenter if it throws cold water on cherished ideas. Having sponsored a limited amount of such experimentation in the Armed Forces during the last war and in the Laboratory with which I am now associated, I have some rather painful memories.

Sociology perhaps never can look forward to the relatively cheap kind of experimentation which grows a thousand different molds simultaneously, hoping to find in one of the thousand jars a new antibiotic. Because of sheer cost, each experiment will have to be preceded by much thinking, to satisfy as well as possible in advance two questions: (1) If the experiment verifies my hypothesis, so what? (2) Can the experiment be done? The answer to the first question calls for a framework of larger theory, even if that theory is a modestly light and shaky scaffolding capable of being blown down by the first gust of serendipity. The answer to the second question calls for a choice of empirical data such that the numerical values obtained by manipulating the variables will not be obfuscated by even larger errors of measurement. We must search hard for our tobacco mosaic molecule which, though not necessarily intrinsically important, has relatively easily observable properties which can be generalized further. One of Conant's main observations from the early history of natural science is of the failures of experimentation because of choice of materials which could not be accurately enough measured on the instruments of the day.

There is another function of controlled

experimentation in sociology which has not received the attention it deserves. That is, to provide an ideal design against which the imperfections of less adequate designs reveal themselves as a caution against over-confident interpretation. We owe much to Chapin for his studies of imperfect facsimiles of experimental design. If we cannot intervene ourselves and introduce the stimulus situation which forces an individual to choose between conflicting role obligations, perhaps we can observe persons whom we infer to be in the throes of such a choice. Bales, for example, has seen in problem-solving small groups that the best initiator of ideas seemingly cannot also remain the best liked member of the group over a sequence of sessions and that he tends to sacrifice his role of instrumental actor rather than sacrifice his role of good friend.[7] Merton has seen in an interracial housing project a trend toward what he calls value homophily, and has observed the strains when a pair of families who are friends feel they either must yield their friendship or else alter their conflicting values with respect to race relations. Lazarsfeld has translated Merton's problem into a simple model based on a 16-fold table, which has the property of specifying what numerical values to look for, given shifts of pairs of variables in two points in time. He also has generalized further to n points in time, showing that certain initial conditions and tendencies to yield in conflict should result more rapidly in an equilibrium than others.[8]

Unless we can control crucial variables, however, mere changes in time between two variables or mere correlation between two variables in a given point in time may be highly deceptive substitutes for changes which we as experimenters might induce on initially matched groups. We who toiled on the researches in *The American Soldier,* only a few of which satisfy the experimental ideal, are keenly sensitized to this problem and, I fear, drove not only ourselves but also some of our readers to distraction in our efforts to control enough variables to get within hailing distance of the ideal even when we could not overtake it. Re-examining some of this material, Lazarsfeld and Kendall performed an office of formalization in the book *Continuities in Social Research,* which should be the precursor of a growing series of logical analyses of approximations to experimentation.[9]

Finally, even if we establish the particular "if then" hypothesis as to role conflict which we have used illustratively in this paper, there will be a most compelling temptation for overgeneralization beyond this one study. There may be no effective immunization against this, but a better knowledge of how inventive processes work may help build up resistance. Let us remember that no one feat, however heralded, is seldom very important in the totality of feats by hundreds, possibly thousands, of different people which usually are necessary before an important development in either science or technology is firmly established. Nothing is more deceptive for example, than to tag the electric light with the single name of Edison. The following question may be more sobering in the future than it has been in the past, when the significance and finality of particular sociological contributions were too hopefully applauded: How can sociologists preserve their humility without losing their enthusiasm?

In my remarks today I have left unsaid much which should have been said. More is needed, for example, about the direct impact on theory of research design involving measurement, by forcing theory to become operational. Problems involved in the invention and development of useful mathematical models have been passed over. Fortunately, on this same program Dodd, one of our most zealous pioneers, has provided examples of such models which are not mere sketches for future testing but which already have been subjected to test. But I do hope that the illustrations and suggestions which I have advanced will encourage others to examine the subject of measurement in sociology more extensively and systematically, as a specimen of the inventive process in Western culture.

Conant's Principle Number 9 reads, "A

[7] Parsons, Bales and Shils, *Working Papers in the Theory of Action,* Glencoe, Illinois: The Free Press, 1953, Chapter 4.

[8] Merton and Lazarsfeld, "The Dynamics of Value Homophily," unpublished manuscript.

[9] A recent and searching general analysis of the problem appears in an unpublished study by Donald T. Campbell entitled "Designs for Social Science Experiments."

scientific discovery must fit the times." Gilfillan's Principle Number 36 reads, "An invention coming before its time remains undeveloped and practically useless." Are the times now ripe for making sociology cumulative by advancing "if then" ideas which can be proved wrong if they are wrong or right if they are right? I think that many, if not all, of the necessary ingredients are now present in our sociological culture. These ingredients are highly complex collections of ideas, of recorded experience, and of research techniques, some of them mathematical.

Who will put these ingredients together in sociology? Not the philosopher, speculating in his arm chair. Not the sensitive artist, watching human activity with a dramatist's eye. Not the statistician who is solely concerned with making a better probability model or measuring device. Rather, the sociologist who combines several of these skills in his own head, or the small sociological team which brings a few specialists together in a concerted enterprise. Then theory will beget research and research will beget theory, and the Malthusian upsweep of sociology will be on its way, slowly—oh, so slowly at first and so painfully—but on its way, with acceleration. To students in our colleges and universities who may hear these remarks or read them: on behalf of those who have permitted me to be spokesman for the American Sociological Society in 1953, may I bid you welcome into a brave new world.

PRESIDENTIAL ADVICE TO YOUNGER SOCIOLOGISTS

(EDITOR'S NOTE: The following consists of the recorded program of the luncheon meeting in Berkeley, August 30, 1953. At the invitation of President Stouffer, each living ex-president of the Society recorded a brief message. These were assembled on a single record, which was played for the first time at the above session. The program is unaltered, except that the brief introduction of each speaker by Dr. Stouffer is omitted.

The records may be obtained from the Executive Office of the Society, New York University, Washington Square, New York 3, New York. For the single record (2 side) standard long-play 33 RPM records, with the forty-five minute (approximately) program, the price is two dollars and fifty cents.)

SAMUEL A. STOUFFER: Fellow sociologists, here in California today at the 1953 annual meeting of the American Sociological Society, we are about to hear the recorded voice of each of the living former presidents of our Society. As the present incumbent, I am happy to introduce my predecessors. Each will speak for about two minutes on the following topic: What are the best words of counsel you can give to a young Ph.D. just launching his or her sociological career?

JOHN L. GILLIN (*University of Wisconsin*): Gentlemen: I do not commiserate you that you begin your careers in sociology at a time in which perplexity and confusion plague men and nations and when fear possesses so many. No, I congratulate you. A review of the periods of history when there was an outburst of great thinkers on social relationships indicates without doubt that these were periods of vast confusion, of fundamental changes which upset the established organization of society and which brought doubt and frustration to many individuals. These conditions presented a challenge. The thinkers answered that challenge with great courage, with formulated diagnoses, and made suggestions as to how the society could be organized to convert the disorganization which challenged them into an organization fitted to the needs of men. We are in the midst of such another period, perhaps in some respects the most complex of any in history. Some of us who are speaking to you today have lived through rapidly changing conditions which have challenged us to think our ways through the maze. We have tried but there is so much yet to be done. Sociology is yet so young, so immature. So much remains to fill even some of the