

UC Irvine

UC Irvine Previously Published Works

Title

Internal Replication and the Systems Concept in Non-Experimental Research

Permalink

<https://escholarship.org/uc/item/2qg587pz>

Authors

White, DR
Pesner, R

Publication Date

1983

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at

<https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Internal Replication, the Systems Concept, And Sources of Validity in Nonexperimental Research

Douglas R. White
and Robert Pesner

A probabilistic and nonparametric measure of replication (White and Pesner 1979) or probability that two bivariate distributions come from the same population, is presented for use in hypothesis testing. It is argued that relationships can be regarded as invariant within a specified system to the extent that they replicate not only regionally but with respect to third factors specified as inflection points or limits on invariant relationships: relationships are intrasystemically invariant and inter-systemically variant. System boundary conditions should show a high density of "interaction effects" for relationships which vary from system to system, or which fail to replicate cross-systemically. A heuristic system is proposed to decompose the variables and cases of an empirical population into component systems with invariant properties. The procedures are based on the measure of replication and group significance tests.

The Problem

Cross-cultural studies, survey analyses and natural observation studies lack the sort of justification for making relational inferences for correlational data which is valid for experimental studies.¹ In the experiment, with random assignment of treatments to cases, the assumption of *other things being equal* besides the experimental treatment is taken as appropriate. In this view the possibility that an observed correlation between experimental treatments and experimental outcomes is a spurious correlation—that is, due to some other between-group difference (third factor)—can plausibly be ruled out. Through experimental replications, the invariance of certain

relationships can be demonstrated for given experimental designs. Such is not the case in non-experimental studies. To the extent that the world is patterned and organized the third factor hypothesis is always more plausible as an explanation of the correlation than the assumption of a direct connection. Given that there are a great many correlates of a given outcome, why should the researcher be allowed any presumption that in demonstrating one or several correlates the correct identifications of invariant relations have been made?

Many comparativists have adopted a more stringent criterion for validating inferences, namely, that for an inference to be made from correlational evidence, the existence of the correlation must be replicated under many different possible test conditions. In this article we argue in favor of taking the problem of replication to its natural limit. For a correlation to have meaning for any sort of inference as to invariance, it must be shown by the researcher that the correlation replicates under all possible test conditions which are theoretically relevant to the topic of study.

By theoretical relevance we mean that one is able to indentify systems of related variables on theoretical grounds, and that variables are theoretically relevant to one another as test conditions if they are part of the same system. Replication of invariant relationships across these conditions is analogous to replication within a given experimental design.² The subsets of cases defined by the values of particular test conditions, over all such conditions for theoretically relevant variables, constitute the *naturally occurring subsamples* for a system of variables. A correlation can be said to replicate for a system of variables if and only if the distribution of replications in naturally occurring subsamples is close to the distribution which is expected to occur randomly in these subsamples.

How does one measure the degree of replication between two correlations? How does one measure the similarity between an expected and an observed distribution of replications? These problems are addressed next.

A Measure of Significance for Replication

Replication is typically tested by a probabilistic significance test. What is the probability that an observation, or an observed distribution, came from a population with given characteristics? For nominal and ordinal levels of measurement we have developed a generalization of Fisher's exact test as a significance measure of replication (White Pesner and Reitz 1983). The statistic will be briefly explicated here.

For simplicity, assume that the observed bivariate relationship in a population is expressed as frequencies in a fourfold table, where the two correlated variables are both dichotomies. Table I shows an example where the values of the four cells are $\frac{1}{11}$ $\frac{8}{5}$ and the correlation between the variables is positive. The population size is 27. We are interested in the replication of this relationship in a sample of the population having some particular characteristic. Table II exemplifies an observed replication sample where the four cells are $\frac{1}{11}$ $\frac{8}{5}$ and the correlation is again positive. The question is what is the probability that the observed sample would be drawn at random for the population? For reasons discussed below, we are primarily interested not in computing the probability of the observed sample as one among many samples of size 15, but as one among many samples with the same row and column marginals, or *comparable* samples—those with the given univariate distributions. The row and column constraint makes the question equivalent to the question of whether there is any *interaction* among the three variables—the two original variables plus the control condition—which would not be predicted from bivariate relationships among the variables. Thus, the replication question is “what are the chances of drawing $\frac{1}{11}$ $\frac{8}{5}$ out of $\frac{1}{11}$ $\frac{8}{5}$ in 15 draws such that there are 6 draws from row 1 and 8 from column 1?”

The replication probability can be derived by consideration of the combinatorial possibilities in drawing a sample from a population. Consider the first, upper left, cell of the sample in Table II (1 case) and the population in Table I (3 cases) from which the sample is drawn. How many ways are there

TABLE I
Bivariate Relationship in a Population
(A Hypothetical Illustration)

Trait B	Trait A	
	Absent	Present
Present	3	8
Absent	11	5

TABLE II
Bivariate Relationship in a Sample of the Population
In Table I (Replication) Having the Characteristic C

Number of Cases where C is present in the population: 15

Trait B	Trait A	
	Absent	Present
Present	1	5
Absent	7	2

of drawing one observation out of 3? There are 3-choose-1 or (3) or 2!1! or in general $(n-m)!m!$ ways of drawing m items out of n , in this case 3 ways of drawing 1 out of 3. The number of ways of drawing the values 1, 5, 7 and 2 out of 3, 8, 11 and 5 is thus 3-choose-1 times 8-choose-5 times 11-choose-7 times 5-choose-2, or 554,400 ways of drawing the observations in Table II out of the population in Table I. These computations are shown in Table III. There are four possible samples with correct marginal totals, and Table III shows the number of ways of drawing each sample from the population. The proportion of ways of drawing a given sample over the ways of drawing any of the four samples gives the probability of each possible distribution. The probability of the observed distribution is .33. The probability of the observed distribution or ones which are equally or less likely to occur by chance is .42. This is the nondirectional cumulative probability explicated in our previous article. The directional probability of the observed distribution or ones which represent a more extreme departure in a given direction—such as more positive correlation—from the most likely distribution is .34. The nondirectional test, however, is the most appropriate for the question of replication: the higher this probability the more likely is replication. Clearly, the best replication would be the $\frac{2}{3} \frac{1}{3}$ distribution, with $p = 1.0$, but one could not reject the hypothesis of replication with the p of .42 for the observed sample.

A pertinent cross-cultural example is shown in Table IV. Here the question is whether the general cross-cultural correlation between partrilineal

decent and payment of bridewealth—which several British social anthropologists contend is valid for African societies—holds good for societies of the Insular Pacific. For Murdock and White’s (1969) standard cross-cultural sample, the gamma correlation between these two features of social organization is .73. For societies outside of the Insular Pacific in the popu-

TABLE III
Replication (Interaction) Tests for a 2 × 2 × 2 Table
formed by Sampling from a 2 × 2 Distribution

Population:

3 8
11 5

Samples:		Ways of Drawing Each Sample	Point P	Cumulative P ^a	Cumulative P ^b
0	6	$\binom{3}{0}\binom{8}{6}\binom{11}{8}\binom{5}{1} = 1\ 28\ 165\ 5 =$	.01	.01	.01
8	1				
1	5	$\binom{3}{1}\binom{8}{5}\binom{11}{7}\binom{5}{2} = 3\ 56\ 330\ 10 =$	.33	.42	.34
7	2				
2	4	$\binom{3}{2}\binom{8}{4}\binom{11}{6}\binom{5}{3} = 3\ 70\ 462\ 10 =$	.58	1.00	.66 ^c
6	3				
3	3	$\binom{3}{3}\binom{8}{3}\binom{11}{5}\binom{5}{4} = 1\ 56\ 462\ 5 =$	.08	.09	.08
5	4				
Totals			1.00		
			1,677,060		

^aNondirectional

^bDirectional

^cThe lesser of two directional values the other being .92.

TABLE IV
Regional Replication of the Association between
Patrilineality and Bridewealth

Insular Pacific Societies			
Patrilineal Groups			
		Absent	Present
Bridewealth	Present	8	5
	Absent	9	9
Gamma = -.23			
Significance Test for Differences: $P = .0003$			
Societies Outside the Insular Pacific			
Patrilineal Groups			
		Absent	Present
Bridewealth	Present	11	46
	Absent	76	22
Gamma = +.87			

lation of 186 societies, the gamma correlation is .87, while for the Insular Pacific the correlation is negative at -.23. What is the likelihood that the Insular Pacific sample is a valid replication of the worldwide bivariate distribution? By our measure of replicability the statistical significance of the differences between the Insular Pacific sample and the world population is $p = .0003$. The hypothesis of replication must be rejected, and the inference of a uniform, world-wide relationship between the two variables — patrilineality and bridewealth—must be withheld. (Further interpretations of results like these are discussed below.)

As a final example, we wish to show the applicability of this measure to replications from bivariate distributions or contingency tables with a larger number of dimensions. Our measure is applicable to replications of any dimensionality, $L \times M \times N$, where there are L exclusive subsamples for the $M \times N$ bivariate distribution in the population. Table V illustrates the measure of replication for a 3x3 bivariate distribution. Cumulative nondirectional probabilities for seven comparable samples are computed as in Table III. The use of gamma to compute a cumulative directional probability is shown to illustrate the differences between the directional and nondirectional notions of cumulative probability. This point has been discussed in our previous paper (White and Pesner 1979; White, Pesner and Reitz 1983).

Replications in Systems of Variables

In comparative research it is no more meaningful to report to a single significance test for a three-variable interaction effect than it is to report a

TABLE V
Interaction Tests for a $2 \times 3 \times 3$ Table
Formed by Sampling from a 3×3 Distribution

Population:

1 4 3
3 2 0
3 2 1

Ordinal Measures

Samples:	Ways of Drawing Each Sample	Point P	Cumulative P^a	Gamma	Cumulative P^b
0 0 3	$\binom{1}{0}\binom{4}{0}\binom{3}{3}\binom{2}{1}\binom{0}{2}\binom{0}{0}\binom{2}{2}\binom{1}{1} = 18$	.04	.15	-.91	.04
1 2 0					
2 1 0					
0 0 3	$\binom{1}{0}\binom{4}{0}\binom{2}{3}\binom{2}{1}\binom{0}{0}\binom{0}{2}\binom{2}{1}\binom{1}{0} = 18$	.04	.15	-.64	.08
2 1 0					
1 2 0					
0 1 2	$\binom{1}{0}\binom{4}{1}\binom{2}{2}\binom{2}{1}\binom{0}{0}\binom{0}{2}\binom{2}{1}\binom{1}{1} = 108$	.22	.48	-.62	.30
1 2 0					
2 0 1					
0 1 2	$\binom{1}{0}\binom{4}{1}\binom{2}{2}\binom{2}{1}\binom{0}{0}\binom{0}{2}\binom{2}{1}\binom{1}{1} = 252$	.52	1.00	-.40	.70 ^c
2 1 0					
1 1 1					
1 0 2	$\binom{1}{1}\binom{4}{0}\binom{2}{2}\binom{2}{1}\binom{0}{0}\binom{0}{2}\binom{2}{1}\binom{1}{1} = 54$	.11	.26	-.20	.18
1 2 0					
1 1 1					
0 1 2	$\binom{1}{0}\binom{4}{1}\binom{2}{2}\binom{2}{1}\binom{0}{0}\binom{0}{2}\binom{2}{1}\binom{1}{1} = 12$	.03	.03	-.13	.07
3 0 0					
0 2 1					
1 0 2	$\binom{1}{0}\binom{4}{0}\binom{2}{2}\binom{2}{1}\binom{0}{0}\binom{0}{2}\binom{2}{1}\binom{1}{1} = 18$	.04	.15	+.05	.04
2 1 0					
0 2 1					

^aNondirectional

^bDirectional

^cThe lesser of two directional values the other being .82.

single significance test and correlation coefficient for a pair of variables. What is of interest are *system-wide* characteristics.

A *homogeneous system* of variables is one where there are first-order interactions—ordinary correlations—among pairs of variables but no interactions (or no more than expected by chance), of a higher order. Typically by homogeneous system we will mean a homogeneous system of order one, as above, unless otherwise specified. *Inferences can be made from the correlational evidence of homogeneous systems* because they satisfy the criterion of all-possible replications holding good considering each variable in the system as a potential control condition for replications of bivariate relationships among pairs of other variables.

In this paper these replications are tested holding *all* univariate and bivariate distributions constant, thus making the question one of second-order interaction. There is no compelling *a priori* reason why this must be done. In fact, the determining factor is the specific formulation of the null hypothesis which is to be tested. In the case of the typical interaction analysis of a three-variable relationship only the original bivariate relationship should be held constant; formulas for this case are given in White and Pesner (1979). It is also possible to only control two of the three bivariate relationships or to mix different kinds of triples in a system analysis; formulas can be modified for that purpose. However, if the focus of investigation is systemic structure modeled on the basis of bivariate relationships, as is the case for example with material entailment analysis or causal modeling (path analysis), then all bivariate relationships in the system should be held constant. The variable acting as control in one test plays the part of an original variable in other tests; bivariate relationships held constant in one test should be held constant in all tests if system characteristics are of interest.

From the viewpoint of the control condition, homogeneity is satisfied when there is no variable in the system which creates second-order interaction effects. Our method essentially calls, therefore, for testing the null hypothesis of no second-order interaction for each variable.

Each variable is potentially interacting with every possible pair of other variables. For each new pair, the formulas discussed above will identify the probability of the observed outcome or a less likely outcome under the null hypothesis. These results can be combined if it is assumed that each is an *independent* test of the same null hypothesis. This assumption of independence is in fact justified only if no four-way interactions are present in the system. Although we assume this here, clearly such an assumption can and should be tested statistically. However, just as a test of two-way interaction (correlation) assumes the absence of three-way interaction and a test of

three-way interaction assumes the absence of four-way interaction, any test of n -way interaction must assume the absence of $(N+1)$ -way interaction. Thus from a logical standpoint, interaction tests should start "at the top" and work down, i.e., in a system of N variables the first test should be for N -way interaction. From a computational standpoint such an approach is impractical.³

Methods for combining multiple independent tests of a given hypothesis were first introduced by Fisher (1932-8), who analyzed the case where each test was based on a continuous distribution. The discrete case was analyzed by Wallis (1942); we use his formulas here. The result is probability which can be compared to conventional levels of significance such as .05, etc. However, application of the Wallis formula is also computationally impractical for even a small system of variables, even for a large computer. Bartlett (1968) gives an approximation formula which makes this solution practical.⁴

Illustrations

We consider as an example the study of the sexual division of labor across 50 subsistence and manufacturing activities in preindustrial societies, conducted by White, Burton and Brudner (1977; also Burton, Brudner and White 1977). They found entailments between which sex does task A and which does task B, for particular pairs of tasks. For this model, the performance of one task acts as a constraint on the performance of other tasks, each constraint being expressed as a bivariate relationship. To test the replicability of these bivariate relationships, we consider the assignment by sex of each other task, C, as a possible control condition. Thus, there are as many possible subsamples for this replication as there are categories of societies which share some particular feature on other tasks. These are the naturally occurring subsamples for this system of variables. The bivariate relationships thus can be said to replicate for this system of variables if and only if the distribution of replications in naturally occurring subsamples is close to that expected to occur by chance under the null hypothesis of no second order interaction.

The results of this process are illustrated by the findings in Table VI. For the 50 division of labor variables, dichotomized at females equal or predominant in the task versus males predominant in the task, and taking each such dichotomy as a control condition for replication, the 2x2x2 tables formed with each other pair of variables have been examined. For 58,800 such tables, we have computed the measure of replication. Naturally, this is

possible only by computer. As can be seen, the actual frequencies of interactions at high levels of significance are no greater than expected by chance for this system of bivariate relationships. Division of labor variables constitute a homogeneous system of variables. Any of the correlations among the variables will replicate in any subsample defined by other variables. Bivariate relationships—correlations or entailments—among the division of labor variables are inferentially meaningful, although this analysis has not told us what kinds of inferences from them will be valid. In our earlier work, we argued for structural inferences or constraint relationships among these variables, rather than causal inferences. Neither form of inference, however, would be valid if the test of homogeneity had not been satisfied. This test, however, is a highly stringent one and not one which it is likely *a priori* that any natural observation dataset would pass.

Many, if not most, nonexperimental datasets will not pass the test of all-possible-replications of system homogeneity. For example, we took the six major world regions plus the first eight social structural variables for Murdock's ethnographic atlas, using the Murdock and White (1969) standard sample. Table VII shows that at high levels of significance, the number of actual interactions exceeded the expected interactions: relationships between social structural variables do not replicate under different

TABLE VI
Expected and Actual Frequencies of Interactions
(by Significance Levels) for 50 Division of Labor Variables

	Expected	Actual	Tables Examined ^a	Binomial Test ^a
$p \leq .001$	.04	0		
$.001 < p \leq .003$	.11	0		
$.003 < p \leq .01$	.6	2 ^a	194	$p = .12$
$.01 < p \leq .03$	2.6	2		
$.03 < p \leq .1$	15.7	14		
$.1 < p \leq .3$	81.8	86		

^aComputed only where actual exceeds expected at a high significance level.

TABLE VII
Expected and Actual Frequencies of Interactions
for 37 Social Structure Variables and Six World Regions

	Expected	Actual	Tables Examined ^a	Binomial Test
$p \leq .001$	.06	1	406	$p = .06$
$.001 < p \leq .003$	.21	2	224	$p = .03$
$.003 < p \leq .01$	.99	1	315	$p = .63$
$.01 < p \leq .03$	3.9	8	426	$p = .40$
$.03 < p \leq .1$	21.9	33		
$.1 < p \leq .3$	112.9	118		

control conditions as well as might be expected by chance in a homogenous system of bivariate relationships. Table IV is an example of a striking lack of replication. The dataset consisting of social structural variables is *heterogeneous*.

Unless a dataset satisfies the criterion of homogeneity, modes of multivariate analysis such as partial correlation, factor analysis, cluster analysis, multidimensional scaling, Guttman scaling and entailment analysis are not valid representations of the data structure, and no valid statistical inferences can be based upon their results.

In order to demonstrate homogeneity in cases where many of the above-mentioned techniques are to be applied, our approach would have to be developed for the continuous case. This would be quite a complicated endeavor. The complications can be hinted at by pointing out that testing interaction involves holding the various bivariate relations constant. This in turn necessitates a transformation of the probability distribution based on the null hypothesis of no second-order interaction (*cf.* White and Pesner, 1979, p. 9-10). Clearly, however, this problem needs attention. Such techniques as are listed above are quite commonly used; until methods for establishing homogeneity in the continuous case are developed, there is a significant theoretical lacunae whenever anyone uses them.

The criterion of homogeneity, although stringent, is not impossible to satisfy in cross-cultural datasets. We have seen one example. Table VIII shows a second example, from an unpublished study to kin role avoidance in 24 kinship dyads in a sample of 218 societies. The criterion of perfect homogeneity or systemic replication is much closer to being satisfied than in the case of the social structural variables.

The Problem of System Boundaries

Every system has its limits. When we make a statement about a system which can be tested in terms of a predicted correlation, the criterion of all-possible-replications demands that the correlation replicate in every theoretically defined condition within the system, but not that the correlation replicate outside the boundaries of the system. Thus, every valid statement about a system must also define the limits of that system.

A system is thus defined by a constraint set of internal relations. Given a model of these internal relations, one can model the boundaries of a system.

TABLE VIII
Expected and Actual Frequencies of Interactions
for 24 Kin Role Avoidance Variables

	Expected	Actual	Tables Examined ^a	Binomial Test
$p = .001$	.0	0		
.001 $p = .003$	.0	0		
.003 $p = .01$	.0	0		
.01 $p = .03$	.0	0		
.03 $p = .1$	.3	1	9	$p = .26$
.1 $p = .3$	.8	1		

Consider the relationship in Table IV. Are there two world systems, in one of which patrilineality and bridewealth are positively connected, while they are negatively related in the other? If so, is Insular Pacific versus Other Regions the correct boundary of the two systems? Or is there some other boundary, such as Pacific Atoll ecology versus other environments which differentiates these two systems even more sharply? If systems are characterized by constant and internally replicable correlations, then system boundaries are defined by parameters which, taken as control variables, have the greatest interaction effects with other variables.

Let us sharpen these definitions. Specification of a system involves (1) parameter conditions which state which cases belong to systems X, Y, Z, and which do not; (2) specification of which attributes belong to a given system as variables; (3) internal specification of a system in terms of relations among attributes (for which there are constant and internally replicable correlations across the cases within the system). Constancy and replication of the latter correlations is not expected among cases which span different systems. This largely corresponds to the conception of comparative research put forth by Prezowski and Teune (1970) in their *Logic of Comparative Social Inquiry* which we consider to be the best work in its class to date. Whereas they are concerned with the identification of systems in terms of multiple level hierarchical organizations (such as nation-states and their component individual members) we adopt the view that any parameter—hierarchical or not—can potentially serve as a system boundary. Climatic regions, for example, might serve as system parameters for systems of social organization variables.

In our view, the systemic replication problem has three complementary uses which are fundamental to comparative research. One use is as a test of homogeneity as a precondition to making meaningful inferences from comparative evidence. The second, implicit in the above, is a partial solution to Galton's problem, which is largely a problem of replication (see also Dow, White and Hansen 1980; White, Burton and Dow 1981; Dow, White and Burton 1982; Dow, White and Burton 1983). The third, drawing on negative results in the homogeneity test, is as a vehicle for the decomposition of a heterogeneous population—containing multiple systems—into homogeneous subsets. If, for example, the social organization variables in Table IX are homogeneous within the Insular Pacific, and homogeneous for the remainder of the world, while heterogeneous for the world as a whole, then it would seem that the Insular Pacific region, or some closely related parameter, is a system boundary which should be taken into account in the formulation of theories of social organization. This distinction cor-

TABLE IX
 Expected and Actual Frequencies of Interactions
 for 37 Social Structure and 6 Regional Variables
 Frequency of Replications at
 $p \leq .003$

	Expected	Actual
1 Africa	.01	0
2 Circum-Mediterranean	.03	0
3 East Eurasia	.02	0
4 Insular Pacific	.03	1
5 North America	.04	1
6 South America	.03	0
7 Bridewealth	.05	2*
8 Brideservice	.01	0
9 Token Brideprice	.01	0
10 Gift Exchange	.01	0
11 Sister Exchange	.00	0
12 No Consideration	.04	0
13 Dowry	.00	0
14 Monogamy	.00	0
15 Limited Polygyny	.05	0
16 Polyandry	.00	0
17 Normal Polygyny	.00	0
18 Abnormal Polygyny	.01	0
19 Stem Families	.00	0
20 Small Extd. Families	.04	1
21 Patrilocal	.05	0
22 Bi/Neolocal	.03	0
23 Matrilocal	.04	0
24 Patrilineality Absent	.05	2*
25 PL Exogamy	.00	0
26 PL Lineages	.01	0
27 PL Sibs	.03	0
28 PL Phratries	.00	0
29 PL Moieties	.00	0
30 Matrilineality Absent	.06	1
31 ML Exogamy	.00	0
32 ML Lineages	.00	0
33 ML Sibs	.01	0
34 ML Phratries	.00	0
35 ML Moieties	.00	0
36 Bilateral Descent	.04	1
37 Kindreds	.03	0
38 Ambilineal Descent	.00	0
39 Ramages	.00	0
40 Exogamous Ramages	.00	0
41 Quasi-lineages	.00	0
42 Pastoral Society	.00	0
	.81	9

*Variables with Greatest Interaction Effects.

Decomposition of a Population

Table IX illustrates the type of evidence which is needed for decomposition of a heterogeneous system into constituent homogeneous systems. The variables with greatest interactive effects are identified, the population partitioned on the basis of these variables, and the homogeneity test is repeated within each partition. In repeating the test, further divisions may be found within partitions. There is no guarantee that the homogeneity which may be achieved, however, is meaningful. The system boundaries which are identified must be externally replicated with new evidence (congruent validation). The procedure has heuristic importance, however, in that it leads towards a major form of validation in comparative research: identification of systems in terms of their invariant properties or internal relations.

Interactional Systems

Invariance of relations among attributes is not the only form of system, however. Systems may also be constituted by interactive relations between units (White, Burton and Dow 1981) and invariance of relations among relations (White 1980). This is the other aspect of the problem of validation, of Galton's problem and of the problem of system identification.

As we stated above, no empirical correlation is inferentially meaningful—i.e., one is not entitled to draw causal or structural inferences—unless the correlation can be internally replicated within all of the constituent (theoretically defined) subunits of the system to which the correlation is ascribed. We will now state a second principle which is explored in a separate paper by Dow, White and Burton (1981). No empirical correlation is properly estimated if either of the variables is significantly autocorrelated in space, time or any network interaction matrix.⁵ Most variables in comparative research, survey research and nonexperimental research in general suffer from one form or another of autocorrelation. The foundations of comparative research must be changed in this regard: our theories and models must specify both the structure of interactional phenomena and internal inter-attribute relations, that is, the between-case and the within-case interactions. Autocorrelation is evidence of interactional phenomena which must be represented in the attempt to explain any given phenomena, or draw causal or structural inferences. Methods for the consistent estimation of properly specified interactive systems in terms of regression equations are now available (Dow, White, and Burton 1982).

Internal replication, the modelling of autocorrelation, and the explicit use of the systems-with-invariance concept in hypothesis testing provide the foundational sources of validity in making inferences from comparative or nonexperimental evidence. These are not inductive methods, however, in that each depends on specification of systems of variables and network processes. At the present stage of comparative research, one good theory is worth a score of techniques.

Weaknesses of the True Experiment

The true experiment is actually inferior in several respects to the more powerful comparative method outlined above. The experimenter cannot specify from a single experiment or even a string of experiments the conditions (type of subjects, type of setting, etc.) under which the experimental effect will take place. Background data on experimental conditions needed for such an evaluation are rarely collected and are a part of non-experimental rather than experimental research. By collecting all such background data as are deemed theoretically relevant and by testing for internal replication under all possible background conditions within the population, the nonexperimental method possesses a source of validation which the experimental method lacks. The experimenter often cannot correctly specify what the experimental effect is, since many of the effects created by the experimenter are unintended. Autocorrelation or interaction problems (interaction between subjects) may be reduced in the experiment (which is probably the chief benefit), but they are rarely eliminated entirely. The large discrepancies between experimental and nonexperimental research on many topics is largely because of the suppression of social interaction effects in experiments as opposed to their predominance in natural situations. The experimental method is thus not very useful in generalizing to natural situations. *The strength of the comparative or nonexperimental method should be its ability to take interactional phenomena into account. This development, however, is only in its infancy.*

There are several signs of progress for interactional studies in comparative research. In Beatrice and John Whiting's (1974) research, which has provided a highly productive theoretical model for psychological anthropology, focus has shifted from relations between attributes such as child training and personality characteristics to the study of specific types of social interactions in relations to personal attributes. In studies of social structure (White 1979) the application of the social networks paradigm is found to be productive in the explanation of kinship role behavior—joking,

avoidance, respect relations and the like. In political science, Prezowski and Teune (1974) have influenced a handful of people who are working with the systems concepts and more interactionally-oriented approaches (although the two do not necessarily go hand in hand). "Context effects" have become a major problem in political and social surveys. Burton and Kirk's work (1979) on the explanation of internal cultural variability using methods of cognitive anthropology is increasingly focusing on social interaction processes. There are many other examples. The 1950's and 60's was a positivistic period in comparative research, with an inductive approach to theory construction, a stress on empirical falsification, and the stripping away of "false" theories in order to leave the fittest survivors (Brudner and White 1979). Some of the directions emerging now in comparative research are different. The principles discussed here can be briefly summarized. In isolation, correlations are inferentially meaningless. They acquire inferential meaning as we identify their invariance across internal replications within theoretically defined systems, and as each variable is checked for autocorrelation effects in terms of theoretically defined interactive systems.

NOTES

¹Campbell and Stanley (1963:3), for example, "recognize experimentation as the basic language of proof, as the only decision court for disagreement between rival theories," but that "the experiments we do today, if successful, will need replication and cross-validation at other times under other conditions before they can become an established part of science." Replication of experimental results is, in short, an established source of validation of theory. In the same book, Campbell and Stanley present 12 major sources of invalidity in experimental research designs. They note that experimental designs are capable of eliminating each of these 12 "rival hypotheses" and thereby lending a greater "degree of confirmation" (p. 36) to the theory. Nonexperimental designs, however, are not capable of eliminating all 12 rival hypotheses or "sources of invalidity" and thus are useful in rejecting hypotheses but not in providing strong confirmations. This one-sided view is adopted by those who argue that cross-cultural research is useful in falsifying theories, but does not provide for the validation of theories. While we do not address each of Campbell and Stanley's 12 sources of invalidity here, our general argument is that there are appropriate tests for each of these "rival" hypotheses in nonexperimental research, and thus "sources of validity" or validation in nonexperimental research as well as "sources of invalidity."

²Replication under within-system conditions does not imply that the relationships will replicate under other conditions outside the system specified. Similarly, the finding that an experimental result replicates within a given experimental design does not imply that it will replicate in other designs.

³It turns out that the necessary computations increase only to some level of interaction determined by the size of the sample; at levels of higher-way interaction the sample size determines all conditional distributions. For more on this see Pesner (1980), where certain methods are presented for determining the highest level on which interaction could be present in a give data set.

⁴While Bartlett's formulas are correct, we did find that his empirical example was incorrectly computed. This was verified by two independent investigators .

⁵In regression equations, we can obtain unbiased estimates of the regression coefficients by ordinary least squares even if significant autocorrelation is present, but estimates of the variance of these coefficients will, in general, be seriously incorrect. Significance tests and correlation coefficients will also be incorrectly estimated.

REFERENCES

- Bartlett, M.S.
 1968 Communication. *Nature* 220:1251.
- Brudner-White, Lilyan A. and Douglas R. White
 1977 Review of Levinson, David, ed. *A Guide to Social Theory: Worldwide Cross-Cultural Tests*. *American Anthropologist* 81:952-954.
- Burton, Michael, Lilyan A. Brudner, and Douglas R. White
 1977 A Model of the Division of Labor by Sex. *American Ethnologist* 4:227-251.
- Burton, Michael, and Lorraine Kirk
 1979 Sex Differences in Maasai: Cognition of Personality & Social Identity. *American Anthropologist* 81.
- Campbell, Donald T., and Julian C. Stanley
 1963 *Experimental and Quasi-Experimental Design for Research*. Chicago: Rand-McNally.
- Dow, Malcolm, Douglas R. White, and Michael Burton
 1981 Multivariate Modeling with Interdependent Network Data. In Press: Behavior Science Research.
- Dow, Malcolm, Douglas R. White, and Michael Burton
 1982 Network Autocorrelation: A Simulation Study of a Foundational Problem in Regression and Survey Research. *Social Networks* 4:169-200.
- Dow, Malcolm, Douglas R. White, and David Hansen
 1980 Regression Analysis and Network Autocorrelation. Ms.
- Fisher, R.A.
 1932 *Statistical Methods for Research Workers*. Fourth and later Editions. London: Section 21.1
- Murdock, G.P. and Douglas R. White
 1969 Standard Cross-Cultural Sample. *Ethnology* 8:329-369.

- Pesner, Robert
1980 On Identifying Possibilities of N-Way Interaction. Ms.
- Przeworski, Adam and Henry Teune
1970 *Logic of Comparative Social Inquiry*. New York: Wiley & Sons.
- Wallis, W. Allen
1941 Compounding Probabilities from Independent Significance Tests. *Econometrika*, 10:229-248.
- White, Douglas, R.
1979 Avoidance in Social Networks.
- White, Douglas R.
1980 Network Structure and Category Theory. Ms.
- White, Douglas, R., Michael Burton, and Brudner, Lilyan
1977 Entailment Theory and Method: A Cross-Cultural Analysis of the Sexual Division of Labor. *Behavior Science Research* 12:1-24.
- White, Douglas R., Michael L. Burton, and Malcolm M. Dow
1981 Sexual Division of Labor in African Agriculture: A Network Autocorrelation Analysis. *American Anthropologist* 83:824-849.
- White, Douglas R., and Robert Pesner
1979 An Exact Significance Test for Second-Order Interaction Effects or Differences Between Discrete Bivariate Distributions. *Social Science Research Reports*, 29. School of Social Sciences, University of California, Irvine.
- White, Douglas R., Robert Pesner, and Karl Reitz
1983 An Exact Significance Test for Interaction Effects. *Behavior Science Research* 18 (in Press).
- Whiting, Beatrice and John Whiting
1974 *Children of Six Cultures: A Psycho-Cultural Analysis*. Cambridge-Harvard University Press.