

A PHILOSOPHY OF SCIENCE PERSPECTIVE ON THE VALIDITY OF RESEARCH CONCLUSIONS: RESPONSE TO ANDERSON AND HAYDEN

PAULINE HOULDEN*

Anderson and Hayden argue that meaningful policy implications cannot be drawn from research which lacks “structural, sociological, and conceptual verisimilitude.” Pointing to my article, “Impact of Procedural Modifications on Preferences for Plea Bargaining” (pp. 267-291) as an example of research which lacks such verisimilitude, they conclude that I am unjustified in drawing conclusions with policy implications. I disagree with them for four reasons. Anderson and Hayden have: 1) based their arguments on a misunderstanding of Lind and Walker’s discussion of the circumstances in which research need replicate reality; 2) misconceived the nature of theory; 3) created an unnecessary and unsupportable distinction between “structural,” “sociological,” and “conceptual” verisimilitude; and 4) misunderstood and misrepresented my article.

I. WHEN MUST THE RESEARCH SETTING REPLICATE REALITY?

Bermant *et al.* (1974) suggest that research lacking in “structural verisimilitude” might not be useful for the practical understanding of actual behavior. They claim that structural verisimilitude is determined by the degree of correspondence between the setting and circumstances of a research study and the problem it examines, as well as the extent of similarity between participants in the research and persons likely to be involved in that role in real life. Their study demonstrated that

* I wish to thank Allan Lind, Susan Kurtz, Steven Balkin, and especially, Stephen A. LaTour, for their careful reading of and comments on this paper.

research results were affected by the degree of structural verisimilitude.

Lind and Walker (1979) believe that this conclusion neglects an important dimension of research—whether a study is intended to test a theory or to investigate a specific situation. They argue that the degree of similarity between a research setting and a real-world setting is important only in the latter case. If the purpose of a research study is to test a theory, then the proper concern is not structural verisimilitude, but whether the theory has been correctly operationalized and tested. Structural verisimilitude is important only if it is an element of the tested theory.

Anderson and Hayden (1981) have, in my judgment, seriously misrepresented Lind and Walker. For example, Anderson and Hayden declare that “simulation of the type discussed by Lind and Walker . . . does assume at least minimal structural verisimilitude” (p. 294). But Lind and Walker do not take this position. They state that if a theory is being tested, research need not replicate an actual situation at all. For example, in their discussion of Fuller’s hypothesis that use of the adversary procedure reduces the likelihood of a quick decision based on “familiar categories” (i.e. decision-making bias), they state:

. . . a proper test of [Fuller’s hypothesis] would require only that some group of subjects be exposed to a preliminary experience that would produce the necessary expectations on which hasty judgments might be based. The hypothesized effect of adversary presentation on this expectancy “bias” could then be tested by presenting information about a case in an adversary fashion to some subjects and presenting the same information in a nonadversary fashion to other subjects It is not necessary that the experiment simulate closely a court trial or that the subjects closely resemble judges or jurors. It is necessary only that the conditions just outlined be met in such a way as to provide a clear test of the basic statement (Lind and Walker, 1979: 10).

Clearly, Lind and Walker do not believe that theory-testing research must mirror to any degree the real-world settings to which one may apply that theory. It is only when research seeks to explain behavior within one particular situation (eg., job dissatisfaction among prison guards at Joliet, Illinois) that the research setting must mirror the real world.

Kruglanski (1975a; 1975b; 1975c; Kruglanski and Kroy, 1976) has written frequently about the confusion of these two types of research and the lack of understanding of the philosophy of science implied by this confusion. According to Kruglanski, theory-testing or “universalistic research” is primarily concerned with creating a test to honestly evaluate the hypothesized relationship between a variable or variables of

some theoretical interest. It is not concerned with any determinants of the investigated phenomenon other than those of theoretical interest. It is, therefore, not concerned with structural verisimilitude. An example of universalistic research would be studies of whether adversary methods of presentation counteract decision-making bias, discussed by Lind and Walker, Hayden and Anderson (1979), and Anderson and Hayden (1981). The intended generalization of such research is universal—in this instance, all decision making. In Kruglanski's words (1975a: 165), for universalistic research,

The sole requirement is to ponder carefully whether the meaning of the total experimental situation . . . captures the essence of the theoretical variable(s). No further attention is paid to aspects of the situation which appear theoretically irrelevant.

Situation-specific, or "particularistic" research, however, aims at understanding a single situation. It is primarily concerned with establishing a test setting that will honestly portray the situation which it seeks to understand and to which its results will be applied. Structural verisimilitude is therefore essential to the achievement of accurate results. An example of particularistic research would be a jury selection study in which lawyers and psychologists sought to predict the members of a venire who would be most favorable to the defense or prosecution in a particular case. The intended generalization of the findings of particularistic research is only the specific set of conditions investigated.

This distinction explains the results obtained by Bermant and his colleagues (1974). They found structural verisimilitude to have an effect because they had conducted a particularistic study¹ (predicting the behavior of a jury for one specific case), precisely the type of research for which structural verisimilitude is thought to be important. Had they conducted a theory-testing study (for example, exploring whether defendant attractiveness affected verdicts), structural verisimilitude would either have been irrelevant or would have been incorporated within the theory.

The purpose of my experiment was to test theory: to ascertain whether participation would increase disputants' preferences for plea bargaining procedures, as it had for traditional adjudicatory procedures. It also sought to test the theory that a mediator is desired in conflict resolution and to

¹ One should note that there was no positive linear relationship between degree of structural verisimilitude and correspondence of research results with the verdict of the actual jury. There was no indication that increasing structural verisimilitude increases the validity of research results.

explore several theories about interactions between these two variables and respondents' perceptions of their own guilt or innocence, perceptions of the strength of their case, and status of respondents as undergraduate students or prison inmates. It is thus a universalistic study to be judged by whether it operationalized its independent and dependent variables in a manner that allowed a test of the theories of interest. Yet Anderson and Hayden, lamenting the lack of structural verisimilitude in my study, have inappropriately sought to apply standards appropriate to a particularistic or situation-specific study.

Whether research is universalistic or particularistic has no implications for whether or not it can be used to generate policy implications. Certainly the route from research to policy implications differs from universalistic and particularistic research (Calder *et al.*, 1981), but each type of research has applications. The policy implications of particularistic research are direct and easily determined: as long as the research procedures simulate the real-world setting, the results are assumed to be similar to the effects that would occur in the real world. The policy application of theory-testing research is less direct. Research is used to test a theory, and if the theory survives strong efforts at disproof, it is used to develop an "intervention." The intervention is a program or activity that those who develop it believe will have an effect in the real world. An intervention may fail (perhaps due to factors in the real world which were not addressed by the theory or mistakes in the application of the theory by those developing the intervention), but its failure demonstrates only that the intervention has flaws. It does not question the validity of the underlying theory or the suitability of the theory for real-world applications.

The theories tested by my study suggested an intervention—that defendants be allowed to participate in plea bargaining and that state-paid mediators not be included. It is this intervention that was proposed in the "policy implications" section of my article. This intervention may or may not work, but philosophy of science certainly supports my derivation of it. Anderson and Hayden correctly noted that, "The ultimate utility of theories can only be judged in studies of real-life situations" (p. 302). It was for that reason that I suggested a possible intervention. Yet, my intervention can only be based on the theories I have tested. Because my theories did not specify aspects of real-world situations which could limit them,

my conclusions cannot speak to those issues. I could have speculated about the possible effects of various untested factors, but my proposed intervention could properly be based only on the results of the theories tested in my study. Anderson and Hayden may disagree with these theories or question whether they were adequately tested, but they may not suggest that the conclusions I derived or the intervention I suggested should have been based on speculations about variables not considered by my research. If Anderson and Hayden wish to propose and test a different theory which they believe would result in development of a different intervention, they are free to do so. If Anderson and Hayden wish to suggest that my use of a laboratory situation or role-playing did not permit proper operationalization and testing of the theoretical variables, that would be a valid criticism. But if they wish to do this, they must explain why they believe my operationalization of participation and presence/absence of a mediator was insufficient. This they have failed to do.

Anderson and Hayden further misperceive that Lind and Walker are inconsistent if they believe that research which is not a perfectly accurate reflection of the world can be used to recommend policy. Neither Lind and Walker nor Kruglanski nor I argue that the results of the one situation which has been experimentally examined can be used to suggest policy for other situations. We believe the theory which is tested in an experimental setting can be used to suggest policy. There is a difference, unacknowledged by Anderson and Hayden, between applying the results of a study and the theory tested in an experiment. There is nothing "inconsistent" in suggesting that theory need not be tested in a real-world setting, but yet, if not falsified, used to derive policy implications for the real world.

Finally, by requiring research to have structural verisimilitude, Anderson and Hayden are ignoring the success achieved by social science in applying theory-testing laboratory studies to real-life policy decisions. For instance, multi-attribute attitude theory has been tested in laboratory settings, yet successfully applied to predicting attitudes of consumers towards products (Lutz and Bettman, 1977). Theories of group polarization and conformity, tested in the laboratory, have been found useful by businesses in guarding against poor management decisions. Theories of helping behavior tested in laboratories have been found to correlate highly with altruism in real-world settings (eg., Piliavin and Piliavin, 1972). Why is it not reasonable (and parsimonious) to assume that theories of

procedural justice tested in less than perfectly realistic settings can also be generalized to the real world?

II. THE NATURE OF THEORY

Anderson and Hayden disagree with Lind and Walker's claim that restrictions on the domain of theory will become evident in the course of testing that theory. They believe that theories cannot be modified by the findings of future research, and they urge researchers to state theories in as restricted a form as possible. For instance, "the adversary system leads to more unbiased decisions in tort cases," when one has experimented only with tort cases, is proposed as a better theory than "the adversary system leads to more unbiased decisions." This suggestion, along with statements such as "the tests upheld the theory" (p. 295) reveal a commitment to inductive rather than deductive theory testing. Yet, Popper (1959) has argued that inductive reasoning is untestable and that only a deductive approach to theory testing is scientifically acceptable. Theories must be stated in general form and accepted as stated until tests, through attempts at falsification, have demonstrated the limits of the theory. Thus, we attempt to disprove theories, not uphold them, and we accept them as stated until the theoretical statement is disproved. Only when additional research reveals a restricting variable, such as type of judge or attorney or defendant, is the theory revised and restricted.

Philosophy of science acknowledges that any theory may be incorrect and that the possibility always exists that variables not yet considered will interact with those currently included (Popper, 1959). Contrary to the assertions of Anderson and Hayden, it is an accepted tenet of our present approach to science that continued research can reveal the shortcomings of theories and that theories can be changed as knowledge expands. In fact, research such as mine is a good example of a study which could reveal the limitations of a theory. It sought to determine whether type of procedure (negotiation as opposed to traditional adjudicatory procedures) limited the theory that participation in a conflict resolution procedure increases preference for that procedure. Why Anderson and Hayden do not expect that the results of my experiment could have modified the theory of participation is, however, unexplained.

There are other problems with suggesting that theories should be expressed in as restrictive a form as possible. First,

there are so many restrictions on any piece of theory-testing research that one would be hard pressed to know which ones to consider. Is the adversary method of presentation likely to lead to more unbiased decisions for tort cases only for male judges to whom facts are presented by males, in the fall of 1974, in the southeastern United States? Which of the endless limitations on this and all research should be "clearly set out"?

Second, as Schlenker (1974: 2) has commented, "If a theory incorporates specifics, it would not possess the generality to satisfactorily explain the required diversity of phenomena." Thus we might note that no such statement as, "the adversary system provides for a more unbiased evaluation of tort cases" would be considered a theory. Even if a series of studies demonstrated that the adversary system resulted in less biased decisions only for tort cases, the theoretical statement would be expressed in as general terms as possible. It would refer to differences between tort cases and all other types of cases that could account for this difference (eg., degree of bias is a function of extent of adversariness and degree of plaintiff injury).

III. DEFINITIONS OF VERISIMILITUDE

There are problems with Anderson and Hayden's definitions. They have employed a definition of "structural verisimilitude" similar to that suggested by Bermant *et al.* (1974) and Vidmar (1979). Yet because Anderson and Hayden define "structural verisimilitude" as only the relative similarity between a research setting and an "institution" under study, it is not as broad, and hence, not as useful a definition as that of their predecessors. Anderson and Hayden have defined "sociological verisimilitude" as recognition that the institution under investigation is part of a "complex social system." "Conceptual verisimilitude," previously suggested by Vidmar (1979), is concerned with whether a research study explores a problem as it would be identified by a person knowledgeable about it.

The need for this tri-partite definition is unclear and seems unnecessarily complicated (cf. Kruglanski and Kroy, 1976). It is difficult to understand why sociological verisimilitude, rather than structural verisimilitude, addresses the relevance of whether individuals are regular participants in plea bargaining. Why should the fact that there are regular participants in plea bargaining, if known only to lawyers, be considered an issue of conceptual and not structural verisimilitude? Does not failure

to include this dimension weaken particularistic research findings in the same way as a dimension known to the general public? In fact, the three verisimilitudes are not mutually exclusive. They are one concept and better identified by the older term “validity” (Campbell, 1957).

Introduction of the concept of “institution” further clouds the utility of the Anderson-Hayden verisimilitudes. For example, what is the relevant “institution” for my research? Is it the plea bargaining process, or the courts in which the plea bargaining occurred, or the types of cases which are customarily plea bargained, or the types of defendants who usually negotiate? Clearly, most of these are not “institutions.” Yet if one is conducting particularistic research, one must accurately reflect all these dimensions. Indeed, an endless number of variables could be included within the concept of verisimilitude. As used by Anderson and Hayden, verisimilitude has become so all-encompassing that it can never be achieved.

Even if these problems did not exist, to what is one to match the experimental setting? Is it to be plea bargaining as it is conducted in Chicago, in one court, at one time, by one prosecutor and one public defender? But this study will not match plea bargaining as it is conducted in any other location at any other time. The demand for verisimilitude implies that plea bargaining, and indeed all social science phenomena, cannot be studied except within specific situations. It denies the possibility of theory and suggests that social science is no more than history. The errors of this belief are discussed by Schlenker (1974).

IV. THE ORIGINAL ARTICLE

Anderson and Hayden state that my research is based on an incorrect assumption—the assumption that formal plea bargaining sessions exist: “If they do not exist, then Houlden’s major policy recommendation is of doubtful utility” (p. 296). “Implicit” plea bargaining may well exist, but the existence of “explicit” formal plea bargaining sessions cannot be denied. Anderson and Hayden themselves discuss the research of Heinz and Kerstetter (1979) in which defendants attended plea bargaining sessions, and they mention the writings of Mather (1979) and Feeley (1979) who describe formal plea bargaining sessions between prosecutors and public defenders.

Anderson and Hayden also claim that my research lacks structural verisimilitude because it presents plea bargaining as

a session in which proposals and counter proposals are made. Such an exchange of proposals does not, they claim, occur. Rather defendants/cases are matched to “standard” cases, and the sentence that has been agreed upon as appropriate for a “standard” case is assented to by both the prosecution and defense. The literature on plea bargaining suggests, however, that plea bargaining assumes different forms in different jurisdictions and within jurisdictions, for cases of varying severity and correspondence to common offense patterns. I therefore wrote a sufficiently general description of the plea bargaining process to apply to sessions that matched defendants/cases to classifications, discussed the strengths and weaknesses of a case in detail, or proposed several dispositions. Respondents were told that plea bargaining would involve the following:

[T]he public defender and the prosecutor will discuss reducing the charge which has been brought against you from murder in the first degree to murder in the second, to murder in the third degree or manslaughter. You will be present at this discussion and may participate whenever you believe it is in your best interest to do so. A mediator will be present at this discussion . . . (Houlden, 1981: 277).

There is no mention of “proposals and counter proposals in this description of plea negotiation.” Rather, the description provides for a broad range of negotiation techniques.

Anderson and Hayden also believe that I have confused plea bargaining and adjudicatory procedures, since I consider plea bargaining “to be a type of fact-finding procedure” (p. 300). In fact, as Malcolm Feeley has observed, fact finding is a crucial aspect of plea bargaining:

Most discussions suggest that the primary object of plea bargaining is administrative convenience; yet some of the more thoughtful examinations of “negotiations” suggest that bargaining sessions are often probing examinations into the facts of the case, facts which in turn can affect the appropriateness of the charges (Feeley, 1979: 29).

Much of what passes for plea bargaining is really negotiation over the meaning of facts . . . Facts are malleable. They must be mobilized and often they are manipulated (Feeley, 1979: 29).

Indeed, Anderson and Hayden themselves acknowledge the importance of fact finding in plea bargaining (p. 301).

Part of Anderson and Hayden’s misunderstanding of the importance of “fact finding” (or as I label it, opportunity for the defendant [or prosecutor] to present evidence) in my research may stem from their misinterpretation of the normalized correlation scores. Anderson and Hayden state, “drawing on the work of Thibaut *et al.*, [Houlden] sees the most important variable in defendant satisfaction as the amount of defendant control over the presentation of evidence” (p. 300). This is not what I said. On the basis of previous research, I devised a list

of attributes of conflict-resolution procedures that disputants might prefer. Respondents' evaluations of those attributes were correlated with their preference scores for the six plea bargaining procedures. The attribute which correlated most highly with the pattern of preference scores was *not*, as Anderson and Hayden wrote, "defendant control" but the two attributes of "defendant opportunity for evidence presentation" and "fairness of the procedure." Although the former attribute may imply defendant control to Anderson and Hayden, the dimension of "defendant control over evidence presentation," which was assessed, did not correlate highly with respondents' preferences.

Anderson and Hayden seem to think that I identified the dimensions of "defendant opportunity for evidence presentation" and "fairness of the procedure" as the most important determinants of preference. In fact, they were identified by the respondents. The importance of these variables emerged from the results presented in Table 2. The variable which Anderson and Hayden see as the goal of negotiation—getting the best deal possible—did not, in my study, correlate highly with preferences for plea bargaining procedures (see Table 2).²

Anderson and Hayden also express concern over the competence of defendants to participate in plea bargaining. This is an interesting question, but irrelevant to the research which was conducted. The primary purpose of my research was to determine whether defendants would prefer plea bargaining if they were allowed to participate and/or a mediator were to be included in the discussion. The issue of competence can only be addressed by additional research. Similarly Anderson and Hayden's speculation that defendants will not understand the process of plea bargaining is irrelevant.

Anderson and Hayden also misconstrue the relevance of the Heinz and Kerstetter study (1979). The issue is whether defendants are relatively less satisfied with plea bargaining than are other participants (police, judges, victims). Perhaps they are, but my study was concerned with whether modifications of the plea bargaining process would render that

² Some readers may note a shift in this paragraph from the deductive terminology of the rest of the paper and to an inductive approach. This is because my study aimed both to test theory and to build it. It tested, for example, the theory of participation and sought to extend this theory by ascertaining which variables mediated the effect of participation on preference. Theory building, unlike theory testing is inductive; thus, the description of this part of the study uses inductive terminology.

procedure more acceptable to defendants. Simply because defendants are less favorably disposed to plea bargaining than are other participants does not mean that modifications might not make plea bargaining more acceptable to them. It is particularly important to examine modifications which might increase the acceptability of plea bargaining to those who must bear its consequences most directly.

It is difficult to understand why Anderson and Hayden would suggest that "the lack of defendant participation in plea bargaining was an important determinant of preference in Heinz and Kerstetter's investigation." Heinz and Kerstetter found no differences in defendants' satisfaction with plea bargaining that depended on their level of participation. My explanation that it was the presence of the state-paid mediator which equalized preferences between defendants who did and did not participate is consistent with the results of both their research and mine.

V. CONCLUSION

Anderson and Hayden may have wished to use their comment to debate the general issue of when researchers should make policy recommendations. There are two points of view: the one that I have taken, that any piece of well-conducted research, be it universalistic or particularistic, may be used; the other, that a body of research must be established before any recommendations can be made. The latter strategy seems beset with difficulties. Who can decide when there is sufficient evidence to warrant a policy recommendation? How is this decision to be made? Who will fund and who will conduct all the replications (traditionally unappealing to funding agencies and most academics) necessary to establish this body of evidence? Is it better simply to allow policy makers to interpret research findings without guidance from those who conducted the research? Certainly there are arguments to be made for each position, but Anderson and Hayden cannot declare unequivocally that policy recommendations should not be based on a single study.

Let me note, however, that the policy implications of my paper were not based solely on the results I reported, but also on extensive research studies in psychology and criminal justice. The relationship between my results and this larger body of research is discussed in the article.

In sum, Anderson and Hayden misunderstand the underlying philosophy of research. As a result they do not

grasp the significance of the difference between situation-specific and theory-testing research, and direct their concern instead to the issue of laboratory versus real-world studies. Yet it does not matter where research is conducted; it is the underlying purpose which is crucial. The purpose determines how the research is designed and how it is to be evaluated.

REFERENCES

- BERMANT, Gordon, Mary MCGUIRE, William MCKINLEY, and Chris SALO (1974) "The Logic of Simulation in Jury Research," 1 *Criminal Justice and Behavior* 224.
- CALDER, Bobby J., Lynn W. PHILIPS, and Alice M. TYBOUT (1981) "Applying Consumer Research." Unpublished manuscript, Northwestern University.
- CAMPBELL, Donald T. (1957) "Factors Relevant to the Validity of Experiments in Social Settings," 54 *Psychological Bulletin* 297.
- CAMPBELL, Donald and Julian C. STANLEY (1966) *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally.
- FEELEY, Malcolm (1979) *The Process is the Punishment: Handling Cases in a Lower Criminal Court*. New York: Russell Sage.
- GERGEN, Kenneth J. (1973) "Social Psychology as History," 26 *Journal of Personality and Social Psychology* 309.
- HAYDEN, Robert M. and Jill K. ANDERSON (1979) "On the Evaluation of Procedural Systems in Laboratory Experiments: A Critique of Thibaut and Walker," 3 *Law and Human Behavior* 21.
- HEINZ, Anne and Wayne KERSTETTER (1979) "Pretrial Settlement Conference: Evaluation of a Reform in Plea Bargaining," 13 *Law & Society Review* 349.
- KRUGLANSKI, Arie W. (1975a) "The Two Meanings of External Invalidity," 28 *Human Relations* 653.
- (1975b) "Context, Meaning and the Validity of Results in Psychological Research," 66 *British Journal of Psychology* 373.
- (1975c) "The Human Subject in the Psychology Experiment: Fact and Artifact," in Leonard Berkowitz (ed.), *Advances in Experimental Social Psychology*, Vol. 8. New York: Academic Press.
- KRUGLANSKI, Arie W. and Moshe KROY (1976) "Outcome Validity in Experimental Research: A Re-conceptualization," 7 *Representative Research in Social Psychology* 166.
- LIND, E. Allan and Laurens WALKER (1979) "Theory Testing, Theory Development, and Laboratory Research on Legal Issues," 3 *Law and Human Behavior* 5.
- LUTZ, Richard J. and James R. BETTMAN (1977) "Multiattribute Models in Marketing: A Bicentennial Review," in Arch G. Woodside, Jagdish N. Sheth, and Peter D. Bennett (eds.), *Consumer and Industrial Buyer Behavior*. New York: North-Holland.
- MATHER, Lynn (1979) *Plea Bargaining or Trial? The Process of Criminal Case Disposition*. Lexington, MA: Lexington Books.
- PILLAVIN, Jane A. and Irving M. PILLAVIN (1972) "Effect of Blood on Reactions to a Victim," 23 *Journal of Personality and Social Psychology* 353.
- POPPER, Karl R. (1959) *The Logic of Scientific Discovery*. New York: Basic Books.
- SCHLENKER, Barry R. (1974) "Social Psychology and Science," 29 *Journal of Personality and Social Psychology* 1.
- VIDMAR, Neil (1979) "The Other Issues in Jury Simulation Research: A Commentary with Particular Reference to Defendant Character Studies," 3 *Law and Human Behavior* 95.