AD 702741

UNCLASSIFIED

TECHNICAL MEMORANDUM 951

NOVEMBER 1968

THE ECHO METHOD AND THE STUDY OF VALUES

T. W. Milburn
R. P. Barthel
R. de Mille

GENERAL RESEARCH CORPORATION
P.O. BOX 3587, SANTA BARBARA, CALIFORNIA 93105

This document has been approved for public release and sale; its distribution is unlimited.

Reproduced by the
CLEARINGHOUSE
for Federal Scientific & Technical
Information Springfield Va. 22151

DDC
RECEIVED
MAR 27 1970
C

UNCLASSIFIED

68

ACCESSION for
CFSTI WHITE SECTION
DDC BUFF SECTION
UNANNOUNCED
JUSTIFICATION
BY
DISTRIBUTION/AVAILABILITY CODES
DIST. AVAIL. and/or SPECIAL
/

Prepared for

Advanced Research Projects Agency
ARPA Order 965, Amendment 2

This research was supported by the Advanced Research Projects Agency of the Department of Defense and administered by U.S. Army Research Office-Durham, Box CM, Duke Station, Durham, North Carolina 27706, under Contract No. DAHC04 68 C 0018

Cleared for Open Publication
4 February 1970
Directorate for Security Review (OASD-PA)
Department of Defense

ABSTRACT

Values are subjective estimates of the worth or desirability of any entity or event, whether concrete or abstract. Explanation and prediction of human behavior can be improved by including value statements among the observed behaviors.

ECHO is a new method of studying values; it differs from most other methods by allowing the respondent great latitude of statement, by relating stated values to stated causes or social influences, and by eliciting multiple statements. Though its antecedents go back seventy-five years, only with the advent of computerized data analysis has the collection of such heterogeneous and complex behavioral data become generally useful. The information provided by ECHO complements the information provided by polls or surveys that use more specific and, usually, more numerous questions.

The concept value and related concepts are defined; the analysis of values and their place in behavioral research is treated. ECHO and five other methods of value study are described and then compared against the criteria of reliability, validity, and utility. Forty testable hypotheses are listed for future exploration of the concept value. Sixty-five references are listed.

BLANK PAGE

CONTENTS

SECTION		PAGE
	ABSTRACT	i
I	INTRODUCTION	1
II	THE CONCEPT "VALUE" IN BEHAVIORAL RESEARCH	5
	A. Definitions of Value and Related Concepts	5
	B. Empirical Relation of the Concepts "Good" and "Bad"	9
	C. Additional Aspects of the Concept "Value"	10
	D. Values in Decision Making and Problem Solving	12
	E. Values as Predictors of Behavior	14
	F. Value Change	17
III	METHODS USED TO STUDY VALUES	20
	A. The ECHO Method	20
	B. Content Analysis	26
	C. Survey	31
	D. The Interview	32
	E. Projective Techniques	34
	F. Systematic Observation	36
	G. The Multiple Research Strategy	37
IV	SELECTED CRITERIA FOR COMPARING METHODS OF VALUE STUDY	39
	A. Reliability	39
	B. Validity	40
	C. Utility and Informational Utility	43
V	EVALUATION OF ECHO AND COMPARISON WITH OTHER METHODS	46
	A. Reliability	46
	B. Validity	49
	C. Utility and Informational Utility	51
	APPENDIX: SOME HYPOTHESES ABOUT VALUES	53
	REFERENCES:	59

I. INTRODUCTION

Investigators who seek new insights in explaining and predicting human behavior have, since World War II, looked beyond what the study of attitude, opinion, belief, custom, norm, and related concepts can provide and have turned increasingly to a systematic study of values. ECHO is a new research tool for surveying values verbally expressed by persons in various groups, cultures, roles, and situations.

ECHO differs in three important ways from most methods currently used to study values: (1) a few very broad questions are asked that elicit responses structured almost entirely by the respondent, (2) questions are asked not only about what is valued and disvalued by the respondent but also about the associated sources of causation or social reinforcement, and (3) the same questions are repeated several times, requiring the respondent to generate multiple responses.

The presentation of a novel method obliges the investigators to relate their particular technique to a relevant background of literature and experience, so that informed readers may quickly grasp its broad implications for methodology and theory. The purpose of this paper is to place ECHO in an appropriate theoretical framework, supported where possible by selective reference to the literature. Since this is the first paper on the topic, some sections contain few citations; these sections, like most of the definitions, represent the considered judgment of the authors and are based on many different sources. The method itself has been described by Barthol and Bridge (1968) and will be briefly described at various points in this paper.

The use of broad, open-ended value questions can be traced back at least to Osborn (1894), who studied the moral values of children nine to eleven years old. Boys wrote as many answers as they liked to two questions: "What must a boy do to be called a good boy?" and "What must he do to be called a bad boy?" Girls answered a corresponding pair of

questions. The responses described the good child as obedient, truthful, helpful to others, gentle, kind, unselfish, loyal, patient, quiet, and polite. The good boy was said not to fight, smoke, or play hookey; the good girl did not whisper in school, make noise on a rainy day, or drum on the piano. Osborn classified the responses into several kinds of behavior and found that more than half of the children contributed to a category involving obedience, while only about a quarter mentioned truthfulness. Intergroup comparisons were made between males and females and between classes in two different schools.

A technique that is the direct forerunner of the ECHO method was introduced by Bavelas (1942), who described its application to school children as follows:

> The first question was, 'What could a child of your age do at school that would be a good thing to do and someone would praise him?' When the child had answered the question, the experimenter asked, 'Who would praise him?' This unit was repeated three times, the second and third time the child being asked, 'What else could a child of your age do and someone would praise him?' On the fourth repetition the experimenter changed the statement by inserting the adjective 'very' before the word 'good,' and by changing the second half of the question to read, 'Who would praise him very much?' On the seventh repetition the experimenter read two 'very's' where one had been. In all, there were nine repetitions of the question asking for praisable activity. This unit of praise questions was followed by a unit of nine scold questions phrased in a corresponding manner. (1942, p. 373)

The Bavelas technique was superior to Osborn's procedure in three important ways:

1. Bavelas presented each question a standard number of times and elicited multiple responses from every subject.
2. The source of praise or blame, implicit in Osborn's questions, was made explicit by Bavelas.

3. Relationships between valued or disvalued acts and the sources of social reinforcement were analyzed.

Kalhorn (1944) applied the Bavelas technique in a study of Mennonite and non-Mennonite school children, finding intergroup differences in values and in sources of approval/disapproval. Havighurst and Neugarten (1955) used the technique to compare midwestern white children with children of ten American Indian communities.

The ECHO method differs from the Bavelas technique in the following ways:

1. Oral interviews are used only with respondents who cannot read and write.
2. Not only are responses classified by the experimenters, but they are also usually classified by members of the pool from which the respondents were drawn.
3. When appropriate, value statements drawn from certain populations are presented, along with comparison stimuli, to samples of the same and different populations to be judged for goodness, acceptability, etc., thus providing a measure of concurrent validity.
4. The intra-question variables of valuation, role, situation, event, source, and reinforcement have been conceptualized and manipulated.
5. The analysis of data has been further developed and largely computerized.
6. In addition to the usual summary tables, full display of the classified responses is readily available for reports.

In this paper, we discuss several aspects of the concept <u>value</u> that are germane to our research interests. We differentiate value from, and indicate some of its relationships to, concepts such as attitude,

opinion, belief, and norm. We use the indispensable criteria of reliability, validity, and utility, to compare ECHO with alternative ways of studying values, pointing out some merits and demerits of various methods. And finally, we discuss the power of ECHO to generate plausible and testable hypotheses, listing in an appendix a number of hypotheses that are worth testing.

II. THE CONCEPT "VALUE" IN BEHAVIORAL RESEARCH

Since ECHO is a method for studying values, its usefulness will depend on whether the concept value is a scientifically fruitful one. This section is a discussion of the concept in behavioral research.

A. DEFINITIONS OF VALUE AND RELATED CONCEPTS

The concept value as used in this paper incorporates the ideas of goodness, attractiveness, interest, preference, satisfaction and their opposites. Stimuli or outcomes which man finds attractive and thus preferable to others have positive values; the reverse is true of negative values. As predictors of behavior, values are independent of concepts concerned with capacity, (e.g., intelligence, power, social skill). Individuals hold values, but values may also belong to groups, organizations, societies, and cultures, or they may be specific to roles and situations. Values result from shared human perceptions and shared meanings. They tend to be stable, but may change in strength and order over time and with changing circumstances.

Value is defined as the implicit or explicit estimate by a person, group, or organization of the merit, excellence, desirability, or worth of a behavior, principle, quality, event, or entity.

A value may be more or less bound to a particular person, role, situation, or object; it may vary along the independent dimensions of desirability and demand; its importance to the holder may be modified by the expectation (subjective probability) of the appearance of what is valued and other events (e.g., success or punishment) that may be associated with it. The investigator bases his determination of the value's specificity on the operations he uses to collect and analyze the data which generate the value statement. The value may be detected by observing behavior directly or by studying written or spoken reports of it.

Values may be further defined by four conceptual variants:

1. Values are selection criteria, bases for judging the merit of objects, persons, ideas, events, or relations. They are bases for choosing or for developing agenda.
2. Value is also used to refer to behavior which has a required or "ought to do" quality. Such usage emphasizes the moral aspect of the concept, and it emphasizes values as behavior rather than as stimuli.
3. By contrast, some psychologists concerned with the experimental analysis of behavior use the term value to refer to those objective stimuli which are apparent goals of an organism (reinforcing stimuli). Thus, a person may be described in terms of those internal or external stimuli that serve to influence his behavior (rather than in terms of his needs, wants, or ambitions).
4. Social psychologists emphasize either the affective (feeling, sentiment) part of values, or the expectancy (set, attitudinal) part of values. In their study of the structure of meaning in various languages, Osgood, et al. (1957) consistently found that the first factor extracted from their matrices of correlations among word qualifiers was the evaluative factor. They see values as the primary component of meaning. Values seen as sentiments, or as the primary component of language-meaning structures, are not very different from attitudes except that they are more general.

The following terms--attitudes, norms, beliefs, habits, customs and folkways, mores, and opinions--are all related to values. They will be defined, but the definitions must be considered somewhat

arbitrary since they are used uniquely by some persons and interchangeably* by others.

Attitudes are defined as predispositions to respond to entities or to situations in specific ways. The object of an attitude may be abstract (democracy, love) or concrete (my dog). An attitude may be inferred from behavior. Attitudes have both intellectual (cognitive) and emotional (affective) components although the affect may be neutral. The social legitimacy of a disposition or preference is irrelevant to its classification as an attitude.

Norms are rules, standards, or patterns of behavior in specific situational contexts shared by two or more persons. Norms tend to reflect consensus and a common value system within a group. The term value may refer to a specific behavior; that same behavior, when commonly performed in a group, may also be called a norm. Behavioral norms can be identified from observation or inferred from verbal reports. Just as publicly held values are unreliable indicators of private behaviors, so norms that are inferred from reports of private behavior are unreliable indicators of publicly held values.

Beliefs are enduring, perceptual, and cognitive structures about the nature of reality; they are statements concerning what is real and credible in and about a person's world. Belief systems may be open or closed, i.e., flexible and open to change on the basis of new evidence or not; ideologies are coherent and tend to be closed and inflexible.

---

* There can be some rationale for this. Values may be seen as attitudes writ large, as more abstract and general attitudes, particularly when values are seen only as preferences or criteria, with no oughtness or legitimation implied. Beliefs are treated here as if analytically distinguishable from values and attitudes because the distinction appears feasible and useful. However, others may reasonably regard beliefs as merely the intellectual (cognitive) components of attitudes. And still others emphasize the cognitive aspect of values including all values under the rubric of beliefs.

Krech and Crutchfield (1948) suggested that beliefs could differ in precision, specificity, strength, and saliency. The strength of beliefs need not depend upon verification or the possibility of verification, e.g., superstitions and beliefs about the existence of a supreme being. Beliefs and attitudes give continuity to a culture as well as to the personality of individuals. They give meaning to daily perceptions and activities and may arise because of the human need for meaning; they determine which solutions to problems are chosen.

Habits are learned actions, feelings or ways of perceiving. Habitual behaviors may be contrasted with other behaviors which are based on choice. To act, feel, and perceive a certain way from habit is to do so because such behavior has been reinforced in the past and has become semi-automatic, not requiring conscious choice or awareness of preference.

Customs and folkways are more than mere aggregates of individual habits; they are concrete social patterns of behavior, the outcome of shared but largely unreflective trial-and-error adjustments. They are always supported by social approval, and deviations from them may be punished.

Mores are customs "that are regarded by general agreement as highly important and obligatory, as evidenced by strong sentiments against deviation and by severe punishment for violation. Incest taboos and rules against in-group murder, rape, cannibalism, and other practices generally regarded as especially heinous," (Williams, 1968, p. 205) are examples of mores; so are the obligations of the scientific community to report results accurately and completely.

An opinion is a view, judgment, or appraisal formed about a particular matter; the term implies a conclusion that is open to contradiction. Opinions have neither the proven property of knowledge nor the unproven

property of faith (a firm belief in something not proven). Opinions may be about beliefs, attitudes, or values, although typically they focus upon fairly specific beliefs or preferences. Opinions are often so specific that the exact wording of questions may considerably influence the answers.

B. EMPIRICAL RELATION OF THE CONCEPTS "GOOD" AND "BAD"

If good and bad were closely and inversely related, not much would be accomplished by employing both concepts. It appears that good and bad, though logical opposites, are not often opposites on linear dimensions of response, but are either scarcely related or related in a nonlinear fashion. Whenever they are unrelated to one another but related to the same criterion behavior, the very lack of relationship will be advantageous, because goods and bads can contribute unique variance to predictions of the criterion. Good values may be linearly related to one another; bad values may similarly be related to one another; but the set of good values and the set of bad values are likely to represent orthogonal dimensions.

In a particularly ingenious and well-designed study an experimental psychologist, Sidney Siegel (1964), found that the use of two value dimensions greatly increased the accuracy of his predictions of behavior. A psychologist and two experimentally-oriented philosophers raised the question whether monetary losses and gains (a bad thing and a good thing) that were objectively equal would also be subjectively equal. They found that for most persons they were not: small losses usually were treated as equivalent to larger gains.

Two other lines of evidence suggest that good and bad will not usually be found in an inverse linear relation. One line of inquiry derives from the work of the industrial psychologists, Herzberg, Mausner, and Snyderman (1959), who asked their interview respondents to describe times when they felt especially good and especially bad about their jobs;

the results have been replicated in a study of male supervisors in a public utility (Schwartz, Jenusaitis, and Stark, 1963). Herzberg, et al., found that sources of job satisfaction were independent of and different from sources of dissatisfaction; the two attributes were not merely opposites. They also found that content factors (e.g., the kinds of problems encountered, self-actualization) tended to be related to satisfaction, whereas contextual factors (e.g., heat, cold, ugliness) tended to be related to dissatisfaction--but that part of the finding may be true only for professionals and managers. While specific findings of statistical independence between content and context may not hold up, the more basic finding that the sources of satisfaction are not the opposite of sources of dissatisfaction but are unrelated to one another appears to be a significant and useful discovery.

A second line of evidence emerges from a survey approach to the study of psychological well being by the National Opinion Research Center (NORC) of the University of Chicago. Bradburn and Caplowitz in their Reports on Happiness (1965) discuss the considerable lack of relation between reported sources of positive feeling (affect) and sources of negative feeling. In none of these studies does the mere lack of goods equal something bad or the mere lack of bads equal something good. A similar finding occurs in several unpublished ECHO studies, particularly the ones concerning what is good to have happen and what is bad to have happen. One implication of these findings is that in exploring what people regard as worth doing and worth having happen one should ask not only about good and better, but also about bad and worse.

C. ADDITIONAL ASPECTS OF THE CONCEPT "VALUE"

Studies of values focus on various aspects of the concept. Besides such relatively simple characteristics as strength or intensity, specificity-diffusion, salience, clarity, and centrality, values may be studied in combination with other variables such as time, situation, culture, and role. A study of value preferences or requirements for specific roles or situations may have considerable practical interest.

Some contributors to the empirical study of values, e.g., philosopher Charles Morris (1956), describe value sets (different styles or ways of life) in part, as sets of means; not as ends. By contrast, psychologist Ralph K. White (1951) would define operationally only those values which appear as ends in that they are themselves not further defended or rationalized. Philospher John Dewey restricted the use of the term value to those situations in which value-preferred actions (behaviors) led to outcomes of interest; in other words means and ends come in sets and both should be preferred for either to be valued. Political scientist Robert C. North (1963) sees the distinction as fairly unimportant since, in the fabric of historical reality, actions conceived as means are likely to be ends for some other means, and ends are legitimately considered means toward other ends.

Students of crises recognize that threats to values or valued entities increase the prominence or visibility of the threatened values, thus increasing their importance--at least temporarily. Values "temporarily" prominent on agendas during crises may, as a result of those actions or of commitments calling for action, become permanently prominent after the crises have passed.

Recently, Milton Rokeach (1968), making a strong case for the utility of the study of values, developed two rank-orderable scales of values (instrumental and consummatory) which emphasize their means-ends distinction. Respondents are given two separate lists of values to rank order in their own value system, within the constraints of a forced-rank distribution. Rokeach has in a sense "validated" his value scale by collecting data to show that values are superordinate or prepotent over less general, more specific attitudes, as he had predicted. He relates values to one another in terms of a rank order scale of preference of attributes of persons and behaviors.

Essentially, Rokeach applied some of Abelson's psycho-logic to the situation. Abelson (1959) argued that rules of psycho-logic would

apply to situations where Ss* were conscious of difficulties. In an unpublished study, Milburn found that Rokeach means (instrumental values) differentiated between college men and women far more readily than did ends (terminal values), suggesting that terminal values are more general and less sex-role related than are instrumental values.

Rokeach instrumental values look like "good to do" items in the ECHO format, while the consummatory or terminal items look like "good to happen" ECHO items; the items are more abstract, however, and clear relationships with ECHO questions have not yet been demonstrated.

Rokeach does not consider negative values, nor does he include the analytic categories relating to perceived antecedent sources and consequent effects of valued behavior. In addition, Rokeach omits the distinction between duty (ought to do) and desire (like to do), an important topic in philosophical writing about value.

D. VALUES IN DECISION MAKING AND PROBLEM SOLVING

Like other scientific concepts, values can be used to explain and relate diverse behaviors. They also play a central role in all decision making, but their effect is modified by information, beliefs, and the estimates of probabilities that certain events will occur, or the conditional probabilities that these events will occur only under certain conditions. A general outline of the decision-making process as affected by values looks something like this:

1. A person observes events within his life space. He perceives these as worth his attention; they are above his threshold of awareness.
2. He places on his agenda events constituting problems or obstacles to the enhancement of values.

---

* The standard abbreviation S will be used to stand for "subject," any individual used in an experiment or survey.

3. He constructs or discovers alternative possible solutions.
4. He selects an alternative for implementation.
5. He implements the selected alternative.

Values play a central role whenever a person makes a rational choice or selection, although they may be less important where behavior is already determined by policy, custom, habit, neurosis, or reflex. Values thus contribute significantly to each phase of the decision-making process. A set of events or circumstances must first be in one's life space, i.e., be perceived as existing. Values operate at this point largely through the mechanisms of selective perception, a process usually not involving conscious choice. However, once problems appear significant enough to be assigned space on the problem agenda, the presence of obstacles to the achievement of positive values adds to the prominence of these problems, and the values they affect.

The strongest emphasis on the role of values has been in the primary choice itself, yet values are not the only determinant. Information, beliefs, and estimates of probabilities are associated in the decision maker's mind with each alternative, along with conditional probabilities that other events will occur, given certain apparently developing sets of events. For example, most scientific hypotheses are based on conditional probabilities.

Values, attitudes, and beliefs in part determine what will be seen as alternatives for the solution of problems. New data, physically available to an individual but contradictory to his beliefs and attitudes, may not even be perceived (Krech and Crutchfield, 1948, p. 190); or once perceived, they may more readily be forgotten. Disvalued sources of information may be forgotten when the information is still remembered. New information has its meaning determined in large part by the existing structures of values, attitudes, and beliefs. Its initial acceptability

depends also on the perceived strength, goodness, and credibility of authorities who are seen as sources.

After the alternative has been selected the choice must be implemented. Whenever the means and character of the implementation are predetermined, values become important only if they are in opposition. If selection is required, values operate as described earlier to affect the choice. The influence of values on the character of implementation has been largely ignored, although implicit recognition is widespread.

E. VALUES AS PREDICTORS OF BEHAVIOR

Much of the time, people do what they like, choosing behavior they value; much of the time, they do what is socially required, behaving according to social norms. The two kinds of behavior are sometimes identical, sometimes not. A given behavior can have positive, negative, or neutral self or social value or any combination. Accurate prediction of behavior is more probable when individual preferences and social norms converge, giving rise to the same behaviors.

As indicated elsewhere in this paper, no simple relationship exists between values and behavior. A given value may underlie many different behaviors, and a given behavior may arise from many different values. Establishing the relationship between a value and a behavior is not necessarily useful. Stopping at a stop sign, a classical example of J-curve behavior* (Allport, 1924), is a behavior that value theory can explain--but the explanations are likely to be multiple and competing, and therefore of little merit. Although many values may be relevant (safety, conformity, respect for law, money, convenience, etc.), forces other than values tend to be overriding. Even in this situation, however,

---

* Some behaviors are performed by almost everyone in a society. When the frequency of certain deviations from one of these behaviors is plotted on a graph, the resultant curve resembles the letter "J."

a knowledge of values could be useful for predicting behavior: when contextual factors are not yet known, a knowledge of values allows the investigator to estimate which factors are relevant and to seek information about these only, rather than all possible factors.*

Further, values that seem to be widely held may not reliably indicate how people will behave, as the Kinsey, et al., studies (1948, 1953) have made clear. One's preference for studying usually leads to his studying in situations that normally elicit such behavior (attending school when school is in session), where studying is approved by peers and authorities; but a high value for studying does not predict studying when one is tired or must earn money or attend to a crisis or to other important yet rare events. Nor does it predict his studying in church or at a cocktail party.

A knowledge of values alone will not make it possible for an investigator to predict effective performance on tasks demanding skill, persistence, or considerable expenditures of energy. Under many circumstances a value for working, effective work habits, and relevant intellectual capacity must all be present to bring about effective performance. Value preferences plus requisite capacities may be predictive of effective performance.

To predict a person's behavior accurately, one must know the person's perception of what attitudes his peers and authorities will have toward the predicted behavior and also what attitudes the person will have

---

* For example, using the stop sign behavior as a paradigm, one might wish to predict behavior at newly installed stop signs in two different countries. If the values of the drivers in Country A are high for courtesy, respect for law, respect for rights of others, and conformity, the investigator need only learn whether drivers stop at already existing signs to predict J-curve behavior. If, on the other hand, the values of the drivers of Country B are for independence, self-interest, laissez-faire, and machismo, then additional information would be needed: the history of law enforcement, attitudes toward police, knowledge of police surveillance by drivers, the location of the stop sign, etc.

toward those peers and authorities. If he believes the behavior will be seen as desirable and he also desires it, then such behavior should appear more often because it will be at least twice rewarded: extrinsically by peers or authorities and intrinsically by task and self. However, if the behavior is desired but peers or authorities will be opposed to it, its appearance will depend on the subjective probability of being caught, the expected severity of punishment, and the person's attitude toward the peers or authorities. Punishment could consist of shame in cultures (such as the Oriental) where interpersonal disapproval constitutes a strong negative sanction. Punishment could also consist of guilt, if the internalized societal values are opposed to desired activities. Curiously, however, a number of studies, as well as clinical experience, indicate that the strength of guilt feelings does not predict strength of resistance to temptation.* Guilt feelings sometimes follow the performance of a socially disapproved act when preference for an object or behavior, which has been strong, declines markedly upon consummation. After a strong desire has been satisfied, the perceived worth of the action may decline for some persons, who are then beset by guilt--contrary to Festinger's (1957) hypothesis of post-decisional reduction of cognitive dissonance.

Even though a person believes that the authority figures in his world see a certain behavior as desirable he still may not act as they prefer. For example, in some situations, behavioral norms may call for aggressive behavior, but if aggressive acts are disvalued they are less likely to be carried out. Peers and authority figures who are perceived as sources of praise or blame may also be seen as causes of the good and bad things that happen; a person is more likely to conform to the norms

---

* These studies have shown low correlations, no relationship, or even negative relationships between values and resistance to temptation (e.g., Hartshorne and Hartshorne, 1929; Mills, 1958) and between guilt and resistance to temptation, (e.g., Allinsmith, 1960; Burton, 1959; Grinder, 1962; Maccoby, 1959; Sears, Race and Albert, 1960); no studies have demonstrated that strength of moral values, resistance to temptation, and proneness to give projective guilt responses are covariant (Pittel and Mendelsohn, 1966).

of his subculture if he positively values these authorities as both good and powerful.

F. VALUE CHANGE

Subjective values and valued entities can be placed in rank order of preference and intensity, but these orderings change from time to time; moreover, the attractiveness of any entity, event, or relation also changes over long or short periods. Reinforcement of a valued behavior is the most effective way of increasing the attractiveness of that behavior or the probability that that behavior will be elicited again when appropriate cues appear.

If the environment elicits and reinforces certain behaviors, those behaviors will appear often. If reinforcement of behavior rarely comes until an organism has produced considerable effort to achieve the valued condition or behavior, the organism also learns to try hard, to persevere.*

For many persons, the stimuli they choose or the behavior they elect in situations of choice afterwards become more valued; those they did not choose become less valued. This shift in values tends to reduce the feelings of discomfort or regret (post-decisional dissonance) occurring when the once possible alternatives are reviewed (Festinger, 1957).

Certain regular psychological phenomena make changes in value ordering inevitable. One common process is that of satiation. When a

---

* Psychologist David Premack (1959) found an interesting relationship in a hierarchy of observed values (i.e., reinforcing stimuli): a given behavior that an organism valued could either reward or punish the organism, depending upon whether the behavior led to more highly valued or to less highly valued behavior. If a student likes chess more than he likes to study (but likes both) and his freedom to play chess becomes conditional upon a certain amount of study, both his capacity for study and the attractiveness of study are likely to increase. (This is similar to the behavior identified by the label "functional autonomy.")

stimulus or experience or behavior occurs with great frequency and intensity it may become aversive. A simple tune which goes on and on at a merry-go-round across the street increases in its tendency to irritate us, despite the fact that we enjoyed it when we first heard it. Politicians and advertisers are sensitive to satiation problems, which they refer to as overexposure. An idea or a person that is "overexposed" may appear less pleasant or may have become somewhat aversive.

The reverse may also occur: much experience with any stimulus object or behavior may lead to habituation. That object or person becomes part of the context of our lives and largely drops out of our awareness. One example of habituation occurs when we no longer notice the temperature of 68° F water in which we are swimming even though the water initially seemed cold. The new member of the Toastmasters Club, terrified by his first speech, eventually looks forward to his Thursday luncheons with the group.

Similarly, a period of isolation from all but low level sensory stimuli (sensory deprivation) can lead adult organisms to find almost any stimulus variation desirable. A period of isolation from other persons which at first provides a pleasant sense of relaxed privacy, can become boring and even disturbing. The isolated normal adult finds that he needs other persons and that their attractiveness has increased markedly for him. A period of absence from familiar events or people can cause their value to increase.

Value orderings may change in conditions of crisis. Some theorists argue that crises do not change the strength of values, they merely render more visible or prominent those values threatened with loss or destruction.

Change in preferred behaviors may occur also if a respected and credible authority presents arguments, evidence, or even merely his

opinion that change should occur. The authority may exert his influence on the preference structures of individuals for belief preferences as well as for behavior. The authority figure may be influential within an organization because he plays a role in the legitimation of production norms or those values associated with moral (interpersonal) behaviors, not merely because he monitors the distribution of information. The change of social values, i.e., the values of a group, may follow behavioral change* but usually must depend upon a consensually accepted procedure for the legitimation of new values, either implicit or explicit. Although such changes occur, the net effect is that values and beliefs tend to be constant, and people perceive their environment to be more constant than it really is; such constancy helps to hold social groups together.

---

* The discrepancy between public sexual norms and private behavior as described to interviewers at the time of the first Kinsey volume (1948) was quite large. Knowledge of actual behavior appears to reshape behavioral norms rather than behavior itself, as a means of reducing the norm-behavior discrepancy. However, the erratic legitimation by commonly accepted authority figures may account for the equally erratic changes in public sexual values.

III. METHODS USED TO STUDY VALUES

A number of research methods have been or might be used to study values. This section will describe ECHO, content analysis, surveys, interviews, projective techniques, systematic observation, and the multiple research strategy. ECHO will be compared with other methods.

A. THE ECHO METHOD

ECHO is one kind of survey used to gather and analyze information about the values of groups. It is a flexible tool that can be adapted to gather information about all aspects of the concept value discussed in the second section of this paper. The fact that ECHO queries specific roles and situations is a particular strength because of the usefulness of the information and the stability of values so delineated. Since these values rise from questions that define specific situations, they have a higher probability (than values elicited by general questions) of being relevant to the defined situations; they tend to remain stable so long as the real situation is unchanged.

When the ECHO method of investigation is employed, assumptions are made that verbal statements are related to other behaviors of individuals or groups and that (when the situation permits, legitimates, and encourages truth telling) most respondents tell the truth to the extent that they can. The unusual nature and pressure of the survey means that respondents typically lack techniques for subtle resistance to the ECHO task.* Thus they reveal "truth" that tends to be about problems relatively prominent or salient to them rather than those that are widely accepted and conventional or merely habitual.

---

* Conspiracy among the respondents would render the findings useless as an estimate of values, but the fact of the conspiracy would be obvious. The pressure of making many responses causes most subjects either to answer honestly or to retreat to facetious statements or non-response. Facetious and other evasive responses are easily identified. The uncooperative subject reduces the N of the study; he rarely distorts the results.

Steps in the ECHO Method. In Step I, general, value-relevant questions are asked.

1. "For a person like you [or in role X], what is a good thing to do? If you did such a thing who would praise you? What would be a bad thing to do? If you did such a thing, who would criticize?"
2. "What is something good [or bad] that has happened [or could happen] to a person like you? Who or what would cause it to happen?"
3. Particular roles or situations may be specified. e.g., "What would be a good thing to do if X were elected?" or "What would be a courageous [or cowardly] thing to do in the village?"

The amount of information ECHO supplies can be increased by asking questions about both preferred and required behavior--to find out whether there is conflict or harmony between what one likes to do and what one ought to do. Such knowledge adds to the investigator's ability to predict behavior.

The categories of either reinforcers of events (those who would praise or censure) or causes of events (those who or that which would bring about significant events) constitute hierarchies of perceived benevolent and malevolent influence in men's lives. In the ECHO method, perceived reinforcers or causes are called sources. (ECHO source questions bear a resemblance to the items of the Implicative Semantic Differential, developed by Harry Triandis [1966].)

In Step II, the value statements, usually written by the respondents on prepunched IBM cards, are classified into a set of categories by indigenous classifiers. Representatives of other subcultures, or professional classifiers, may also classify the same responses. Role-playing classification is also used, in which members of one group classify as they believe members of another group would.

Theoretically designed categories, to elicit certain kinds of information, can be constructed in a variety of ways, such as common inferred cause, common consequence, common goal, similar essential features, value level, etc; teams of classifiers may employ a variety of conceptual approaches, boundaries between concepts, or topological constructions. Personality features of classifiers and of groups may be systematically related to the structure as well as to the content of categories.

A particular subculture in contrast to another will, by putting sets of items together in a unique way, reflect a subtle difference in its view of reality.* When categories are well constructed, they possess a useful degree of invariance across various transformations, reflecting stable aspects of the internal and external environment of a group or subculture. Indigenous classifications can be compared with professional and/or theoretical classifications.

In Step III, computer programs are used to analyze the data and present it for ready interpretation.

ECHO information can be enriched by combining it with information gathered in other ways. If ECHO categories of value and influence are treated as concepts to be rated on semantic differential forms, additional information can be obtained about relations between preferred and required behavior; the semantic differential dimensions (orthogonal factors) of intensity and activity can be added to descriptions of these concepts; and descriptions of the interrelationships among values and sources can be expanded.

---

* Such a distinction was apparent in two different classifications of the values of freshman nurses, one by freshmen, the other by senior nurses. The two sets of categories were related to one another as idealistic-perfectionistic is related to pragmatic-efficient. In another investigation one group of college students produced more differentiations among interpersonal and self-actualizing behaviors than did the other.

To add information, conventional closed-ended survey items can be used along with the ECHO questions. Indigenous informants can state when the interpretations of responses and categories are incorrect and can explain the prevailing consensual rationale for socially required behavior. Ratings of behavior and documentary information about ECHO populations are useful for validation. Conversely, ECHO can add information about Ss' perceptions of the situation in psychological or sociological laboratories, in simulations utilizing interactions between human beings and computers, or in other observational settings.

One of the main strengths of ECHO lies in the open-ended question, which prevents the investigator's premature conceptualizing of what is important. Spontaneous value statements are likely to reflect concerns of the respondents rather than expectations of the researchers. The ECHO question also makes it difficult for respondents who choose to provide answers they think the investigators would like to bias the data systematically. Other deliberate distortions are themselves revealing.

Standard procedures have been developed for data collection, analysis, interpretation, and presentation; ECHO observations made in various settings can, therefore, be compared. The system of classified responses and their associated category titles form a rich, well-differentiated basis for intercultural comparisons. Data have already been obtained from a number of populations, especially from college populations. Collection of data from illiterates is feasible though tedious. Costs of the various procedures have been brought to practical levels.

The statistical reliability of ECHO categories is quite high, especially with professional classifiers. The lower reliability inherent in indigenous classifications is an acceptable cost for the intrinsic and corrective renditions that indigens give of their own value hierarchies.

Computer-based data analysis, statistical tests, information display, and document preparation have been developed to make ECHO results

prompt, accurate, and understandable. Prior to the availability of high speed computers, the practical and adequate use of data resulting from open-ended questions like the ECHO questions would not have been feasible. Automated organization and display of information allows the ECHO report reader to participate in the analysis, critically evaluating the report in detail, and bringing to bear on the report any special information he possesses. This procedure is in contrast to the clinical mystique, in which detailed data may be either unavailable to or uninterpretable by the reader. In the ECHO report, all data and every step of the analysis and interpretation can be made explicit.

A critical assumption of the ECHO method--that there will be high incremental utility from repeating one question a number of times (7 or 10)--has received empirical support. ECHO Ss seldom give the same response to a value question more than twice or three times; the mode is one contribution per S per category.

Limits of ECHO's Ability to Predict Behavior. Like other methods, ECHO is limited in its ability to predict behavior. Some important limiting factors are listed below.

1. Values are not norms. The commonly used ECHO questions produce statements of value; norms refer to behavior under specific circumstances of role and situation. However, values can be used to predict or hypothesize behavioral norms, and vice versa.
2. Factors other than values help to determine behavior. Some examples are role, situation and other contextual factors, specific beliefs, specific motives, the individual's perception of his world, incentives, and threats.
3. Even if intended as predictions by the respondents, value statements may prove inaccurate because people

often cannot predict what they might do or feel in circumstances they have not encountered before.

4. Persons may lie or not be aware of the values that determine their behavior. Some people are more sensitive or vulnerable to the evaluations of others; they may seek to describe more attractive sets of values and norms than they actually possess. The degree to which they do so will depend on their perception of their anonymity, and on their assumptions about the motives and expectations of the investigator as well as the demands of the situation. For these reasons, the analysis must include internal checks and cross-group comparisons.
5. Values which are not sufficiently visible or salient may not be mentioned even though they serve as criteria for preference and actions.
6. A person's values are usually part of complex cultural value systems. Knowledge that he holds certain values is not enough; the relations among them must be specified. Oppositions and contradictions among values are not unusual; individuals and groups must inescapably face choices from time to time. Even the most harmonious systems of values require selectivity in the balancing of different claims to time, energy, and other resources.
7. Specific acts or sequences of acts are usually steered by multiple and changing clusters of values.
8. Any one method of observation or inference contributes its own specific inherent error, bias, or distortion. For example, the specified or implicit role of the respondent, the situation in which the responses are given, interactions between respondents and interviewer or administrator--all could introduce distortion in a hierarchy of values that could mislead the unwary investigator.

B. CONTENT ANALYSIS

Content analysis consists of description in terms of an economical set of categories of written or observed material and its systematic analysis and interpretation in whole or part. Content analysis refers mostly to verbal materials, but some non-verbal material may also be treated. A wide variety of material has been content analyzed.* Several of the studies are cross cultural. Angell and Singer, for example, compared U.S. and Soviet periodicals directed to six comparable elite groups upon some fifty value dimensions. Weakland's analysis is of values expressed in Chinese communist films as compared with those of a control group. McClelland used the Yale-developed Human Relations Area Files to explore several values in many nonliterate cultures.

Content analysis has proved most successful (i.e., made the most valid prediction) when predictions from it have been based on contextual and situational information in addition to analysis of the main corpus of material. Content analysis has points of similarity with ECHO, especially the classification of data, intended in both cases to reduce the complexity of data without sacrificing meaning.

There are three different approaches to content analysis: quantitative, inferential, and predictive.

1. Quantitative Content Analysis

In quantitative content analysis, words are counted according to some set of rules without inferences about underlying meanings and without going outside the context of the material in the documents. Lasswell (1938) was one of the pioneers of this approach.

---

* The list includes myths and children's stories (McClelland, 1953; Atkinson, et al. 1958) films (Weakland, 1966); comic strips (Wolfenstein, 1954); private letters (Allport, 1960); elite periodicals (Angell, 1964; Singer, 1964); and editorial material (North, 1965).

An early innovation in quantitative analysis of manifest content was the development by Ralph K. White (1951) of value analysis:

> Value-analysis is a form of content analysis in which a value, the unit of analysis, was defined as '<u>any goal or standard of judgment which in a given culture is ordinarily referred to as if it were self-evidently desirable</u>' (White, 1951, p. 21).
>
> Value-analysis essentially is a method of categorizing what people say about what they want [which] has the advantage of reducing various and diverse materials to somewhat comparable terms. Whether these 'values' reflect sincere beliefs, or how they will be expressed in action, cannot be determined by value analysis as such, but is a matter of inference and interpretation (Eckhardt and White, 1967, p. 326).

Reliabilities reported on the technique vary from the .60's to a correlation of .93 on the ranking of 50 values analyzed in a ten-page autobiography.*

While not widely used, White's form of content analysis does have adequate reliability and has produced at least some evidence of construct validity.** On the other hand, data scanning and analysis can be

---

* Eckhardt has checked his own analysis-reanalysis reliability on the same material over a six-month interval, and has obtained a correlation of .80, significant at the .01 level of confidence, between two sets of frequencies on 32 values. He has also checked the reliability of subject matter in the case of Winston Churchill's speeches, whose expressed values in 1933-1938, 1938-1940, 1940-1941, and 1944 were all correlated with each other about .80, significant at the .01 level of confidence. (Eckhardt, 1965, p. 346). Eckhardt's average reliability with ten other analysts on a variety of material was .84, significant at the .01 level of confidence.

** White's analysis of speeches of Kennedy and Khrushchev serve to confirm Bronfenbrenner's (1961) mirror-image hypothesis of US-Soviet relations. He found that "Kennedy's and Krushchev's overtly stated values were positively and significantly correlated with each other" (Eckhardt and White, 1967, p. 331).

laborious. Evidence of the predictive validity of his approach has not been reported.

Angell (1964) developed a scheme for comparing Soviet and American values. A team of coders compared values expressed in three years of periodicals addressed to six American and six Soviet elite groups, e.g., military, intellectual, political. (An earlier year was used to develop the technique, train raters, etc.) His coders coded materials along 4 to 6 point dimensions of 42 value categories.

According to Angell:

> The greatest difficulty...was undoubtedly the achievement of intercoder reliability. Our experience would indicate that it would take months of training to obtain 90 percent reliability on material such as that coded in this study. (p. 384)
>
> The problem is not in securing reliability on the value position within a dimension once the coders agree that that dimension is being discussed. However, since all degrees of allusiveness are found in written material, it is almost impossible to find a rule that will discriminate cases where a value position is being taken from cases where there is a veiled and uncertain reference to it. (p. 384-385)

The Angell exercise was worthwhile, and postdictive evidence was accumulated of its predictive validity concerning US and Soviet behavior toward each other. However, as indicated above, it was also a most laborious exercise, costly in time if not in money, and has not been, nor does it appear likely to be, widely used by policy makers although scholars find it of interest.

Robert C. North has also developed a form of quantitative analysis of the manifest content of verbal material. He and his team (e.g., Holsti, 1963; Zaninovich, 1962) have analyzed documents from countries in the 1914 pre-World War I crisis, as well as more contemporary materials

produced by the Soviet Union and communist China which bear upon their relations with one another and with the United States. Using situational base lines through a Q-sort technique initially suggested by Milburn and developed by North (1963), positive and negative affect (values), threats, and perceived threats were scored. However, months of training and much hard work were required to get inter-coder reliabilities beyond .50. Validity of the analysis in terms of behavior of the parties has proved exceedingly high. More recently, North has used a computer-based approach to content analysis, a modification of one developed by Stone, et. al, (1966). This last development is expensive at this stage, but promises to be quite useful once adequate dictionaries (one per scale) and other modifications have been developed. The Soviet-Chinese materials also bear close relations to behaviors and politically contextual events within and outside of crisis.

The increased use of high speed computers will make quantitative content analysis practical. The reduction of huge quantities of raw data now requires more time and manpower than most investigators are willing to expend. Computer theory and programming have advanced sufficiently so that fast and reliable analysis should be feasible within a few years.

2. Inferential Content Analysis

A second form of content analysis goes beyond manifest content to infer the values that are latent in the material. Thus, John Weakland (1960) has analyzed and compared two sets of Chinese films, one produced on the mainland and the other from Taiwan, and compared them for manifest and latent themes, values, outcomes, the nature of heroes and villains, etc. The communist Chinese in particular are strongly anti-individualist at both levels of analysis. Weakland makes his inference on the basis of a variant of communication theory developed by him and his colleagues (Bateson, Haley, Jackson; 1956) and upon psychodynamic considerations. His approach is consistent with that of Wolfenstein

(1954), who analyzed latent values and themes in the humor of children. No attempt was made to achieve quantitative indices of reliability in either of these studies although Weakland essentially replicated his studies over a number of films, thus achieving a kind of reliability. The results of these studies are highly plausible but there have been no attempts so far to gather validational data.

3. Predictive Content Analysis

In a third form of content analysis, used during World War II to make various predictions, political and situational contextual factors were employed extensively to provide bases for going beyond the manifest content of the material being studied. After the war it was possible to check the validity of many, though not all, of the inferences through interviews and extensive documentation. Alexander George (1959) describes attempts to articulate and to validate the method of making inferences of values and behavior. Of those predictions which could be checked, some 81 percent were correct although in some areas of prediction results were better or worse than that. Impressive also was the capability of the team of analysts to make correct interpretations of several month-long series of behavior. The inferential connections were made largely upon the basis of theoretical considerations that are primarily psychodynamic.

It should be emphasized that content analysis appears to be an effective way of exploring written material for values. While developmental costs for scoring schemes and computer dictionaries or coder training may be fairly high, costs per document gradually decrease; and total costs may be quite low and time short compared with the costs of doing field work. Content analysis is a particularly effective way--and sometimes the only way--to deal with hostile information sources. Many hypotheses not initially conceived by data collectors can sometimes be checked through content analysis. Purely quantitative approaches to content analysis like those of Stone et. al. (1966) typically yield high

reliability, but the results are harder to interpret than are those which also employ context as a basis for predicting behavior.

C. SURVEY

Surveys are quantitative studies in which standardized information about predefined populations is collected from samples that are large enough to provide bases for statistical inference. The individual is usually the unit of analysis, although the union, the corporation, the community, fraternity, or the tribe, could just as well be the unit. The data must be available in a form to allow primary manipulation by an analyst. Data may be based on face-to-face or telephone interview or mailed questionnaires. The cross-sectional survey is the basic survey design. Survey analysis may be directed toward answering descriptive, correlational, or explanatory questions.*

The survey interview is a truncated version of the interview employed naturalistically. The survey sacrifices many kinds of information for standardization of questions and for the achievement of statistically reliable scales, so that the data collected may prove comparable across persons. The survey is the most popular form of standardized data collection and analysis from representative samples.

Opinion questionnaires tend not to have their items scaled. As a result the relationships of groups of items is not known and the reponses to one question cannot readily and reliably be related to others. Frequently, the across-time or across-item reliability of items is not ascertainable, and other information is generated that is difficult to interpret. Used cross culturally, questionnaires are

---

* Famous studies based on surveys include Campbell, et al, The American Voter (1960); Stouffer, et. al., volumes 1, Adjustments During Army Life, and 2, Combat and Its Aftermath, of the American Soldier (1949-50); and the Kinsey studies (1948, 1953).

confounded by translation, making item and situation context not comparable across cultures. Literal translation loses equivalence of meaning; idiomatic translation is difficult. To achieve even inter-individual comparability, most items should be closed ended. Though such items permit easy comparisons among various cross-tabulations of data, they reflect hypotheses that the investigator establishes before data collection and thus may exclude information. Werner and Campbell (1959) have listed some of these problems and have suggested solutions. Whether closed-ended or open-ended items are used, the interpretation of survey results frequently requires long periods of time. Lerner's (1961) cross-cultural survey of European elites and their attitudes toward weapon systems and related cold war matters took an additional year to interpret after the interviews had been completed, in part because content analysis and scaling of answers had to be accomplished before analysis and interpretation were possible. Hadley Cantril (1962) met the problem in a different way. The base lines he utilized referred only to in-country circumstances; base lines against which the hopes and fears of the people of different countries could be meaningfully compared were not developed.

D. THE INTERVIEW

An interview is an organized, purposive, direct confrontation in which conversation is usually focused upon one person, the interviewee or respondent. Interviews are typically structured for purpose, formality, location, time, duration, and limits of behavior. Information is usually shared asymmetrically: The interviewer asks and the interviewee answers or provides information. In treatment or counseling interviews, the interviewee tells about himself and perhaps about his problems, but he does not expect to learn about the interviewer as he would in an ordinary conversation. In other kinds of interviews, the purpose may be to gather information, to provide information, or to arrive at a decision; the purpose determines the organization, structure, and limits of the interview. This section is limited to the interview used to acquire information about values.

The interviewer overtly or tacitly structures the interview, instructing the respondent about what to expect and to answer. The interviewer is a primary stimulus to the respondent and an important condition for each response. In the absence of electronic recording, the interviewer is a fallible observer and inconsistent recorder of the responses. As an interacting participant, he is part of the situation he both observes and influences. His role may be subject to change in ways that he may not notice or understand. Most articles about the interview (e.g., Krech and Crutchfield, 1948) point out that the mere presence of the interviewer and the way he is perceived by the respondent (role, age, sex, dress, attitudes, etc.) have some usually unknown effect upon the responses. This tends to reduce reliability and validity.

The interview is an exceptionally rich source of data and hypotheses. Cronbach and Gleser (1965), among others, point out that it can be a basis for generating a wide range of hypotheses, none of them necessarily reliable or valid, and that many separate bases for decisions are generated at low cost per hypothesis or prediction. The richness of the interview as a source of data and hypotheses is due in part to its paralinguistic qualities (Weick, 1968). The starts, stops, hesitations, rate of speech, and tonal changes all communicate. So gestures, facial expressions, bodily tension, and spatial orientation vis-a-vis the interviewer, may serve as partially redundant, partially independent, channels of communication, providing information that may be useful or misleading, depending on the training and skill of the interviewer. In addition, the answers and comments upon the interviewer's questions may be concise and specific, or digressive, diffuse, vague, and avoidant.

The interviewer can learn about cognitive categories of the respondent, how his ideas are organized, and the rationalizations of his positions. The interviewer can report the context of the interview in general (e.g., the setting, time of day) or for particular statements (e.g., tone of voice, contiguous events). Such indications of microcontext can improve interpretations and inferences.

Interviews may be divided into two kinds, self-report and informant. In the former, the respondent answers questions about himself, his own behavior, attitudes, or feelings; in the latter he discusses groups or other individuals in various situations, describing their behavior, attitudes, or feelings. Anthropologists (Albert, 1958) typically utilize informants, particularly when values are seen as norms of conduct or as moral or social obligations, whereas sociologists typically seek information about the respondent himself. Thus the sociologist utilizing the self-report interview more often collects data about preferences, expectations, perceptions, or attitudes than about norms. Informant interviews often are presumed to have more reliability because the data are less variable than self-report data across time or within a culture. No validity studies of informant interviews were found except for the data quality control techniques suggested for anthropologists by Naroll (1962).

Sociologists commonly employ the self-report interview, which normally is highly structured; the information sought is limited, which increases reliability and comparability, but sharply curtails the variety of data and hypotheses. These characteristics are discussed in the section on surveys.

E. PROJECTIVE TECHNIQUES

The typical projective test presents an ambiguous stimulus which requires behavior, usually verbal, that presumably stems from the psychodynamics of the respondent. The responses (words, sentences, stories, drawings, or action) are analyzed, usually within a theoretical structure, following specific decision rules (e.g., the presence or omission of attention to detail or structure). The interpretation usually requires considerable time of a highly trained clinician. A few individuals can give convincing demonstrations of uncanny psychological insight using these techniques.

The assumption that responses to projective tests mirror actual behaviors and actual emotional responses seems often to be taken as proven by projective test users. The few attempts to validate projective tests indices, signs, and indicators in behavioral terms on Rorschach, Sentence Completion, or Thematic Apperception Test (TAT) have mostly been unsuccessful. Very little emphasis has been placed on predictive validity. "The adequacy of these measures has been judged by their effectiveness in yielding results consistent with psychoanalytic and learning theory models of personality development and by their face validity" (Pittel and Mendelsohn, 1966, p. 33).

Projective techniques can be useful in inferring what is not manifest in content or structure. Martha Wolfenstein (1954) treats humor as if it represented projective output, somewhat as McClelland (1953) used myths from around the world.

Projective techniques are employed by those who regard indirect measurement as the best and most accurate measurement of values and motives. If a man does not realize what he is telling, he will inadvertently reveal more, it is thought. Such methods are also used by skeptics who, although believing them to be unreliable and of low validity, find them rich sources of hypotheses.

A major disadvantage of projective techniques in their full-length format is that administration, scoring and interpretation are each time-consuming and costly. In their group forms (they may be given as multiple choice questionnaires) they are changed and truncated.

Lindzey (1961) in discussing the cross-cultural replicability of projective tests states:

> Very little is known <u>of the extent to which personality inferences reported by one investigator...</u> can be <u>reproduced by another investigator</u> if he is given the opportunity to observe the same subjects or even the same raw data.

As with content analysis and the interview, knowledge of the context in which communications are elicited yields more interpretable results:

> The importance of 'blind' analysis of projective techniques is easily overestimated... One cannot expect such instruments to function even close to their normal level of effectiveness when the interpretations are made with little or no contextual information concerning the subject, the testing situation, and the culture. In spite of this, it is somewhat surprising...[that not one study has considered] random factors and bias in the comparison of [blind] projective-technique inferences with inferences derived from traditional data sources.... A striking and very surprising shortcoming that characterizes most of the studies we have discussed is a failure to provide a full description of the circumstances under which the test is administered. One might expect that the anthropologist, with his traditional focus upon external constraint and social determinism, would show an unusual sensitivity to the role of situational and cultural influences on response to projective techniques. Not only do the anthropologists fail to give full consideration to these factors, but most of those studies where the testing situation is described relatively satisfactorily have actually been conducted by psychologists, or else they have involved a psychologist as a collaborator. Presumably the resolution to this paradox lies not in the psychologist's appreciation for the general importance of cultural-situational factors, but rather in the anthropologist's lack of appreciation for the extreme importance of immediate or situational factors in determining projective technique response. The anthropologist often appears to have accepted the projective technique as a relatively immutable device that can be administered in a standard manner and scored in an objective fashion, with the consequence of specifiable results quite independent of variation in the surrounding world. As our discussion has already made clear, this is by no means the case; a sensitive and dependable outcome from projective-technique data depends upon a full understanding of the context within which the test was administered (Lindzey, 1961, pp. 297-298).

F. SYSTEMATIC OBSERVATION

Observation is the foundation of empirical science: The subject matter, the methodology, and the purpose of the observations distinguish

the various disciplines. Observing behavior in different settings is one way to discover values. Systematic observation is the principal tool of the anthropologist working as a participant observer who shares native customs and who may have gone partially native. Sociologist Howard Becker (1958) has argued very strongly for careful observation in field settings, but he emphasizes the phenomenological method for understanding the roots of behavior. The psychologist who systematically observes behavior, typically observes only those behaviors that he has previously chosen to observe, thus increasing the likelihood of missing events known to be significant and also decreasing the likelihood of making unexpected and useful discoveries.

Values, then, are inferred from observations, particularly observations of behavior displayed in making choices. Just as interviewers may draw incorrect inferences because respondents lie to them and mislead them, so can observers be mistaken, since a single set of observed behaviors are essentially ambiguous about the motivation that produced them.

Behavior may be predicted from a knowledge of values and then observed, but unless we are told a man's values (and he is both honest and insightful) we cannot know them directly.

G. THE MULTIPLE RESEARCH STRATEGY

A combination of methods can assist in the study of values. When several value dimensions and measures that partially overlap are used, each aspect of value is measured more than once. This combination, in addition to increasing the amount of information (the more dimensions, the more information), increases confidence in the reliability of the measures. The same combination of measures used to check the validity of the concepts and to provide empirical support for the hypotheses can also supply less ambiguous and more solid bases for interpretation of findings.

There is a second gain in using multiple research strategy: Through a multi-method, multi-trait matrix, as described by Campbell and Fiske (1964) in their convergent and discriminant approach to validity, it is possible to discern which methods yield the most and which methods yield the least method-effect (as contrasted with substantive-effect) at different points in the two sequential research strategies of inquiry and verification. The investigator is thus able to be more objective and to make consistent predictions (reliability) with greater accuracy (validity) and utility.

Among the tools which might be used in a multiple research project investigating values are those instruments designed to assess individually held values. Although projective tests can be used for this purpose, the focus is much broader. Two well-known and respected instruments have been used to study individual values: the Allport-Vernon-Lindzey (1960) Study of Values test (A-V-L) and the Strong Vocational Interest Blank (SVIB) (Strong 1943). Both have adequate stability; the SVIB has had remarkable success demonstrating the usefulness of values (restricted here to preferences) as determiners of future behavior when the results have been used to predict job satisfaction and academic success; the A-V-L has been used as a criterion measure and the results, though mixed, have indicated that broad value groupings (e.g., economic, theoretical) are related to specific behaviors. Both tests must be restricted to the populations on which they have been standardized; neither instrument has been usefully applied to determine group values, though either could be.

IV. SELECTED CRITERIA FOR COMPARING METHODS OF VALUE STUDY

The criteria used to compare approaches to the study of values are reliability, validity, and utility. The factors that are critical for effective design of a study (sampling, experimental or quasi-experimental design, levels of significance, etc.) are omitted because they are common to all approaches.

A. RELIABILITY

Reliability may be defined as the consistency with which an instrument measures whatever it is measuring. The reliability of a behavioral measure or set of observations is a sine qua non for other forms of merit. Reliability is measured by a proportion which describes the consistency or agreement of two highly similar measures, or of two (or more) observers at the same instant, or one observer on two (or more) occasions. The common coefficients of reliability can vary from plus one (complete agreement) through zero to minus one (complete disagreement). If the reliability of an instrument or observer does not differ significantly from zero, its agreement with a behavior criterion (validity) can prove no higher. If the interobserver reliability is one, then knowing the observations of one member of a pair of observers permits perfect prediction of the observations of the other. The reliability of psychological instruments or observations is usually estimated by measuring internal consistency (split-half or item-statistic estimates) or consistency over time or trials (test-retest or equivalent-form estimates). In systematic, objective observation or in processes of content analysis a coder must be compared with himself at two points in time or against two comparable samples of behavior or material. Several (and at least two) observers or coders are customarily employed and their independent judgments are compared. If observers both categorize (assign to a category or dimension) and assign to location on a scale or dimension, separate reliability coefficients for the two operations are computed.

High reliability, while a logical prerequisite and limit of validity, does not state what behavior will be predicted by such a measure. A measure

with high reliability may accrue various validity coefficients, some fairly high, some not significantly different from zero. It is obvious that no measure could predict all behavior equally well. A reliability coefficient does not in itself give any hint of the nature of behavioral criterion measures to be predicted, nor does it give any indication of whether the measure under consideration might be said to have utility or a favorable ratio of benefits to costs.

Replicability is one kind of reliability. If an experiment, simulation, or other set of observations is repeated and the outcomes essentially agree with the original, the phenomenon or procedure is described as being reliable or replicable. Some psychological phenomena have been replicated with very high reliability, e.g., the remarkable systematic behavior produced by one or another schedule of reinforcement. Others, such as the para-psychological phenomena, appear to be almost impossible to replicate. Still other phenomena, such as those of repression, may occur regularly in day-to-day behavior or in clinical settings, but have not been produced in controlled form (without confounding artifacts) in the psychological laboratory.

It is not difficult to measure the reliability of observations or judgments of values. Obtaining adequately high reliabilities is sometimes difficult.

B. VALIDITY

Validity may be defined as the degree to which an instrument measures what it is expected to measure. While reliability concerns the extent of similarity between two highly similar sets of observations or measures of the same groups, process or entities, validity* concerns

* Some writers (e.g., Campbell and Stanley, 1963) use the terms internal and external validity to refer respectively to the power of an experimental design and the generalizability of the findings. Those meanings are not intended here, although external validity and construct validity are related.

relations between two different processes or entities, e.g., a group's ratings of truth as a value and ratings of their behavior as truthful. The same coefficients* which are computed for reliability are also computed for validity and vary from zero to plus or minus one. Validity coefficients indicate a degree of relation between two different measures, usually a predictor and a criterion. A set of observations may serve several different uses or purposes; each use requires a separate criterion measure and validity coefficient. While the reliability coefficient (r_{11}) is an indicator of consistency of observation, judgement or measurement, the validity coefficient (r_{12}) is an indicator of the extent of agreement between a set of observations of verbal (or non-verbal) behavior and some different verbal or non-verbal behavior that it is expected to predict or measure.

Several validity terms may be associated with a set of observations.

1. Face Validity

Face validity means that a measure appears to be appropriate, that it should measure what it purports to measure; no demonstration is needed.

2. Intrinsic Validity

Intrinsic validity exists when the predictor is also the criterion, e.g., an algebra test given at the end of a class in algebra may have intrinsic validity.

3. Concurrent Validity

The concurrent validity of a set of observations is its correlation with another set of observations. The concurrent validity of a verbal ethnocentrism scale could be estimated by correlating the scores with judgments or ratings of actual discriminatory behavior toward ethnic groups.

---

* In many studies, the experimenter is not attempting to determine the extent of the relationship, but rather that a particular relationship is not a chance result. In these studies, more sophisticated designs and more powerful statistics can be used.

4. Predictive Validity

Predictive validity typically is harder to achieve than concurrent validity because it concerns the degree of association between observations now and behavior later on. An example might be relations between value preferences expressed now and job performance or job satisfaction at later points in time. For some jobs, validity coefficients between value (or interest) preferences now and outcomes later may be fairly high; whereas for jobs where capacity or talent outweighs value preferences, the latter predict less well. The distinction between concurrent and predictive validity is not always clear. Predictive validity refers always to the relationship between two sets of measurements, both designed to estimate criteria. Validity coefficients ideally would be computed between a predictor and an ultimate criterion, e.g., something a practical man really wants to predict; but often researchers settle for second or third best proximate criteria.*

In the laboratory, hypotheses are tested and measurements validated by the prediction and subsequent observation of cause-and-effect relationships that are under the experimenter's control. This kind of rigorous testing cannot be achieved when survey methods are used under field conditions; however, quasi-experimental designs have been developed (Campbell & Stanley, 1966) and can be used to eliminate rival hypotheses and establish the validity of measurements.

5. Construct Validity

Construct validity refers to the explanatory power of a concept, the degree to which it is unambiguously related to other concepts.

---

* They might, for example, have to settle for a salesman's grades in a sales training course as a criterion, when it is not possible to use as an ultimate criterion his success in the field. There might, however, be little correlation between the two criteria. Contribution of a new tactical missile to the winning of battles may be the ultimate criterion needed in operations research designed to evaluate the missile, but the investigator might settle for more easily obtained measures of merit such as the Circular Error of Probability.

We have just seen that viewed empirically, a useful concept is related to the operations (observations or measurements) that define it; this relationship facilitates predictions, which can be confirmed or disconfirmed. Viewed theoretically, a concept becomes more useful as it is related to other theoretically significant and fruitful concepts; this relationship multiplies the explanatory power of the concept and increases confidence in its meaning and generality. Such conceptual interrelatedness helps to organize as well as to explain data. When a concept (construct) has a pivotal role in a theoretical framework, and when predictions generated by that concept or by its surrounding framework tend to be empirically confirmed, the concept is said to have construct validity.

Different measures of values have different levels of validity. But values recently have been shown to predict not only voting behavior, classroom grades, delinquency, vocational choice, satisfaction, and success, but also more specific attitudes toward particular social objects and situations, including something as complex as behavior rated as creative, and even leadership style. The construct validity of the measures of value has not yet been clearly established, but the number of hypotheses generated which employ the concept, together with the philosophic centrality of the concept, leads us to believe that value estimates will prove to have considerable construct validity.

C. UTILITY AND INFORMATIONAL UTILITY

1. Utility

The term utility, when applied to a scientific concept or a set of observations, refers to perceived usefulness, whether direct or derived, absolute or comparative. Economists use the term to refer to perceived desirability and imply (but do not state) that negative qualities have been taken into account. Here we intend the meaning to be explicit: utility refers to both benefits and costs, which we call

the benefit/costs ratio (without suggesting a simple mathematical operation). The utility of an information-gathering technique is a function of the relevance of the obtained information to a problem and the costs (e.g., manpower, time, and money) of obtaining the information. For any given problem, the information might be pertinent to the solution, might be irrelevant to the solution, or might be misleading in which case utility is negative.

A technique for gathering information about the organization of the values of an individual may be expensive when it takes much time of an expert plus the time of a subject. A projective test such as the Rorschach takes one or two hours to administer and up to seven or eight additional hours to score, interpret, and report--all per single client. The reliability of Rorschach signs and interpretations and the validity of hypotheses produced from the Rorschach are typically quite low. The test is still widely used because it is believed to have much utility in clinical settings: it permits the inference of latent psychodynamics, the organization of the perceptual-conceptual defenses, coping mechanisms, and values.

Cronbach and Gleser (1965) have argued persuasively that techniques such as the Rorschach or open-ended interview and questionnaire can produce many working hypotheses, not readily obtainable in other ways. The techniques may have high utility despite their often low reliability and validity since the cost per hypothesis is low. Cronbach and Gleser contrast these multi-indicator techniques with scales measuring only one concept such as responsibility, attitude toward religion, or intelligence. These latter scales must have high reliability and validity or their utility may tend toward zero under some circumstances.

2. Informational Utility

Information theory suggests that unexpected information has more utility than expected information, which merely serves to increase the

credibility of a source; a technique or theory that produces more unexpected information has more information utility than a technique that produces less. Unexpected information that is relevant to important problems, or to problems that can be handled, has high utility.

Informational utility is different from and partially independent of validity. An instrument may state with great accuracy what it is supposed to state, but the information may already be known; on the other hand, a technique, such as the interview, which tends to be low on reliability and validity may provide information which is unexpected and which the user could have discovered in no other way. Even if such information must for the moment be considered tentative (until validated by other approaches), it may be invaluable. If a theory or a concept suggests interesting questions for research, questions which otherwise would not have been raised, this concept or theory has heuristic worth, i.e., informational utility.

In general, multi-indicator techniques (which Cronbach and Gleser call wide bandwidth techniques) may be expected to have greater informational utility than a univariate technique. Multi-response, open-ended, or partially structured techniques should have more informational utility than convergent, or highly structured ones.

Now that a number of methods for the study of values have been discussed, and criteria for evaluating them have been explained, it is possible to proceed to an explicit evaluation of the methods to see the strengths and weaknesses of each in terms of the criteria.

V. EVALUATION OF ECHO AND COMPARISON WITH OTHER METHODS

The six methods discussed earlier will be compared under each of the three criteria; multi-research strategy will not be included since by definition any combination of investigative strategies might be used, and the relationships to the criteria would differ for each one. Intelligent use of multi-research strategy will usually increase reliability and validity, but utility and informational utility must be estimated for each project. One reason that multi-research strategy, which clearly is a powerful investigative tool, is not regularly used is that the costs may be greatly increased and the utility thereby reduced. Another reason is that investigators tend to use the methods with which they are familiar; thus a survey organization will tend to define all problems so that survey research is the only appropriate investigative tool.

A. RELIABILITY

1. The ECHO Method

Satisfactory coefficients of reliability for group data collection have been obtained on test-retest measures; when responses were collected from the same Ss on two occasions and classified together, conservative reliability estimates were between .71 and .84. These estimates reflect the consistency with which a group of Ss is represented in the same set of categories on two occasions.

The reliability of classification has been variable. Re-sorting of data cards into existing categories yields r's of .90 and above. Reclassifications into new category arrays have been judged as highly equivalent, but the discrepancy between the numbers of categories used for the different classifications has occasionally been so great that the legitimacy of an estimate of correlation is open to question.

Classifiers can usually agree that a given response card belongs in a particular category; the reliability problem arises with the free determination of the number and kind of categories. A team focusing on

abstractions is likely to develop categories that are different from those of a team focusing on concrete variables; the distribution of items by a team using a large number of categories is necessarily different from that of a team using a small number of categories. A reliability coefficient determined under these conditions is ambiguous; in various investigations whenever judgments were made of the semantic comparability of categories, the reliability was high--but a coefficient determined by subjective manipulation does not carry the weight of one determined by mechanical manipulation. (Classification reliability can be deliberately increased by conferences between the two classification teams.)

In general then, the ECHO method yields reliable results, whether one evaluates the consistency of the respondents or of the classifiers; but under some conditions the reliability of a classification cannot be unambiguously determined.

2. Content Analysis

Well-trained analysts, working with pre-determined common categories, can produce adequately reliable data. An "unstructured" classification, such as ECHO uses, is normally not used in content analysis; if it were, the reliability would undoubtedly be very low. Naive indigenous classifiers, working with ECHO data, provide reliability coefficients of .90 and up when classifying data into existing categories; only highly trained content analyzers can approximate the same level.

3. Survey

Under good conditions survey data can be highly reliable. However, many pitfalls exist, some of them not known, which require careful design and continuous monitoring. Adequate reliability can never be assumed. The data collected in surveys vary from quite objective (age, sex, height, etc.) to completely subjective (opinions, attitudes, etc.). Reliability

tends to drop as the data become more subjective. Since the phrasing of the questions can have great impact on the answers, reliability across surveys is difficult to assess. (ECHO questions, though variable, are much more stable than survey questions.)

The reduction and analysis of survey data obtained in quantitative form is highly reliable. The more open-ended the questions, the less reliable the analysis.

4. Interview*

Studies of the interview indicate very low reliability for data collection and interpretation. Data reduction and analysis can be reliable if relatively simple categories are predetermined and the raters are well-trained. If the interviews are extensive, content analysis can be used, achieving the same kind of data reduction reliability described in that section.

5. Projective Techniques

For both data collection and data analysis, inter-investigator reliability is generally low, but acceptable levels are possible with very skilled and uniformly trained analysts. In addition, intra-subject reliability is also low, probably because projective tests are particularly sensitive to transitory fluctuations in mood and set of the respondents.** Evidence for inter-group reliability is non-existent.

6. Systematic Observation

The concept of reliability is meaningless for systematic observation as a general procedure because of the many variations in procedure and conditions. This much, however, can be inferred from a knowledge

---

* The interview used to collect survey data will be discussed in the section on surveys.

** True also for ECHO, but ECHO is designed for group analysis and the responses for a given group have high reliability from one occasion to another.

of psychology: data collected by observation tends to be quite unreliable, unless well-trained observers are agreed on exactly what is to be observed and the probable psychological distorting mechanisms are known and controlled. Thus, two job analysts can produce reliable information about relatively simple jobs in which they are not emotionally involved; but, it is unlikely that two social scientists conducting phenomenological investigations of a culture would present the same findings, other than about the most compelling phenomena, and even these could be grossly different. A team of scientists, working together, could agree about their findings, and the appearance of high reliability could be achieved. However, if the scientists undergo careful training and are agreed in advance about what to look for and how to interpret what they find, then adequate reliability can really be achieved.

B. VALIDITY

1. The ECHO Method

Although not enough validity estimates have been attempted to allow a confident statement about validity, the high reliability of the method means that high validity is at least possible. Two estimates of concurrent validity show excellent promise. Studies of predictive and construct validity are in process; preliminary findings are satisfactory.

2. Content Analysis

A few studies have shown acceptable concurrent validity. No information is available for predictive or construct validity. Since the evidence for reliability is unclear, no prediction of validity can be made.

3. Survey

The high validity coefficients found for survey data really validate the sampling procedures, since many surveys (e.g., voting behavior), are designed to estimate the quantity of some variable in a population, when the situation makes the total population unavailable because of cost or some other factor.

Other predictions, such as consumer or political preferences, have shown remarkable variation in validity, e.g., the marketing of the Edsel as contrasted with the Mustang, or the J.F. Kennedy presidential primary campaign in West Virginia as contrasted with the R. M. Nixon gubernatorial campaign in California. All of these were based on surveys. Although it is possible that judgment was the controlling factor, it is more likely that the specific questions asked yielded in one case useful, in the other misleading, information.

4. Interview

Low reliabilities mean that overall high validities are not possible. Specific cases may show high validity, but the prediction of which ones are valid is based solely on faith.

5. Projective Techniques

Projective techniques should be considered to have unknown validity. They are primarily useful for the generation of hypotheses; validity under these conditions is almost irrelevant since the criterion measure would be the veridicality of the hypothesis. Those studies that have used projective techniques to assay groups have not presented evidence of validity.

6. Systematic Observation

The concept of validity is appropriate only for each particular case or kind of observation. Reliability, described as mixed, determines the upper limits of validity; in addition, the factors described in the section on reliability apply independently to validity criteria. Observation, used to estimate values, is particularly susceptible to psychological variables. If observation continues for a long period of time, the observer is increasingly prone to become involved and non-objective, and validity will be reduced. When observation, per se, has been measured for validity, the resulting coefficients have been low for everything except simple counting or identification.

C. UTILITY AND INFORMATIONAL UTILITY

1. The ECHO Method

Utility is probably high. The cost of data collection is low whenever groups are available but is considerably higher when the target population must be surveyed as individuals. The cost of data analysis is low and the potential benefits are high. The larger the population to be predicted the greater the potential utility, since samples can remain relatively small.

Informational utility can be high. Large scale and pilot studies show that usable unexpected information is developed.

2. Content Analysis

At the present time, utility is low. Data collection costs are frequently low, but cost of analysis is high. Under conditions when content analysis is the only method possible, the concept of utility may be meaningless. The development of computer analysis will increase the utility.

Informational utility is moderate. The analyst must work with a limited array of uniform concepts and the probability of getting unexpected information is low. However, usable information could be elicited with the proper design.

3. Survey

For certain questions, utility is high; for other questions it is very low. Since the survey can give misleading information, a poorly done or inappropriate survey can lead to results that are worse than chance. Utility, under these conditions, would be negative.

Informational utility is low for unexpected information since surveys normally are designed to yield only expected results, but it should be high for usable information, for the same reason.

4. Interview

Utility is probably low. The cost can be high and the benefits, as with systematic observations, are not known.

Informational utility is not determinable but is probably high, unless the interview is completely structured. The interview is potentially the richest source of unexpected information.

5. Projective Techniques

Utility is low. The cost of data collection and analysis is normally high and the benefits are unknown.

Informational utility is not determinable but probably is high. Unexpected and potentially usable information may be generated but only in the form of hypotheses to be tested.

6. Systematic Observation

Utility is variable. The cost can vary from low to high and the benefits, because of the unknown reliability, cannot be estimated until after decisions are made.

Informational utility can be very high. Paradoxically, the more systematic the observation, the more likely that informational utility will be low. If observation is restricted to predetermined categories, the observer is restricted. However, a skilled observer can be both systematic and open.

APPENDIX

SOME HYPOTHESES ABOUT VALUES

PREMISE: Values are arranged hierarchically for individuals or groups.

1. Hypothesis: Higher ranking values will outweigh lower ones when choices involved are relevant to both.

2. Hypothesis: The ordering of values will predict attitudes if Ss are aware of fact and value consistency. Value-attitude discrepancies will be resolved in favor of values.

3. Hypothesis: Changes in the skills, talents or capacities of a group will be the best single predictor of changes in value orderings.

4. Hypothesis: Changes in the incentive structure within an organization will be the next most effective predictor of changed value orderings.

5. Hypothesis: Women who are comfortable in roles subordinate to men and who perform well there will have different value orderings, but not necessarily different values, than women without such tendencies or inclinations.

6. Hypothesis: Values are cognitive and are influenced by affect. If the cognitive content of a value remains constant while affective investment goes up (one feels more emotionally committed to a value) that value will rise in the value hierarchy.

7. Hypothesis: Values seen as essential (necessary to the preservation of a culture or person) will be higher in a value hierarchy than values regarded as less essential.

8. Hypothesis: Values which are regarded as important, but not essential, to the preservation of a group or person will tend to stand high in rank though less high than ones seen as essential.

9. Hypothesis: In general, instrumental values (good to do) which are high in a value hierarchy (rated highest on semantic differential potency, activity, and evaluation factors, for example) will follow the classic "J" curve (Allport, 1923) of behavior distribution, and these values will show the J-curve effect far more than will values rated lower on value hierarchy.

10. Hypothesis: Values higher in the value hierarchy of an individual or group will resist change more than will values lower in the hierarchy.

11. Hypothesis: Values high in a value hierarchy can be employed more effectively to change values lower in the value hierarchy than conversely.

12. Hypothesis: If two groups have similar values, differently ordered, the groups will tend to be incompatible.

13. Hypothesis: Events representing terminal (good to happen) values that are high on the value hierarchy will reinforce those behaviors perceived as necessary for their occurrences.

14. Hypothesis: Good to happen (ECHO) or consummatory (Rokeach) values that are unrealized will motivate behavior strongly even when relatively low in the value hierarchy.

PREMISE: Although values tend to have either a prescriptive quality (ought to do) or a desired quality (like to do), the two qualities are strongly related.

15. Hypothesis: When behavior is perceived as causing conflict between prescriptions (what one ought to do) and preferences (what one would like to do) the frequency of such behavior will decrease.

16. Hypothesis: Prescriptions and preferences will tend to be in harmony. Value prescriptions will tend to predict what people like to do (i.e., ought will predict prefer) and vice versa.

17. Hypothesis: Agreement between preferred and prescribed behavior will considerably increase the probability that such behavior will occur.

18. Hypothesis: The "J" curve of conformity will be sharper when group values support institutional norms, e.g., being on time to work (institutional) when workers place high value on their jobs.

PREMISE: Values are more general than attitudes.

19. Hypothesis: Values will show wider predictive applicability than attitudes.

20. Hypothesis: Semantic differential activity (factor III of Osgood's semantic differential) will correlate with saliency

of value, whereas potency and evaluation correlate with ranked importance.

PREMISE: Hierarchical value orderings can be used to determine various behaviors among groups.

21. Hypothesis: Adult socialization (the learning and internalizing of new adult roles) typically will involve both shifts in the rank ordering of some values and the appearance or disappearance of others.

22. Hypothesis: Given the capacities or abilities of a group or organization, values and value orderings will predict preferences for persons (friendships), activities, and effectiveness.

23. Hypothesis: The perception of dominant family dyads will correlate with value orderings more closely than with the presence or absence of values.

24. Hypothesis: Value categories that appear uniquely for a group whose value ordering otherwise resembles those of other groups, will prove particularly useful, i.e., possess high informational utility, for predicting behaviors of the groups. (An example of this may be the unique categories appearing in the Howard University data, i.e., Good to do "Have pride of Race;" bad to do "deny race." While these items did not predict the fact of conflict with the University administration [ECHO was administered over six weeks before the sit-in] they clearly reflected the issues.)

25. Hypothesis: Differences in value ordering between parents and their children will predict intergenerational conflict.

26. Hypothesis: Most people value the approval of others, but the rank of this value will vary considerably both within and between groups.

27. Hypothesis: Groups that do well with one another, and like one another in spite of heterogeneous values will tend to be particularly adaptable and creative.

28. Hypothesis: Similarities of value orderings will predict compatibility between groups more accurately than will similarities of content of beliefs.

29. Hypothesis: Value orderings will differentiate creative from conforming and non-creative groups.

30. Hypothesis: Two groups with different value orderings will tend to find communication difficult.

31. Hypothesis: Values predict morale: if two groups have similar value hierarchies, but the members of one group have more homogeneous value orderings, the more homogeneous group will show higher morale.

32. Hypothesis: Shared value orderings will play a larger role in predicting friendship or compatibility than will values themselves.

33. Hypothesis: Chinese, Japanese, and Soviet value orderings will be more similar to one another than to US or British value orderings (e.g., respect, authority, history, individualism).

PREMISE: Credibility of messages is in part a function of the perceived power of the sender.

34. Hypothesis: Messages from powerful sources of praise and good events should appear credible across a distribution of messages.

35. Hypothesis: Messages from powerful sources of praise and censure and of good and bad events will appear even more credible across a distribution of messages.

PREMISE: Values affect and are affected by crisis and stress situations.

36. Hypothesis: Crises will temporarily increase the visibility of certain values, thereby increasing their hierarchical levels.

37. Hypothesis: Long-term or particularly intense crises will tend to change value orderings permanently.

38. Hypothesis: Threats to particular values serve to heighten the saliency of those values.

39. Hypothesis: Values influence crisis decisions but do so most effectively when information is ambiguous or inadequate.

40. Hypothesis: In the limiting situation of no information, values alone may serve to suggest reactions to ambiguous stress.

REFERENCES

Abelson, R.P. Modes of resolution of belief dilemmas. Journal of Conflict Resolution, 1959, 3, 343-352.

Albert, E.M. The classification of values: a method and illustration. American Anthropologist New Series, 1958, 221-248.

Allinsmith, W. Moral Standards: II The learning of moral standards. In D.R. Miller and G.E. Swanson (Eds.), Inner conflict and defense, New York: Holt, 1960, 141-176.

Allport, F.H. Social Psychology. New York: Houghton Mifflin Co., 1924.

Allport, G.W., Vernon, P.E., and Lindzey, G. A study of values. Boston, Massachusetts: Houghton Mifflin, 1960.

Angell, R.C. Social values of Soviet and American elites: content analysis of elite media. Journal of Conflict Resolution, 1964, 8, 329-385.

Atkinson, J.W. Motives in fantasy, action and society: a method of assessment and study. Princeton, New Jersey: Van Nostrand, 1958.

Barthol, R.P. and Bridge, R.G. The ECHO multi-response method for surveying value and influence patterns in groups. Psychological Reports, 1968, 22, 1345-1354.

Bateson, G., Jackson, D.D., Haley, J., and Weakland, J. Toward a theory of schizophrenia. Behavioral Science, 1956, 1, 251-264.

Bavelas, A. A method for investigating individual and group ideology. Sociometry, 1942, 5, 371-377.

Becker, H.S. Problems of Inference and Proof in Participant Observation. American Sociological Review, 1958, 23, 652-660.

Bradburn, N.M. and Caplovitz, D. Reports on happiness. Chicago, Illinois: Aldine, 1965.

Bronfenbrenner, U. The mirror-image in Soviet-American relations. Journal of Social Issues, 1961, 17, 45-46.

Burton, R.V. Some factors related to resistance to temptation in four-year-old children. Unpublished doctoral dissertation, Harvard University, 1959.

Campbell, D.T. and Fiske, D.W. Covergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 1959, 56, 81-105.

Campbell, A. et al. The American Voter. New York: Wiley & Sons, Inc., 1960.

Campbell, D.T. and Stanley, J.C. Experimental and quasi-experimental designs for research and teaching. In N.L. Gage (Ed.), Handbook of research on teaching: Chicago, Illinois: Rand McNally & Co., 1963, 171-241.

Campbell, D.T. and Stanley, J.C. Experimental and quasi-experimental designs for research. Chicago: Rand McNally, 1966.

Cantril, H. and Free, L.A. Hopes and fears for self and country: the self-anchoring striving scale in cross-cultural research. American Behavioral Scientist, 1962, 6, supplement.

Cronbach, L.J. and Gleser, G.C. Psychological tests and personnel decisions. Urbana, Illinois: University of Illinois Press, 1965.

Dewey, J. Theory of valuation. In International encyclopedia of unified science. Chicago, Illinois: University of Chicago Press, 1939, 2, 72, 75-78.

Eckhardt, W. and White, R.K. A test for the mirror-image hypothesis: Kennedy and Krushchev. Journal of Conflict Resolution, 1967, 11, 325-334.

Eckhardt, W. War propaganda, welfare values, and political ideologies. Journal of Conflict Resolution, 1965, 9, 345-358.

Festinger, L. A theory of cognitive dissonance. Evanston, Illinois: Row, Peterson, 1957.

George, A. Propaganda analysis: a study of inferences made from Nazis in World War II. Evanston, Illinois: Row, Peterson, 1959.

Grinder, R.E. Parental childrearing practices, conscience, and resistance to temptation of six grade children. Child Development, 1962, 33, 802-820.

Hagen, E.E. On the theory of social change. Homewood, Illinois: The Dorsey Press, Inc., 1962.

Hartshorne, H. and May, M.A. Studies in the nature of character: Volume I, studies in deceit. New York: McMillan, 1929.

Havighurst, R.J, and Neugarten, B.L. American Indian and white children, a sociopsychological investigation. Chicago: University of Chicago Press, 1955. 84-123.

Herzberg, F., Mausner, B. and Snyderman, B. The motivation to work. New York: Wiley, 1959.

Holsti, O.R. The values of international tension measurement. Journal of Conflict Resolution, 1963, 7, 608-617.

Kalhorn, J. Values and sources of authority among rural children. In K. Lewin, C.E. Meyers, J. Kalhorn, M.L. Farber and J.R.P. French, Authority and frustration. Iowa City: University of Iowa Press, 1944, 101-151.

Kinsey, A.C. et al. Sexual behavior in the human male. Philadelphia, Pennsylvania: Saunders, 1948.

Kinsey, A.C. et al. Sexual behavior in the human female. Philadelphia, Pennsylvania: Saunders, 1953.

Krech, D., and Crutchfield, R. Theory and problems of social psychology. New York: McGraw-Hill, 1948.

Lasswell, H. A provisional classification of symbol data. Psychiatry, 1938. 1, 197-204.

Lerner, D. Europe and Polaris: a case study. European attitudes toward weapons system characteristics. In D. Lerner and M. Gorden, (Eds.), European community and Atlantic security. Cambridge, Massachusetts: Massachusetts Institute of Technology, Center of International Studies, 1961.

Lindzey, G. Projective Techniques and Cross-Cultural Research. New York: Appleton-Croft, 1961.

Maccoby, E.E. The generality of moral behavior. Paper read at American Psychological Association, Cincinnati, Ohio, 1959.

McClelland, D.C. The achievement motive. New York: Appleton-Century-Croft, 1953.

Mills, J. Changes in moral attitudes following temptation. Journal of Personality. 1958, 26, 517-531.

Morris, C.W. Varieties of human value. Chicago, Illinois: University of Chicago Press, 1956.

Naroll, R. Data quality control: a new research technique. New York: Free Press, 1962.

North, R.C., Holsti, O.R., Zaninovich, G.M. and Zinnes, D.A. Content analysis: a handbook with application for the study of international crisis. Evanston, Illinois: Northwestern University Press, 1965.

North, R.C., Triska, J.F. and Holsti, O.R. Studies in deterrence: the analysis of international tension. China Lake, California: NOTS, 1963 (NOTS TP 3154).

Osborn, F.W. The ethical content of children's minds. Educational Review. 1894, 8, 143-146.

Osgood, C.E., Suci, G.J. and Tannenbaum, P.H. The measurement of meaning. Urbana, Illinois: University of Illinois Press, 1957.

Pittel, S.M. and Mendelsohn, G.A. Measurement of moral values: a review and critique. Psychological Bulletin, 1966, 66, 22-35.

Premack, D. Toward empirical behavioral laws: I. positive reinforcement. Psychological Review, 1959, 66, 219-233.

Rokeach, M. The nature of attitudes. In D. L. Sills (Ed.), International encyclopedia of the social sciences. New York: MacMillan, 1968.

Schwartz, M.M., Jenusaitis, E., and Stark, H. Motivational factors among supervisors in the utility industry. Journal of Personnel Psychology, 1963, 16, 45-53.

Sears, R.R., Race, L. and Alpert, R. Identification and child training: the development of conscience. Interim research report. Presented at the American Psychological Association, Chicago, Illinois.

Siegel, S., Siegel, A.E. and Andrews, J. Choice, strategy and utility. New York: McGraw-Hill, 1964.

Singer, J.D. Soviet and American foreign policy attitudes: content analysis of elite articulations. Journal of Conflict Resolution, 1964, 8, 424-485.

Stone, P.J., Dunphy, D.C, Smith, M.S. and Ogilvie, D.M. The general inquirer: a computer approach to content analysis. Cambridge, Massachusetts: The M.I.T. Press, 1966.

Strong, E.K., Jr. Vocational interests of men and women. Stanford, California: Stanford University Press, 1943.

Stouffer, S.A. The American soldier. Princeton University Press. 1949-50.

Triandis, H.C. Toward an analysis of the components of interpersonal attitudes. University of Illinois Technical Report No. 35, 1966.

Weakland, J.H. Studies in deterrence: XIV Chinese political and cultural themes; a study of Chinese Communist films. China Lake, California: NOTS, 1966 (NOTS TP 4029).

Weick, Karl. Systematic observation. In Lindzey and Aronson (Eds.), Handbook of Social Psychology (Revised Edition), Volume II, Chapter 13. Reading: Addison-Wesley, 1968.

Werner, O. and Campbell, D.T. Translating, working through interpreters, and the problem of decentering. In R. Naroll and R. Cohen (Eds.). A handbook of method in cultural anthropology. New York: American Museum of Natural History, 1959.

White, R.K. Value analysis: the nature and use of the methods. Glen Gardner, New Jersey: Libertarian Press, 1951.

Williams, R.M. Values. In Sills, D.L. (Ed.), International encyclopedia of the social sciences. New York: MacMillan, 1968.

Wolfenstein, M. Children's humor: a psychological analysis. Glenco, Illinois; The Free Press, 1954.

Zaninovich, G. Pattern analysis of variables within the international system: the Sino-Soviet example. Journal of Conflict Resolution, 1962, 6, 253-268.

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* General Research Corporation P.O. Box 3587, Santa Barbara, California 93105	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP

3. REPORT TITLE

The ECHO Method and the Study of Values

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

Technical Memorandum

5. AUTHOR(S) *(First name, middle initial, last name)*

T. W. Milburn, R. P. Barthol, R. deMille

6. REPORT DATE November 1968	7a. TOTAL NO. OF PAGES 68	7b. NO. OF REFS 65
8a. CONTRACT OR GRANT NO. DAHC04-68-C-0018 b. PROJECT NO. ARPA Order 965, Amendment 2 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) TM-951 9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*	

10. DISTRIBUTION STATEMENT

Cleared for open publication 4 February 1970, Directorate for Security Review (OASD-PA), Department of Defense.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Advanced Research Projects Agency The Pentagon Washington, D.C.

13. ABSTRACT

Values are subjective estimates of the worth or desirability of any entity or event, whether concrete or abstract. Explanation and prediction of human behavior can be improved by including value statements among the observed behaviors.

ECHO is a new method of studying values; it differs from most other methods by allowing the respondent great latitude of statement, by relating stated values to stated causes or social influences, and by eliciting multiple statements. Though its antecedents go back seventy-five years, only with the advent of computerized data analysis has the collection of such heterogeneous and complex behavioral data become generally useful. The information provided by ECHO complements the information provided by polls or surveys that use more specific and, usually, more numerous questions.

The concept value and related concepts are defined; the analysis of values and their place in behavioral research is treated. ECHO and five other methods of value study are described and then compared against the criteria of reliability, validity, and utility. Forty testable hypotheses are listed for future exploration of the concept value. Sixty-five references are listed.

DD FORM 1 NOV 65 1473

UNCLASSIFIED

Security Classification

14 KEY WORDS	LINK A ROLE	LINK A WT	LINK B ROLE	LINK B WT	LINK C ROLE	LINK C WT
ECHO Method						
Value Statement						
Survey						
Questionnaire						
Multiresponse						
Social Influence						
Causation						