Beyond the Two Disciplines of Scientific Psychology

LEE J. CRONBACH *Stanford University*

The historic separation of experimental psychology from the study of individual differences impeded psychological research. So I argued when last I had occasion to address the APA audience (Cronbach, 1957). It was time, I said, for the manipulating and the correlating schools of research to crossbreed, to bring forth a science of Aptitude × Treatment interactions (ATIs).

As that hybrid discipline is now flourishing, a progress report on ATI studies is the appropriate first business of this article. It is not practical to treat here the studies of ATIs in social behavior (e.g., Fiedler, 1973; McGuire, 1969), ATIs in response to drugs and therapy (e.g., Insel & Moos, 1974; Lasagna, 1972; Schildkraut, 1970), or ATIs in learning and motivation generally. I confine myself to ATIs related to instruction, drawing on a comprehensive review Richard Snow and I have just completed (Cronbach & Snow, in press). In that field, several research programs have brought us a long way; particularly to be acknowledged are the sustained inquiries of Bill McKeachie, Jack Atkinson, Russ Kropp and Fred King, George Stern, David Hunt, Victor Bunderson and Jack Dunham, and Snow and his graduate students.

Important as ATIs are proving to be, the line of investigation I advocated in 1957 no longer seems sufficient. Interactions are not confined to the first order; the dimensions of the situation and of the person enter into complex interactions. This complexity forces us to ask once again, Should social science aspire to reduce behavior to laws?

This article was presented as a Distinguished Scientific Contribution Award address at the meeting of the American Psychological Association, New Orleans, September 2, 1974.

I thank Denis Phillips, Robert Calfee, and Lee Shulman for critical comments. My thinking owes much to Richard Snow's enthusiastic collaboration over the years, but I do not speak for him in this article.

Requests for reprints should be sent to Lee J. Cronbach, School of Education, Stanford University, Stanford, California 94305.

Some 30 years ago, research in psychology became dedicated to the quest for nomothetic theory (Hilgard & Lerner, 1951; Koch, 1959; Merton, 1949). Model building and hypothesis testing became the ruling ideal, and research problems were increasingly chosen to fit that mode. Taking stock today, I think most of us judge theoretical progress to have been disappointing. Many are uneasy with the intellectual style of psychological research (Gergen, 1973; Glass, 1972; Israel & Tajfel, 1972; McGuire, 1973; Newell, 1972). Here I shall cut short my comments on ATIs as such, in order to join in that discussion. I shall express some pessimism about our predominant norms and strategies and offer tentative thoughts about an alternative style of work. My sense of the importance of this discussion is heightened by Don Campbell's Lewin Memorial Award address (see Campbell, in press). I would not accuse him of agreeing with me, but if you put our articles side by side you have a binocular view of the scene.

First, we can take a look at ATIs themselves. The typical ATI study is a two-group experiment. The measure of outcome is regressed onto a score recorded prior to treatment. If the regression lines in the two treatments differ in slope, that is evidence of Aptitude × Treatment interaction. Treatment A in Figure 1, though best on the average, is not *equally* superior all along the aptitude scale.

Snow and I give a general meaning to the term *aptitude,* letting it embrace *any* characteristic of the person that affects his response to the treatment. To illustrate research designs and results, I shall describe work by Domino and by Majasan. Both of these have to do with stylistic variables rather than abilities.

Personal Style or Belief as an Interacting Variable

Domino (1968) investigated the relation of college success to the Ai (Achievement via Independence)

and Ac (Achievement via Conformance) scores of Gough's California Personality Inventory. A student scores high on Ai if he says, in effect, "I do good work when I can set tasks for myself." Domino anticipated that instructors who dominated the class—who pressed for conformity—would get poorer results from high-Ai students than instructors who pressed for independent work. A naturalistic review of student grades confirmed this. And it confirmed the reverse relation for high-Ac students, the ones who say "I do well in meeting requirements others set for me." Domino (1971) added to the evidence with a manipulative experiment. He assembled four classes, filling two with High Ai's and two with High Ac's. The same instructor taught introductory psychology to all four sections, pressing for conformity in two and for independence in the other two. Outcomes were indeed better when the student's style of learning matched the instructor's press (Table 1). All but one of Domino's dependent variables showed the ATI he had predicted. (The exceptional variable was a measure of originality. On this, the independent students had an advantage no matter how the class was operated.) Others have found rather similar results (Dowaliby & Schumer, 1973; McKeachie, Isaacson, & Milholland, 1964, Sec. VI-A-3). And, according to a personal communication, Goldberg's (1972) large experiment turned up such an interaction in only one of two courses.

A quite different hunch about instructor–student match was pursued by Majasan (1972). He suspected that an instructor communicates better to students whose beliefs on key matters concur with his. In introductory psychology, beliefs about the intellectual character of psychology would be pertinent. Majasan developed a short bipolar scale, each item offering a "behavioristic" (B) and a "humanistic" (H) alternative. For example:

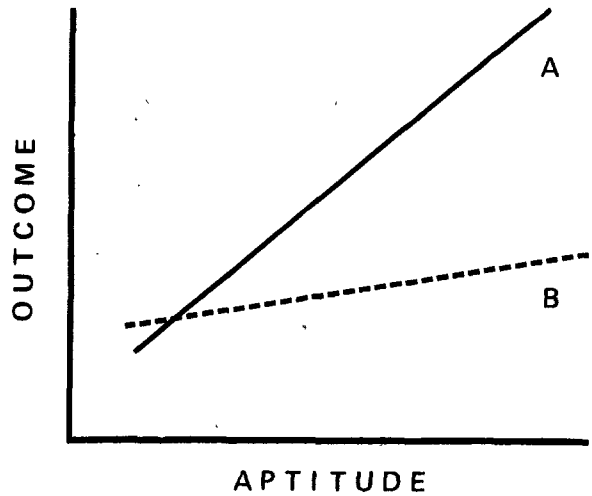

Figure 1. Regressions within two treatments.

The central focus of the study of human behavior should be:

(a) the specific principles that apply to unique individuals. (H)
(b) the general principles that apply to all individuals. (B)

(a) People's observable actions capable of objective interpretation should be the primary concern of psychology. (B)
(b) Psychologists should be primarily concerned with the subjective experiences underlying people's actions. (H)

The instructor and the students filled out the scale at the start of the course. Majasan predicted that students who responded much like the instructor would do best. The criterion was the student's total score over all the course examinations; these examinations were usually assembled from multiple-choice items provided by the textbook publisher.

TABLE 1

Outcomes of a Psychology Course under Four Combinations of Personality and Treatment

Student pattern	Instructor press	Mean outcome (rescaled)[a]			
		Exam	Course grade	Originality of thought	Student satisfaction
Independent (High Ai, Low Ac)	Independence	98	100	99	100
	Conformity	87	83	100	88
Conforming (Low Ai, High Ac)	Independence	78	66	65	82
	Conformity	100	89	59	94

Note. Data are from Domino (1971).
[a] The value of 100 was assigned to the average score of the highest ranking group, and other averages were scaled proportionately. Here, Exam combines a factual multiple-choice test and a quality score for the final essay exam that Domino (1971) reported separately. Likewise, two satisfaction measures have been pooled.

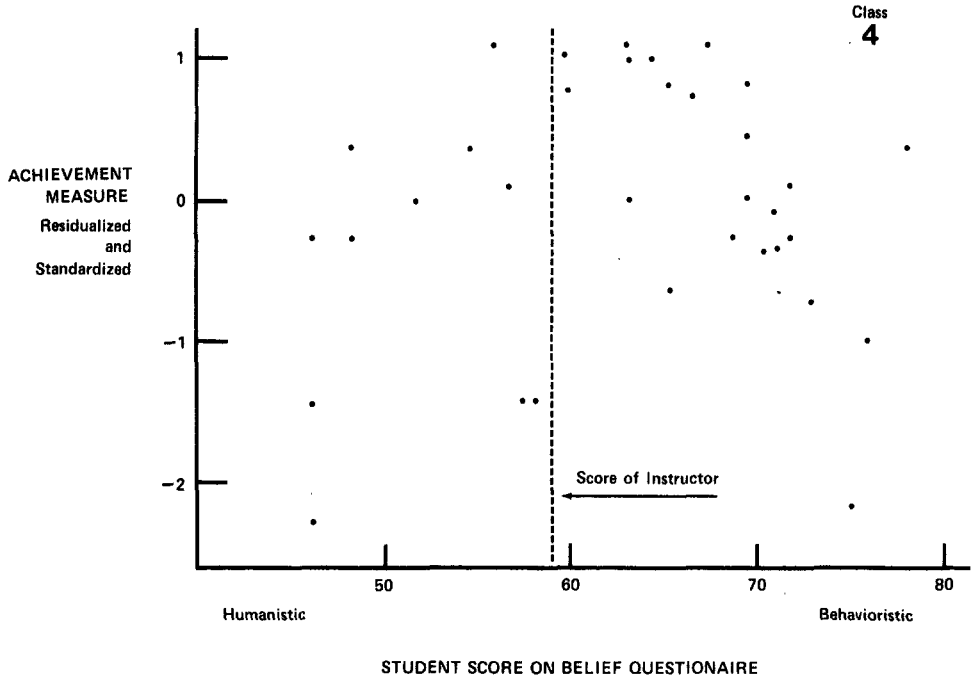

Figure 2. Outcome in a psychology class as a function of the student's belief (after Majasan, 1972).

In Figure 2 we see results on one class. Instructor 4 held intermediate views. Hence Majasan expected peak achievement from middle-of-the-scale students, and that is what he found. The interaction hypothesis is that the arch will shift to the left with a humanistic instructor, to the right with a behaviorist. Examining data from 12 classes in various colleges, Majasan found the trend he had predicted in 11 of the classes. The regression lines in Figure 3 are representative. (The one exception, which is not shown here, was a class in which no examinations were given. The only available criterion, the grade on an independent project, had no relation to the predictors.)

Some Broad Conclusions about ATI

Let me turn next to some synoptic statements about ATI findings. It is not possible to qualify these statements adequately, nor to describe the strength of the evidence.

Quite a lot of work shows student personality interacting with teacher press (see Cronbach & Snow, in press, chaps. 12–13). Thelen (1967) established that the traits that make a high school student "teachable" vary idiosyncratically from teacher to teacher. McKeachie and his colleagues (McKeachie, Isaacson, & Milholland, 1964; McKeachie, Milholland, Mann, & Isaacson, 1968) found many interactive relations between instructor character-

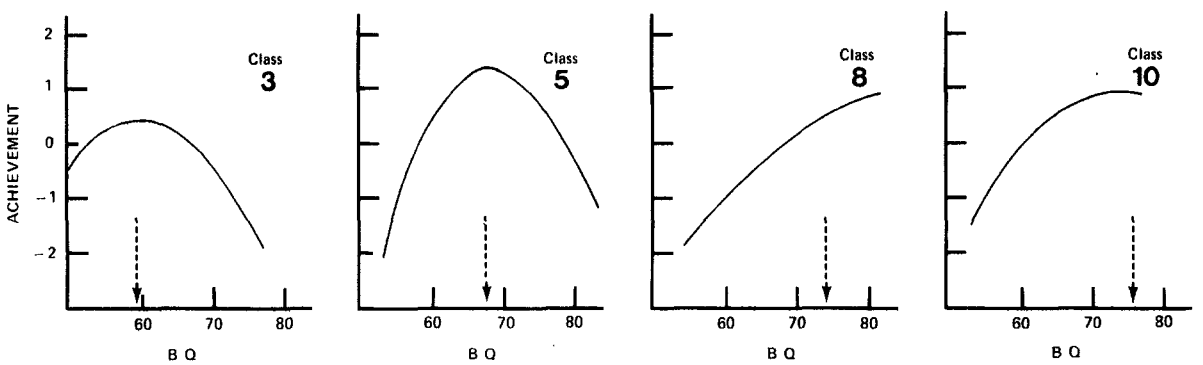

Figure 3. Relation of outcome to student belief in four representative classes (data from Majasan, 1972). (Arrow indicates instructor's BQ score.)

istics and the college student's need for power, need for affiliation, etc. But results were strangely inconsistent from year to year and from course to course. Some effects were significantly moderated by sex or ability of the student. Insofar as a generalization can be glimpsed through the tangle of evidence, it is this: The constructively motivated student (who seeks challenges and takes responsibility) is at his best when an instructor challenges him and then leaves him to pursue his own thoughts and projects. In contrast, the defensive student tends to profit when the instructor lays out the work in detail.

As for abilities, the interactions did not turn out as we had anticipated. If we are to have warrant for instructing some students one way and some another, regression slopes *must* differ from treatment to treatment. When Goldine Gleser and I first came to realize this, we forecast that traditional scholastic aptitude would not be a source of interactions (Cronbach & Gleser, 1957, pp. 125–127). Just because it is general, we expected it to have the same predictive validity for almost any kind of instruction. Only specialized aptitude measures could be expected to forecast differential success, we thought.

The special-ability hypothesis got off on the wrong foot. Between 1960 and 1970, many of us searched fruitlessly for interactions of abilities in the Thurstone or Guilford systems. One hypothesis Snow and I pursued ran like this: "High spatial ability makes for success when the instruction uses diagrams as much as possible, and minimizes words." No interaction of this sort was found, in our shop (Markle, 1968) or elsewhere. With hindsight, we can see that the low-spatial student ought to profit from diagrams if they display relations that the high-spatial student can visualize without help. Conversely, the diagrams can put the high-spatial student ahead when the instructional diagram is complex and has to be transformed mentally to be understood. This more closely reasoned hypothesis has not been tested. We expect a close scrutiny of cognitive processes to be a profitable next phase of work on ATIs.

Contrary to our original view, conventional tests of mental ability or level of educational development do interact. They predict how much is learned from most instruction of fixed duration; but whether the regression slope is steep or shallow depends on the instructional procedure (Cronbach & Snow, in press, chaps. 5–11). One way to reduce the effect of general ability is to bring in pictures or diagrams. Another is to make lessons more didactic, less inductive. On the whole, the regression of outcome onto general ability tends to be relatively steep when the instruction requires the learner to actively transform information, and it tends to be shallow when the demands are less. But the generalization is weak, with many studies running counter to the trend.

Inconsistencies as Higher Order Interactions

In attempting to generalize from the literature, Snow and I have been thwarted by the inconsistent findings coming from roughly similar inquiries. Successive studies employing the same treatment variable find different outcome-on-aptitude slopes. Some fraction of this inconsistency arises from statistical sampling error, but the remainder is evidence of unidentified interactions (McGuire, 1968).

When ATIs are present, a general statement about a treatment effect is misleading because the effect will come or go depending on the kind of person treated. When ATIs are present, a generalization about aptitude is an uncertain basis for prediction because the regression slope will depend on the treatment chosen. Having said this much in 1957, I was shortsighted not to apply the same argument to interaction effects themselves. An ATI result can be taken as a general conclusion only if it is not in turn moderated by further variables. If Aptitude × Treatment × Sex interact, for example, then the Aptitude × Treatment effect does not tell the story. Once we attend to interactions, we enter a hall of mirrors that extends to infinity. However far we carry our analysis—to third order or fifth order or any other—untested interactions of a still higher order can be envisioned.

When I say something like that, some colleague is likely to reply: "In my experience, interaction effects are not large." To check that out, let us look at the magnitude of various effects in one ecology. The last four volumes of the *Journal of Personality and Social Psychology* contain 17 studies with the same design: A × B × C, persons

TABLE 2

Magnitudes of Main Effects and Interaction Effects in 17 Social-Psychological Studies

Standardized estimate of variance component	Cumulative percentage of components				
	Main effect			Interaction effect	
	Largest	Second largest	Smallest	First order	Second order
2.00–4.00	12 (7)				
1.00–1.99	29 (20)				
.50– .99	47 (33)	6 (7)		19 (26)	12 (22)
.30– .49	65 (53)	12 (13)		35 (31)	12 (22)
.10– .29	94 (93)	41 (40)	6 (0)	54 (49)	47 (55)
.00– .09	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)

Note. Components are scaled so that the two smaller main effects and the four interaction effects sum to 1.00 within any study. Each study had an A × B × C design. Cumulative proportions in parentheses take into account only those effects having a single degree of freedom.

nested within cells.[1] Each effect size can be described by the estimated variance component. For Table 2 I brought the studies to a common scale; interactions, we see, were in the same range of magnitude as the second-largest main effect. If these main effects were worth attention, so were the interactions.

Most branches of psychology have to be concerned with higher order interactions. According to McGuire's (1968, 1969) reviews, response to a persuasive communication, for example, is determined by interaction among characteristics of persuader, listener, message, and setting.

In the personality field, it is neglect of interactions that has kept alive the battle between the "situationists" and the trait theorists (Bowers, 1973). Propositions about traits are actuarial statements, valid over situations in the aggregate. Thus, a total score on a religious attitude questionnaire turns out to be a fine predictor of response in real life, when the criterion is an average over 100 kinds of relevant activity—saying grace, voicing conscientious objection, etc. (Fishbein & Azjen, 1974). The trait measure, however, has negligible power to forecast what the high scorer is likely to do in any one situation. The contention that behavior is determined by the situation alone is equally wrong. In studies Bowers reviewed, the Person × Situation interaction usually accounted for more variance than the situation effect.

Mischel (1973) argued that research in personality cannot but become the study of higher order interactions:

> For example, to predict a subject's voluntary delay of gratification, one may have to know how old he is, his sex, the experimenter's sex, the particular objects for which he is waiting, the consequences of not waiting, the models to whom he was just exposed, his immediately prior experience—the list gets almost endless. (p. 256)

Those seven variables can give rise to 120 interactions, a number beyond the practical reach of a direct experiment. In his work on these variables, Mischel has directly observed only interactions of the lower orders. It is by inference from inconsistencies across experiments that he makes a case for some relations at about the fourth order (Mischel & Moore, 1973). If reactions are so complexly conditioned, it is not even faintly surprising that we get contradictory conclusions from experiments taking only two or three factors into account.

The problem is as pressing for cognitive psychology as for personality. Newell (1972) lamented the current fragmentation in the study of information processing. He tallied 59 colonies of investigators, each collecting data on its own narrow task. Because the fine structure of the task and the person's characteristics influence outcomes, results obtained under such disparate conditions cannot be linked up. Newell doubted that the usual experimental strategy of narrowing condi-

[1] The interactions were not all of the ATI type. Some were interactions of situational characteristics, and rarely there was an analysis of two characteristics of the subject. I passed over studies having "replicates" as a factor, and two multivariate analyses of variance. Where a continuous characteristic of subjects had been blocked in two levels, I made a correction for continuity (Cronbach & Snow, chap. 4) to increase the component of variance for that factor and its interactions. In each analysis I assumed factors to be fixed. In standardizing, I set 100% equal to the sum of the two lesser main effects and the four interactions. The largest main effect is often little more than a demonstration of the obvious, and I therefore did not include it in the scaling unit. (The residual component, which includes person main effects and numerous interactions, was typically many times the size of the other effects combined.)

tions, refining results, and looking for generalizations of limited range can generate adequate cognitive theory.

A laboratory generalization, once achieved, may not be a good first approximation to real-world relationships. Psychology has no regularity more venerable than that of E. H. Weber. A person's difference limen is smaller for stimuli of smaller magnitude—who could doubt it? Adelbert Ford and I were once engaged in training sonarmen to detect pitch differences, not in pure tones but in sea echoes. When the echo coming back off a submarine becomes higher in pitch than the signal the sonar put into the water, the submarine has turned toward the sonar vessel. The sonarman of those days had to hear this Doppler effect promptly, so that his vessel could counteract the sub's maneuver. Somewhat casually, the Navy had chosen 800 cps as the frequency at which the sonarman heard the transmitted signal. It was fairly difficult to discriminate a 15-cycle difference between the echo and the transmission reverberating through the water; but 15 cycles of Doppler was equivalent to one important knot of relative movement. Ford and I, invoking Weber, proposed that the Navy relocate the base frequency at 500 cycles. If the difference between 800 cps and 815 cps is near threshold, the 500–515 difference ought to be discriminated easily. To test the idea, we heterodyned sea reverberations and echoes down to the 500-cycle base. But sonarmen did no better than on an 800-cycle series. Any improved ability to discriminate was masked by a systematic bias. The rumbling, 500-cycle background somehow made it difficult for the men to concentrate on the difference itself, and they consistently reported "down Doppler" more often than "up Doppler." As responses at 800 cycles were free from this bias, the proposal was dropped.[2] I do not say that Weber's law was falsified when we took it to sea, but in the presence of the Frequency × Tonal Quality interaction it no longer accounted for appreciable variance.

McKeachie (1974) made the same point in lamenting the undependability of the "laws of learning" when carried into the classroom. Many of the problems, he said, derive "from failure to take account of important variables in natural educational settings" (p. 9). As one example he mentioned the emerging evidence that social class moderates instructional effects (see also Brophy & Evertson, 1973). It begins to appear that the lower-class child responds better (on the average) to didactic teaching, with explicit requirements and close-coupled rewards. Problem-oriented, ego-motivated, supportive methods of teaching, which educational theorists have long been advocating, seem to benefit only a middle-class clientele.

Even the animal experimenter is not exempt from problems of interaction. (I am indebted to Neal Miller for the following example.) Investigators checking on how animals metabolize drugs found that results differed mysteriously from laboratory to laboratory. The most startling inconsistency of all occurred after a refurbishing of a National Institutes of Health (NIH) animal room brought in new cages and new supplies. Previously, a mouse would sleep for about 35 minutes after a standard injection of hexobarbital. In their new homes, the NIH mice came miraculously back to their feet just 16 minutes after receiving a shot of the drug. Detective work proved that red-cedar bedding made the difference, stepping up the activity of several enzymes that metabolize hexobarbital. Pine shavings had the same effect. When the softwood was replaced with birch or maple bedding like that originally used, drug response came back in line with previous experience (Vesell, 1967).

The aim of social and behavioral science, since Comte, has been to establish lawful relations comparable to those of the traditional natural sciences. The program has not been without success, else it would not have commanded loyalty for so long. We need to reflect on what it means to establish empirical generalizations in a world in which most effects are interactive (Campbell, 1973).

The experimenter studying the isolated organism has two things working for him in his quest for dependable effects. He arranges the conditions under which he observes to a far greater extent than the social scientist can. When he says that such-and-such a relation is true, "other things being equal," he is speaking from the experience of having made a lot of things equal. When pine versus hardwood interacts, he can get rid of the interaction; investigators of drug effects need only agree hereafter to bed animals down on maple shavings. The results will be specific to a world

[2] We were not prepared to complicate the wartime training program by extra drills to remove systematic error. This might overcome the bias and enable the Weber effect to operate.

of maple shavings, but they will be orderly. I shall argue later in this article for close observation of effects the hypothesis under test says nothing about. One of the strong points of the laboratory sciences is that investigators find out about these matters before they "run a study." They bring the subject into a somewhat standard condition, they fine-tune their apparatus, and they check out details of the stimulus—all this before any subjects are run for the record. In educational, social, and developmental experiments, I fear that the common practice is to allow one casual pilot run to make sure the procedures can be carried out at all, and then to move at once into a formal experiment in which only the variables that enter the hypothesis are observed.

The second asset of the animal experimenter is that the system he investigates can usually be isolated. Effects are rarely sensitive to what is happening outside the laboratory room. What happens to one animal is not usually allowed to influence the behavior of the others. But the human subject's reaction in the experiment is influenced by his past and recent experiences elsewhere, and by what he has heard about psychologists (Freedman, Cohen, & Hennessy, 1973; Gergen, 1973).

Some social scientists nowadays are eager to establish rigorous generalizations about social policy by conducting experiments in the field. We have already seen mammoth federal experiments on performance contracting in education, on alternative rules for making "negative income tax" payments, and on alternative practices in compensatory education. As these experiments have moved toward completion, their advocates have become increasingly pensive. Alice Rivlin, a leader among those advocates, has just reiterated her belief that formal social experiments are worth their cost. But she also (Rivlin, 1973) entertains the thought that their proper use is to compare alternative rules, rules so formal that the winning competitor can be embodied in an act of Congress and enforced uniformly over the nation. The welfare alternatives are of this sort. Rivlin doubts that gross experimental comparison can produce useful rules for schooling, where a treatment is multifaceted, cannot be standardized, and interacts with pupil background. Under these circumstances, the between-school variation in practices swamps out any generalized effect of the specified treatment variable.

The Times as a Source of Interactions

Not even the income experiments establish a secure generalization. They tell us what rule produced the best results during the three tryout years. This is an observation about recent history, not an enduring conclusion. A decade from now, with changes in the economy, in social morale, in the family structure, and in aspirations, community attitudes will be different. The rules tested in 1970 might compare very differently if tested again in 1985. We tend to speak of a scientific conclusion as if it were eternal, but in every field empirical relations change (Gergen, 1973; Schlenker, 1974). It was once a sound generalization that DDT kills mosquitos, until the evolutionary mechanism acted rapidly to make mosquitos resistant to DDT. Further along the time scale, consider star maps for navigation. Those maps have not changed perceptibly during recorded history. And yet the stars are shifting in their courses, and it is only a matter of time—a lot of time—before the map changes beyond recognition. All that can be considered eternal is the laws. The gravitational "constant" will change from today's value, but we can assume that s will continue to equal $\frac{1}{2}\,gt^2$.

In psychology, Ghiselli (1974) suggested that even such a reliable finding as the superiority of distributed practice over massed practice may not remain valid from one generation to another. Similarly, J. W. Atkinson (1974) pointed out that when a substantial relation is found between personality variables, it describes only "the modal personality of a particular society at a particular time in history" (p. 408). He went on to say:

> I believe that the early success of Lewin et al. (1944) in the study of level of aspiration can be attributed largely to the fact that their subject samples, drawn from in and around German, and later American universities in the decades prior to World War II, were homogeneously high in achievement and low in anxiety. (p. 409)

Of a piece with this observation is the recognition that the California F Scale is obsolescent (Ghiselli, 1974; Lake, Miles, & Earle, 1973). The 25-year-old research supporting its construct validity gives us little warrant for interpreting scores today because with new times the items carry new implications. Perhaps the best example of all is Bronfenbrenner's (1958) backward look at research comparing middle-class and lower-class parenting. Class differences observed in the 1950s were sometimes just the reverse of what had been observed in 1930.

Generalizations decay. At one time a conclusion

describes the existing situation well, at a later time it accounts for rather little variance, and ultimately it is valid only as history. The half-life of an empirical proposition may be great or small. The more open a system, the shorter the half-life of relations within it are likely to be.

This puts construct validation (Cronbach, 1971; Cronbach & Meehl, 1955) in a new light. Because Meehl and I were importing into psychology a rationale developed out of physical science, we spoke as if a fixed reality is to be accounted for. Events are accounted for—and predicted—by a network of propositions connecting abstract constructs. The network is patiently revised until it gives a good account of the original data, and of new data as they come in. Propositions describing atoms and electrons have a long half-life, and the physical theorist can regard the processes in his world as steady. Rarely is a social or behavioral phenomenon isolated enough to have this steady-process property. Hence the explanations we live by will perhaps always remain partial, and distant from real events (Scriven, 1956, 1959b), and rather short lived. The atheoretical regularities of the actuary are even more time bound. An actuarial table describing human affairs changes from science into history before it can be set in type.

Our troubles do not arise because human events are in principle unlawful; man and his creations *are* part of the natural world. The trouble, as I see it, is that we cannot store up generalizations and constructs for ultimate assembly into a network. It is as if we needed a gross of dry cells to power an engine and could only make one a month. The energy would leak out of the first cells before we had half the battery completed. So it is with the potency of our generalizations. If the effect of a treatment changes over a few decades, that inconsistency is an effect, a Treatment × Decade interaction that must itself be regulated by whatever laws there be. Such interactions frustrate any would-be theorist who mixes data from several decades indiscriminately into the phenomenal picture he tries to explain.

The obvious example of success in coming to explanatory grips with interactions involving time is evolutionary theory in biology (Scriven, 1959a). Darwin considered observations on species against the background of ecologies and viewed his data in Galapagos as only the latest snapshot of an ever-changing ecology. The positivistic strategy of fixing conditions in order to reach strong generalizations (Allport, 1964, p. 550) fits with the concept that processes are steady and can be fragmented into nearly independent systems. Psychologists toward the physiological end of our investigative range probably can live with that as their principal strategy. Those of us toward the social end of the range cannot.

Interpretation in Context, Contrasted with Generalization

Social science has been dedicated to formal testing of nomothetic propositions. Given the difficulties interactions create for social science, what might a better strategy be? This has recently been discussed by Gergen (1973),[3] Glass (1972), Newell (1972), McGuire (1973), Snow (1974), and Campbell (in press); all of them, in one way or another, propose to break away from the preoccupation with fixed-condition experiments that seek generalizations.

I endorse many points of these several authors that I lack time to echo here. My only major divergence is that most of the others expect more progress toward enduring theoretical structures than I do. Advocates of "theory" mean many things. I am as prepared as anyone to endorse the value of such model building as we see McGuire or Jack Atkinson doing. But I cannot go so far as Suppes (1974), who exhorted us that theorizing is our principal duty and that in the fullness of time our successors will erect "theoretical palaces." Suppes and I both subscribe to the position that our abstract concepts perform a great service in altering the prevailing view of man (Cronbach & Suppes, 1969, pp. 122–134; cf. Gergen, 1973, on "sensitization"). But a point of view is not a theory, capable of sharp predictions to new conditions.

The experimental strategy dominant in psychology since 1950 has only limited ability to detect interactions. Typically, the investigator delimits the range of situations considered in his research program by fixing many aspects of the conditions under which the subject is observed. The interactions of any fixed aspect are thereby concealed, being pulled into the main effect or into the interactions of other variables. The concealed interaction may even wipe out a real main effect of the

[3] Note Schlenker's (1974) dissent, especially on points on which I echo Gergen.

variable that chiefly concerns the investigator. It makes sense for drug investigators to standardize on softwood bedding, since bedding is a variable outside the system they wish to understand. (But even they had to first establish that softwood was preferable to hardwood for their purposes.) When the system of interest cannot be constrained to fit a limited model, the function of research in highly standardized conditions is primarily to identify pertinent variables and to suggest possible mechanisms to study in more natural situations. Dick Atkinson has demonstrated how laboratory work on learning can suggest profitable lines for investigation in a case study of instruction (R. Atkinson, 1974; R. Atkinson & Paulson, 1972).

The investigator who employs a factorial design can detect some interactions of those conditions he allows to vary, but sizable interactions are likely to be suppressed, just because any interaction that does not produce a significant F ratio is treated as nonexistent. Unfortunately, enormous volumes of data are required to pin down higher interactions as significant, unless one is guided by strong prior knowledge. When the facets of the design have more than two levels, the sample size required for establishing complex interactions, at least in instructional research, becomes prohibitive (Cronbach & Snow, in press, chap. 4).

The time has come to exorcise the null hypothesis. We cannot afford to pour costly data down the drain whenever effects present in the sample "fail to reach significance." Originally, the psychologist saw his role as the scientific observation of human behavior. When hypothesis testing became paramount, observation was neglected, and even actively discouraged by editorial policies of journals. Some authors now report nothing save F ratios. Hereafter, let us see estimates of variance components and raw-score regression coefficients instead. Confidence intervals will serve adequately to keep us cautious. Let the author file descriptive information, at least in an archive, instead of reporting only those selected differences and correlations that are nominally "greater than chance." Descriptions encourage us to think constructively about results from quasi-replications, whereas the dichotomy significant/nonsignificant implies only a hopeless inconsistency.

The canon of parsimony, misinterpreted, has led us into the habit of accepting Type II errors at every turn, for the sake of holding Type I errors in check. There *are* more things in heaven and earth than are dreamt of in our hypotheses, and our observations should be open to them (Cronbach, 1954). From Occam to Lloyd Morgan, the canon has referred to parsimony in theorizing, not in observing. The theorist performs a dramatist's function; if a plot with a few characters will tell the story, it is more satisfying than one with a crowded stage. But the observer should be a journalist, not a dramatist. To suppress a variation that might not recur is bad observing.

Correlational research is distinguished from manipulative research in that it accepts the natural range of variables, instead of shaping conditions to represent a hypothesis. By sampling from a population of persons, or from a domain of situations in the Brunswikian sense, one puts himself in a somewhat better position to generalize. Majasan's parallel inquiries in 12 classrooms formed a correlational study of representative instructors and teaching conditions. Representative design gives us more information about instruction, but it does not carry us far in the study of interactions. It is rarely practical to obtain information in a large number of situations. And the statistical estimates typically describe the gross aggregation of conditions instead of pinning down just what joint action of situational variables produces a particular effect.

In the investigation of complex practical and social phenomena, I am sure we will continue to employ manipulative experiments and to test hypotheses stated in advance about the fixed conditions. I am sure we can make better use of Brunswikian correlational analysis. But I believe that in past research the psychologist has been too willing to stop as soon as he has calculated the statistics stating the strength of the relationships he specified a priori. The experimenter or the correlational researcher can and should look within his data for local effects arising from uncontrolled conditions and intermediate responses (Edwards & Cronbach, 1952). He can do so, of course, only if he collected adequate protocols from the start.

Instead of making generalization the ruling consideration in our research, I suggest that we reverse our priorities. An observer collecting data in one particular situation is in a position to appraise a practice or proposition in that setting, observing effects in context. In trying to describe and account for what happened, he will give attention to whatever variables were controlled, but he will give equally careful attention to uncontrolled conditions, to personal characteristics, and to events that oc-

curred during treatment and measurement. As he goes from situation to situation, his first task is to describe and interpret the effect anew in each locale, perhaps taking into account factors unique to that locale of series of events (cf. Geertz, 1973, chap. 1, on "thick description"). As results accumulate, a person who seeks understanding will do his best to trace how the uncontrolled factors could have caused local departures from the modal effect. That is, generalization comes late, and the exception is taken as seriously as the rule.

Majasan reached a fine statistical generalization, but he had no way to go behind his pretest and posttest to learn what mediated the effect. To carry the work further, Katharine Baker at Stanford has gathered dissertation data on a wide range of variables. She replicated the Majasan procedure (in a smaller number of classes), adding classroom observations, and collected information on course content and assignments. Whatever regressions appear in her several classes, it is my hope that her data will enable her to give a plausible account of the mediating events that generated them.

When we give proper weight to local conditions, any generalization is a working hypothesis, not a conclusion. The personnel tester, for example, long ago discovered the hazard in generalizing about predictive validity, because test validity varies with the labor pool, the conditions of the job, and the criterion. To select salesmen using a test found valid in *other* firms is indefensible in the absence of solid knowledge about how aptitudes interact with the parameters along which sales jobs vary. Hence, personnel testers are taught to collect local data before putting a selection scheme into operation, and periodically thereafter. Likewise, positive results obtained with a new procedure for early education in one community warrant another community trying it. But instead of trusting that those results generalize, the next community needs its own local evaluation.

These examples are from applied work, but the same style can be used when one's motives are pure. Mischel's (1974; Mischel & Moore, 1973) work on delay of gratification shows how generalized thinking can be enriched by local observation. He considered data from Trinidad or Uganda or the United States against the community background. He found out what his subjects were saying to themselves during the delay interval and used that to explain their scores. He varied his procedures and used the inconsistencies from experiment to experiment to make the case for specific interactions.

The two scientific disciplines, experimental control and systematic correlation, answer formal questions stated in advance. Intensive local observation goes beyond discipline to an open-eyed, open-minded appreciation of the surprises nature deposits in the investigative net. This kind of interpretation is historical more than scientific. I suspect that if the psychologist were to read more widely in history, ethnology, and the centuries of humanistic writings on man and society, he would be better prepared for this part of his work.

Realizable Aspirations for Social Inquiry

Social scientists generally, and psychologists in particular, have modeled their work on physical science, aspiring to amass empirical generalizations, to restructure them into more general laws, and to weld scattered laws into coherent theory. That lofty aspiration is far from realization. A nomothetic theory would ideally tell us the necessary and sufficient conditions for a particular result. Supplied the situational parameters A, B, and C, a theory would forecast outcome Y with a modest margin of error. But parameters D, E, F, and so on, also influence results, and hence a prediction from A, B, and C alone cannot be strong when D, E, and F vary freely. Theorists are reminded from time to time that the person who states a principle must also state the boundary conditions that limit its application. The psychologist can describe the conditions under which his generalizations have held, or the domain of which they provide an actuarial summary. He cannot often state the boundaries defining how far they will hold (Donagan, 1962). How could Weber, or anyone down to 1940, have told us whether the pitch "law" would apply to sonar echoes? No one had heard such echoes until the sonar apparatus itself was invented, and there was no reason for theorists to consider in advance the variables characteristic of the sonar reverberation.

The forecast of Y from A, B, and C will be valid enough, if conditions D, E, F, etc., are held constant in establishing and in applying the law. It will be actuarially valid, valid on the average, if it was established in a representative sample from a universe of situations, as long as the universe remains constant. When the universe changes, we have to go beyond our actuarial rule. As Meehl

(1957) has said, when we step outside the range of our experience, we have to use our heads.

Though enduring systematic theories about man in society are not likely to be achieved, systematic inquiry can realistically hope to make two contributions. One reasonable aspiration is to assess local events accurately, to improve short-run control (Glass, 1972). The other reasonable aspiration is to develop explanatory concepts, concepts that will help people use their heads.

Short-run empiricism takes soundings as one proceeds into unfamiliar waters. Only by testing pitch discrimination on sea echoes could Ford and I learn that a 500-cycle base has no advantage. Any evaluator is engaged in monitoring an operation in context. Though from persistent work in many contexts he may reach an actuarial generalization of some power, this will rarely be a basis for direct control of any single operation. Given Domino's (1971) impressive generalization, a college counselor might classify instructors as pressing for conformity or individuality, and might then advise the high-Ac student as to which course section to enroll in. But the generalization gives no guarantee of his individual success because it ignores additional variables such as those of Majasan. Hence the Domino interaction ought not to control irreversible assignments; rather, it offers the student a hypothesis about choice of section, one that he can confirm or reverse after he is two weeks into the term. Short-run empiricism is "response sensitive" (R. Atkinson & Paulson, 1971); one monitors responses to the treatment and adjusts it, instead of prescribing a fixed treatment on the basis of a generalization from prior experience with other persons or in other locales.

In order to give a wide reach to our explanations, we make experience cumulative by abstracting from it. The explanatory constructs that we find fruitful combine into a view of man, his institutions, and his behavior. The informed public projects each new circumstance against that background, and so is able to react more intelligently. As time passes, the prevailing view will not necessarily progress from hazy vision to crude sketch to articulate blueprint. A general statement can be highly accurate only if it specifies interactive effects that it takes a large amount of data to pin down. Some effects in the network will change in form, in a span of one or two generations, even before enough qualifying clauses have been added to describe the effect accurately. Though our sketch of man may become more elaborate, it will remain a sketch.

Social scientists are rightly proud of the discipline we draw from the natural-science side of our ancestry. Scientific discipline is what we uniquely add to the time-honored ways of studying man. Too narrow an identification with science, however, has fixed our eyes upon an inappropriate goal. The goal of our work, I have argued here, is not to amass generalizations atop which a theoretical tower can someday be erected (cf. Scriven, 1959b, p. 471). The special task of the social scientist in each generation is to pin down the contemporary facts. Beyond that, he shares with the humanistic scholar and the artist in the effort to gain insight into contemporary relationships, and to realign the culture's view of man with present realities. To know man as he is is no mean aspiration.

REFERENCES

Allport, G. W. *Pattern and growth in personality.* New York: Holt, Rinehart & Winston, 1964.

Atkinson, J. W. Motivational determinants of intellective performance and cumulative achievement. In J. W. Atkinson & J. O. Raynor (Eds.), *Motivation and achievement.* Washington, D.C.: Winston, 1974.

Atkinson, R. C. Teaching children to read using a computer. *American Psychologist,* 1974, *29,* 169–178.

Atkinson, R. C., & Paulson, J. A. An approach to the psychology of instruction. *Psychological Bulletin,* 1972, *78,* 49–61.

Bowers, K. S. Situationism in psychology: An analysis and a critique. *Psychological Review,* 1973, *80,* 307–336.

Bronfenbrenner, U. Socialization and social class through time and space. In E. E. Maccoby, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology* (3rd ed). New York: Holt, Rinehart & Winston, 1958.

Brophy, J. E., & Evertson, C. M. *Low-inference observational coding measures and teacher effectiveness.* Austin: University of Texas, 1973. (ERIC Document Reproduction Service No. ED 077879)

Campbell, D. T. The social scientist as methodological servant of the experimenting society. *Policy Studies Journal,* 1973, *2,* 72–75.

Campbell, D. T. Qualitative knowing in action research. *Journal of Social Issues,* in press.

Cronbach, L. J. Report on a psychometric mission to Clinicia. *Psychometrika,* 1954, *19,* 263–270.

Cronbach, L. J. The two disciplines of scientific psychology. *American Psychologist,* 1957, *12,* 671–684.

Cronbach, L. J. Test validation. In R. L. Thorndike (Ed.), *Educational measurement.* Washington, D.C.: American Council on Education, 1971.

Cronbach, L. J., & Gleser, G. C. *Psychological tests and personnel decisions.* Urbana: University of Illinois Press, 1957.

Cronbach, L. J., & Meehl, P. E. Construct validation in psychological tests. *Psychological Bulletin,* 1955, *52,* 281–302.

Cronbach, L. J., & Snow, R. E. *Aptitudes and instructional methods.* New York: Irvington, in press.

Cronbach, L. J., & Suppes, P. (Eds.). *Research for tomorrow's schools.* New York: Macmillan, 1969.

Domino, G. Differential prediction of academic achievement in conforming and independent settings. *Journal of Educational Psychology,* 1968, *59,* 256–260.

Domino, G. Interactive effects of achievement orientation and teaching style on academic achievement. *Journal of Educational Psychology,* 1971, *62,* 427–431.

Donagan, A. Are the social sciences really historical? In B. Baumrin (Ed.), *Philosophy of science: The Delaware Seminar*(Vol. 1). New York: Interscience, 1963.

Dowaliby, F. J., & Schumer, H. Teacher-centered versus student-centered mode of college classroom instruction as related to manifest anxiety. *Journal of Educational Psychology,* 1973, *64,* 125–132.

Edwards, A. L., & Cronbach, L. J. Experimental design for research in psychotherapy. *Journal of Clinical Psychology,* 1952, *8,* 51–59.

Fiedler, F. E. The effects of leadership training and experience: A contingency model interpretation. *Administrative Science Quarterly,* 1973, *18,* 453–570.

Fishbein, M., & Ajzen, I. Attitudes toward objects as predictors of single and multiple behavioral criteria. *Psychological Review,* 1974, *81,* 59–74.

Freedman, P. E., Cohen, M., & Hennessy, J. Learning theory: Two trials and tribulations. *American Psychologist,* 1974, *29,* 204–206.

Geertz, C. *The interpretation of cultures.* New York: Basic Books, 1973.

Gergen, K. J. Social psychology as history. *Journal of Personality and Social Psychology,* 1973, *26,* 309–320.

Ghiselli, E. E. Some perspectives for industrial psychology. *American Psychologist,* 1974, *29,* 80–87.

Glass, G. V. The wisdom of scientific inquiry on education. *Journal of Research in Science Teaching,* 1972, *9,* 3–18.

Goldberg, L. R. Student personality characteristics and optimal college learning conditions: An extensive search for trait-by-treatment interaction effects. *Instructional Science,* 1972, *1,* 153–210.

Hilgard, E. R., & Lerner, D. The person: Subject and object of science and policy. In D. Lerner & H. D. Lasswell (Eds.), *The policy sciences: Recent developments in scope and method.* Stanford, Calif.: Stanford University Press, 1951.

Insel, P. M., & Moos, R. H. Psychological environments: Expanding the scope of human ecology. *American Psychologist,* 1974, *29,* 179–188.

Israel, J., & Tajfel, H. (Eds.). *The context of social psychology.* New York: Academic Press, 1972.

Koch, S. Some trends of Study I. Conceptual and systematic. In S. Koch (Ed.), *Psychology: A study of a science*(Vol. 3). New York: McGraw-Hill, 1959.

Lake, D. G., Miles, M. B., & Earle, R. B., Jr. (Eds.). *Measuring human behavior.* New York: Teachers College, Columbia University, 1973.

Lasagna, L. The impact of scientific models on clinical psychopharmacology: A pharmacologist's view. *Seminars in Psychiatry,* 1972, *4,* 271–282.

Majasan, J. K. College students' achievement as a function of the congruence between their beliefs and their instructor's beliefs. Unpublished doctoral dissertation, Stanford University, 1972.

Markle, N. H. Differential response to instruction designed to call upon spatial and verbal aptitudes. Unpublished doctoral dissertation, Stanford University, 1968.

McGuire, W. J. Personality and susceptibility to social influence. In E. F. Borgatta & W. W. Lambert (Eds.), *Handbook of personality theory and research.* Chicago: Rand McNally, 1968.

McGuire, W. J. The nature of attitudes and attitude change. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (Vol. 3, 2nd ed.). Reading, Mass.: Addison-Wesley, 1969.

McGuire, W. J. The Yin and Yang of progress in social psychology: Seven koan. *Journal of Personality and Social Psychology,* 1973, *28,* 446–456.

McKeachie, W. J. The decline and fall of the laws of learning. *Educational Researcher,* 1974, *3*(3), 7–11.

McKeachie, W. J., Isaacson, R. L., & Milholland, J. E. *Research on the characteristics of effective college teaching.* Ann Arbor: University of Michigan, 1964. (ERIC Document Reproduction Service No. ED 002948)

McKeachie, W. J., Milholland, J. E., Mann, R., & Isaacson, R. *Research on the characteristics of effective teaching.* Ann Arbor: University of Michigan, 1968. (ERIC Document Reproduction Service No. ED 024347)

Meehl, P. E. When shall we use our heads instead of the formula? *Journal of Counseling Psychology,* 1957, *4,* 268–273.

Merton, R. K. *Social theory and social structure.* New York: Free Press of Glencoe, 1949.

Mischel, W. Toward a cognitive social learning reconceptualization of personality. *Psychological Review,* 1973, *80,* 252–283.

Mischel, W. Processes in delay of gratification. In L. Berkowitz (Ed.), *Advances in social psychology* (Vol. 7). New York: Academic Press, 1974.

Mischel, W., & Moore, B. Effects of attention to symbolically presented rewards on self-control. *Journal of Personality and Social Psychology,* 1973, *28,* 172–179.

Newell, A. You can't play 20 questions with nature and win. In W. G. Chase (Ed.), *Visual information processing.* New York: Academic Press, 1972.

Rivlin, A. M. Social experiments: The promise and the problem. *Brookings Bulletin,* 1973, *10*(3), 6–9.

Schildkraut, J. J. *Neuropsychopharmacology and the affective disorders.* Boston: Little, Brown, 1970.

Schlenker, B. R. Social psychology and science. *Journal of Personality and Social Psychology,* 1974, *29,* 1–15.

Scriven, M. A possible distinction between traditional scientific disciplines and the study of human behavior. In H. Feigl & M. Scriven (Eds.), *Minnesota studies in the philosophy of science* (Vol. 1). Minneapolis: University of Minnesota Press, 1956.

Scriven, M. Explanation and prediction in evolutionary theory. *Science,* 1959, *130,* 477–482. (a)

Scriven, M. Truisms as the grounds for historical explanation. In P. L. Gardiner (Ed.), *Theories of history.* Glencoe, Ill.: Free Press, 1959. (b)

Snow, R. E. Representative and quasi-representative designs for research in teaching. *Review of Educational Research,* 1974, *44,* 265–291.

Suppes, P. The place of theory in educational research. *Educational Researcher,* 1974, *3*(6), 3–10.

Thelen, H. A. *Classroom grouping for teachability.* New York: Wiley, 1967.

Vesell, E. S. Induction of drug-metabolizing enzymes in liver microsomes of mice and rats by softwood bedding. *Science,* 1967, *157,* 1057–1058.