Baker Service inquiries: (phone) 510-642-2538; (email) baker@library.berkeley.edu **Order# 062906-2**

UCB Order via: (Pathfinder) http://sunsite5.berkeley.edu:8000/ ; CDL Request; (Gladis) "Page Baker"; (Email) requests@library.berkeley.edu; (Fax) 510-643-5598

Baker is your proxy borrower and checks out books directly to you. Return books by the due date to the lending library. To renew materials, contact the lending library directly. In accordance with library circulation policy you are responsible for the safe handling and return of materials and for any fines or replacement bills.

PATHFINDER
David Collier **WEB DELI**
0053914
Political Science--210 Barrows Hall
O:2-6323 M:2-7910 dcollier@berkeley.edu
status: 3(faculty)

no. 1-3 = **PHOTOCOPY**

Loc/Call No: **Main Stack HM258 .S76**
Article Title: Avoiding Regression Effects in panel studies ....
Article Author: Campbell and Clayton
Volume: 3 Pages: 99-118
Title: Studies in public communication. Publisher: Committee on Communication, University of Chicago, 1957-1962.
ISSN: 0585-704X

NRLF B 3 992 428

Gladis ☑ Bmc
Melvyl ☐ ______
WCAT/RCAT ☐ ______
OTHER ☐ ______

Action Taken ______

Date Filled: ______
Filled from Web Resource ☐

Report Sent: ______

☐ This is non circulating at: ______
☐ This is on ______ reserve at: ______
☐ Out to another borrower, recalls have been placed at: ______
☐ Not on the shelf, searches have been placed at: ______

☐ This book/serial not owned by UCB
☐ This book/issue is missing at: ______
☐ This issue is not owned/not yet received /at bindery at: ______
☐ This request has been forwarded to Interlibrary Borrowing automatically.
☐ Notify us if you wish this request to be referred to Interlibrary Borrowing.

☐ We cannot verify this publication/article. Please return this slip with any corrections or additional information available. If possible send us a copy of your original citation.
☐ Blocked/Autoblocked
☐ OTHER: ______

DTD/Ariel
NRLF Access Services
Date 6/30/06
Initials ______
PS 7KA
Pages 21

baker. 30

lished
go and
grams
N has
of
and

er and
of . . .

er 1,
where
ders,
to:

Studies in Public Communication

AVOIDING REGRESSION EFFECTS IN PANEL STUDIES OF COMMUNICATION IMPACT*

Donald T. Campbell and Keith N. Clayton**

The panel technique in public opinion surveys has offered the social sciences an opportunity to go beyond the usual static correlational data with their interminable ambiguity as to direction of effect, into something like experimentation. By employing measurements extended in time, asymmetries of antecedence-subsequence in correlation can be noted which make possible some interpretation of the direction of effect between variables. Two modes of analysis have been developed by the Bureau of Applied Social Research for this purpose. While the most ingenious may well be the "sixteen-fold table", the most straightforward is the panel analysis of the effects of an intervening communication, as in the studies by Kitt and Gleicher [1] and by Glock [2], designated in this paper as the "panel-impact" design. It is this design that receives primary attention.

Along with other experimental designs making use of naturally assembled groups (rather than experimentally assembled groups equated by randomization), the panel-impact design is liable to the confounding of statistical regression effects with changes due to the experimental variable. The "turnover table" form of data presentation employed in the panel-impact design turns out to be particularly vulnerable, as will be illustrated after a brief review of the problem of statistical regression as an artifact confusable with more substantive processes.

---

*The preparation of this paper was made possible by the authors' participation in Northwestern University's Psychology-Education Project, sponsored by the Carnegie Corporation. It represents a part of a general survey by the senior author of the quasi-experimental research designs available in natural field situations, and hence available for research in educational settings. The paper is also intended to be a chapter in a projected volume on panel studies to be edited by Paul F. Lazarsfeld and to be entitled The Study of Short-Run Social Change. The authors have had the advantage of suggestions made to a previous draft by C.I. Hovland, O. D. Duncan, J. A. Davis, R. F. Winch, D. L. Thistlethwaite, and especially Thomas J. Banta.

**Dr. Campbell is Professor of Psychology at Northwestern University, where he has been since 1953, and has taught at Ohio State University (1947-50), the University of Chicago (1950-53), and Yale University (Visiting, 1955). While at the University of Chicago he was a member of the Committee on Communication. Although he does not regard himself as primarily a methodologist, he is probably best known for his several contributions in this area. His most recent work is Experimenting, Validating, Knowing: Problems of Method in the Social Sciences (New York: McGraw-Hill, in press).

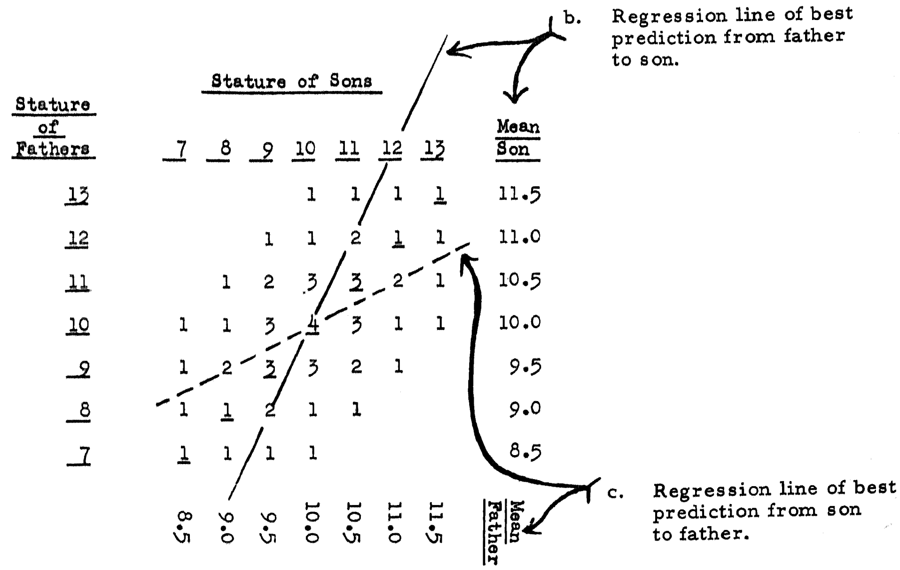
Dr. Clayton received his Ph. D. in Psychology at Northwestern University in 1960, and is presently Assistant Professor of Psychology at Vanderbilt University, where he is teaching in the field of experimental psychology.

Review of the Regression Problem

Two-variable regression problems are most frequently encountered. The simple fact of imperfect correlation between two variables implies as a tautology that those units selected as most deviant upon one measure will average nearer to the mean of another. If a time or stage difference distinguishes the two variables, and if the investigator is interested in examining the "fate" of those initially extreme in one direction or the other, this inevitable accompaniment of a correlation less than unity is frequently misinterpreted as a biological or sociological function of the temporal or ordinal sequence. For example, when intelligence tests are given one year apart in an orphan asylum, the children brightest on the first test are, on the average, somewhat duller on the second (though of course still above average), while the dullest now average brighter than they did initially. This is an inevitable finding if the test-retest correlation is less than unity, but it has been frequently misinterpreted as evidence of a "leveling" effect of the homogeneous environment.

While Galton's understanding of the phenomenon was more complete and subtle than that of many of those following him, he encountered regression before the development of the symmetrical correlation coefficient and his mode of presentation led to an initial perpetration of the mistaken inference. Thus in his 1877 address [3] he said, "the progeny of all exceptional individuals tends to 'revert' towards mediocrity", and he presented but a single "reversion" or "regression" line, showing that the average progeny of sweet peas reverts from the parent size about one-third the way back to the mean of the group. However, even then he noted one fact which most misinterpreters of two-variable regression have failed to examine, i. e., the variance of the filial generation was as great as that of the parent. (This fact is at odds with the interpretations of "conformity" and "leveling", etc., which imply a reduction in variance on the second measure.) But to handle this maintenance of variety, Galton hypothesized a second mechanism working opposite to and compensating the effects of regression. His second major paper on the problem bore the connotatively misleading title "Regression towards Mediocrity in Hereditary Stature" [4] and retained the two compensating processes of regression and dispersion, but did present the two complementary regression lines, i. e., adding the "regression" of parents on offspring, and thus recognized that not only did exceptional parents have offspring more mediocre than themselves, but also exceptional offspring came from parentage more mediocre than they. In an epoch in which the average stature is not increasing, the sons of tall fathers, for example, average shorter than their fathers, but the fathers of tall sons average shorter than their sons. Fig. 1 attempts to provide an intuitive understanding of why such regression is a tautological restatement of imperfect correlation.

In Fig. 1a, there is shown for each group of fathers, classified by height, the distribution of sons' heights. Thus, of the four fathers 13 units high, one has a son 10 units, one 11, one 12, and one a son 13 units high. If the relationship were perfect, all of the cases would fall along the underlined diagonal values. Instead, there is considerable scatter. While the data here are hypothetical, they do illustrate a typical father-son correlation of .50, under conditions where fathers and sons have equal means and variabilities. Note that a by-product of the fact of scatter is that our best prediction of a son's height is not



1. a. Frequency Scatter of Sons' Stature for Each Class of Fathers, and Vice Versa.

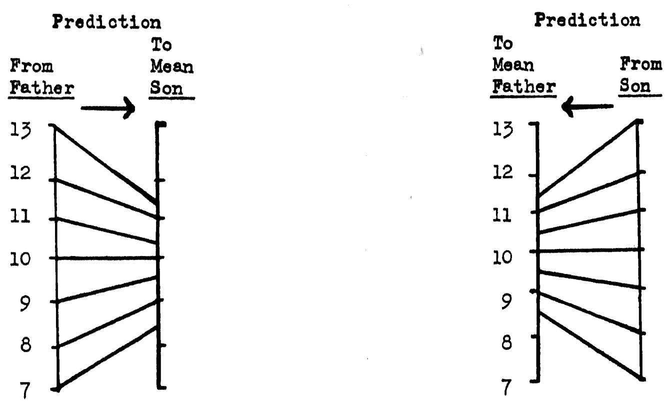


FIGURE 1. REGRESSION IN THE PREDICTION OF SONS' STATURE FROM FATHERS' AND VICE VERSA.

his father's height, but rather a value somewhat closer to the average of the class of sons as a whole than his father was to the class of fathers as a whole. This automatic by-product of scatter or imperfect relationship is what is meant by regression. Galton was interested in inheritance from fathers to sons, thus tried to predict from fathers to sons, and thus first noted regression line b in Fig. 1a. Corresponding to the correlation value of .50, the mean son has regressed halfway to the group mean of 10. (Had the correlation been zero, the regression line would have been vertical, the total group mean being the best prediction of the son's height no matter what the father's height was.) Fathers of 13 have sons that average 11.5, fathers of 12 have sons that average 11, fathers of 10 have sons that average 10, etc. This has alternately been diagrammed in Fig. 1b, in which form Galton's dynamic-process misinterpretation is easily made. The apparent reduction in the variability of sons is specious, as reference to 1a shows. Presentation 1b is misleading because it does not make clear the important qualitative difference in the values plotted for fathers (selected for those values) and those plotted for sons (resultant averages). This qualitative difference is seen clearly in the rows of Fig. 1a, in which fathers have been selected in homogeneous groups while the sons show the scatter. Entering the table column by column, of course, reverses the picture. Those making the dynamic-process misinterpretation are unprepared for Fig. 1c, which seems to show the reverse effect when one starts with selected groups of sons and then finds the corresponding average father. In Fig. 1a, line c shows this regression. (The artificial values in Fig. 1a have been selected so as to make visual determination of row (and column) means easy.)

Hovland _et al_[5] have discussed the problem as it relates to the amount of attitude _change_ shown by persons of various initial attitudes, where the pretest-posttest correlation is less then 1.00. McNemar[6] has illustrated in detail how this error has operated in studies of institutional effects upon the intelligence of children. For these two-variable (i.e., test and retest) regression situations the two features most readily illustrative of the fallaciousness of "dynamic" interpretations are the examination for reduction in population variability and the examination of reverse direction regression. In the studies which McNemar criticized these two analyses had been omitted.

Differential Regression in Three-Variable Problems

To complete the background for the understanding of regression-confounds in panel-impact studies, we need to examine three-variable problems and the occurrence of differential regression upon the part of "matched" subgroups, for which a third variable has been "controlled". We will attempt to approach the problem first in a setting in which the reader's intuition will be in agreement with the statistical analysis. Let us consider a course in which three similar one-hour examinations, _A_, _B_, and _C_, have been given, and in which each correlates .50 _with_ each _of_ the others. Let us now try to study the relation between _A_ and _C_, as qualified by _B_. Without this qualification, we would expect _students_ scoring high on _A_ to be above average, though not so high on _C_. But if we break _those_ scoring high on _A_ into those who scored high _also_ on B, and compare them with those _rare_ persons scoring high on _A_ but low on _B_, we would find the first group higher on _C_ than the _second_. From a _common_ sense point of view, we can see _that_ the "consistent performers" on _A_ and _B_ are more likely to do well on C, and this is indeed what both regression algebra and actual classroom experience does show. From the point of view of the statistics of regression, the situation which is intuitively obvious here is the same as that one ubiquitously found in the social sciences today under such names as "qualifier" analysis, or "elaboration", even when the three variables involved include dichotomous ones and are designated by dissimilar labels. The differential outcome on C for the two subgroups "equal" on A but differing on B says nothing more than the initial statement that all three correlated positively. The more dynamic interpretations, with the implications of additional "discovery" through the "elaboration", are in most instances mistaken, and could only represent discovery in cases of nonlinear relationships, usually absent. For situations in which three variables are positively but imperfectly correlated, it is a tautology that of two groups equally high on one variable, those also high on a second will have regressed less on the third than those low on the second.

The situation is sufficiently complex and overlooked to justify a more detailed statement in regression algebra. In the hypothetical instance of the three examinations with all intercorrelations .50, let us express the scores in z scores (i. e., transformed so that for each examination, the mean = 0, the standard deviation 1.00). As in any case of scores of equal means and standard deviations, the correlation coefficient states the amount of regression. Taking as our case those + 1.0 on A, we get:

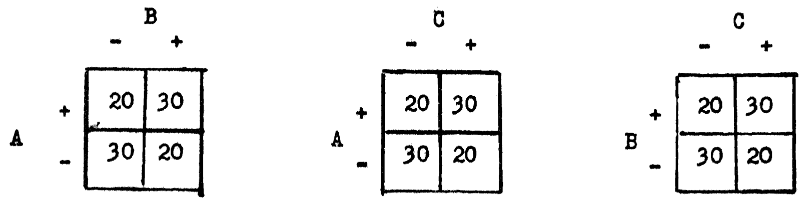
For all those +1.00 on A, the predicted mean on C is +.50
For those +1.00 on A and +1.00 on B, mean C is +.67
For those +1.00 on A and -1.00 on B, mean C is .00 (a rare group)

The regression values for the three-variable cases can be computed in several ways. Perhaps the simplest is to note that not only do A and B separately predict C, but also that the sum of A + B predicts C. This correlation is actually higher than for A or B separately, and by the multiple correlation formula is found to be .58. The A + B sum for the first subgroup is +2.00, that for the second 0.00. The standard deviation of the composite A + B is of course larger than that for A or B alone, and is in fact 1.73 $(=\sqrt{3} = \sqrt{\sigma_a^2 + \sigma_b^2 + 2r_{ab}\sigma_a\sigma_b})$. The mean of the composite is still zero. The sum of 2.0 thus lies 1.16 standard deviation from the mean of the composite, which, when regressed .58 to C, gives a value of .67. Note that because of the positive correlation, it would be hard to find many cases +1.00 on A and -1.00 on B. A general caution is that the harder it is to find cases in a given cell, the more regression is to be expected. A more general rule is that any variable (e. g. C) is better predicted from the sum of two predictors than from either alone. Reinterpreting qualifier "matchings", or "partialings", or other multiple-variable "elaborations" as the summing of predictors would obviate most of the overinterpretation of such analyses now current.

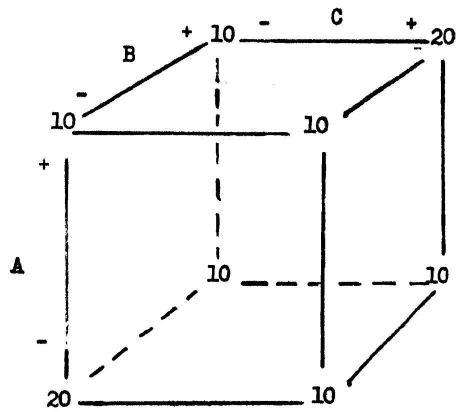
Thus a student's intelligence, his likelihood of completing high school, and his later income represent three intercorrelated variables. If one examines the "effects" of completing high school by selecting two groups of students, completing and not completing, who are matched on I. Q. and then finds that those completing high school earn more in later life, one is apt to infer from the incidental temporal order a causal asymmetry that is not justified. All that such evidence shows, if it goes no further, is that the three variables are positively but less than perfectly related. Many findings having to do with cognitive consistency and attitude stability can be similarly interpreted.

The three variables in such cases might be an initial measure of an attitude, a related value, and a second measure of the attitude. Those high on both of the first two symptoms will be higher on the third, no matter in what temporal order one takes the variables. Stating this as greater attitude stability for those with congruent values and attitudes overdresses the simple illustration of imperfect positive correlation among the three variables. Analogously, a respondent's reports on the opinions of his acquaintances correlates with his expressions of his own opinions on both of two separate occasions, generating the over-specific, misleadingly "dynamic" interpretation of greater attitude stability among those with social support for their attitudes. This is not to deny the truth of this allegation, but merely to point out that three-variable regression results in trends which can be mistakenly taken as supporting it. For another example, with positive but imperfect correlations among ascribed social status, class identification, and nonanomie one could find a significant relationship between class identification and nonanomie even with social class "held constant", or within subgroups "identical" in social class, generating misleading interpretations of factorial complexity which would not have resulted from a factor analysis of the original correlations. Note the way in which the labeling of variables contributes here: Let us suppose that ascribed social status represents an interviewer's rating, and that this rating has been done independently by two raters, resulting in positive but imperfect correlation between the two, and between each and a questionnaire measure of nonanomie. If we then were to equate groups on the first rating, we would still find a residual significant relationship between the second rater's ratings and nonanomie, and in this case would readily accept the interpretation that two raters are better than one, that the sum of two imperfect predictors is better than either alone. Yet in the original presentation we overlooked the fact that the results were perfectly consistent with the interpretation that class identification was just another imperfect measure of social class.

Such results are <u>not</u> specific to conditions in which underlying continuous measures and normal curve statistics may be assumed. They are rather common to probabilistic relationships in general. Consider three dichotomous variables all related to the same degree, the three two-by-two plots all taking the form shown in Fig. 2a, and the three-dimensional plot taking the form of Fig. 2b. Focus now on the relation between <u>A</u> and <u>C</u> as qualified by <u>B</u>. Fifty persons are high on <u>A</u>. Without regard to <u>B</u>, there is regression on <u>C</u> to the extent that 60% (30/50) are high on <u>C</u>. These fifty can be divided into 30 who are also high on <u>B</u> and 20 who are low on <u>B</u>. When their values on <u>C</u> are examined, the "consistent" 30 have "regressed" less, 67% (20/30) of them being high on <u>C</u>, while the other 20 have "regressed" farther, only 50% being high on <u>C</u>.



2. a. Two-by-Two Plots.



2. b. Three-Dimensional Plot.

FIGURE 2. ILLUSTRATION OF THREE-VARIABLE REGRESSION FOR DICHOTOMOUS DATA

Three-Dimensional Regression Effects in Quasi-Experimental Designs

When one has a field experiment in which the three variables are pretest, exposure to the persuasive message, and posttest, and where exposure itself is correlated with the test scores, one meets the three-variable regression problem in a particularly subtle form. Thorndike[7] has analysed the regression fallacy generated where the experimenter attempts to correct for the mean differences in pretest between his experimental and control group respondents by employing a matched subset of each. As he explains it, the matched pretest cases tend to be from the opposite extremes of their respective groups, and upon re-testing can thus be expected to regress in opposite ways, i. e., to the means of their respective groups, thus generating second-test differences between the two groups which would have occurred even if the "second" test had been administered before the one upon which the matching was done. Or, in terms of the mode of exposition here employed, group membership can be treated as the third variable, positively but imperfectly correlated with both pretest and posttest. Panel studies of the effect of an intervening communication are similarly vulnerable. Two well-known examples will be examined in this regard.

In Glock's [8] study of the effect of the motion picture Gentlemen's Agreement upon attitudes, the major results are presented in Table 1. To make expectation and outcome more apparent to the reader, we

Table 1

Glock's Turnover Table for Effect of "Gentlemen's Agreement" (GA)

	Original Level of Anti-Semitism in November					
	High		Medium		Low	
Level of Anti-Semitism in the following May	Saw GA	Did Not	Saw GA	Did Not	Saw GA	Did Not
High	63	< 70 +	16	< 45 +	5	< 14 +
Medium	19	> 15 +	46	20	9	< 16 +
Low	18	> 15 +	38	> 35 +	86	> 70 +
Total	100%	100%	100%	100%	100%	100%
N	32	132	26	76	57	173

have added to Glock's table inequality signs to indicate the direction of difference which would seem to confirm the hypothesis of effect, and the + and - signs to indicate the confirmations of expectations, 8 out of 8 in this instance. But if exposure to the movie is itself a symptom of attitude, correlated with both pretest and posttest (if those who go to the movie or report going have lower antisemitism in general), then regression effects alone could give a similar pattern: those initially responding in a non-antisemitic manner who also showed non-antisemitism by reporting going to the film would be more certainly non-antisemitic and would regress less on the second measurement. If the apparent effect were only this simple regression, then one should get a similar picture by reversing the temporal arrangement of the table. Reconstructing the table in raw frequencies, one gets Table 2:

Table 2

Glock's Raw Data (Inferred)

	Level of Anti-Semitism in November					
	High		Medium		Low	
Level in following May	Saw GA	Did Not	Saw GA	Did Not	Saw GA	Did Not
High	20	92	4	34	3	24
Medium	6	20	12	15	5	28
Low	6	20	10	27	49	121

Turning this table on its side, and computing percentages, Table 3.

Table 3

Temporally Inverted Turnover Table

	Level of Anti-Semitism in May					
	High		Medium		Low	
Level in preceding November	Saw GA	Did Not	Saw GA	Did Not	Saw GA	Did Not
High	74	< 61 -	26	< 32 +	9	< 12 +
Medium	15	> 23 -	52	24	15	< 16 +
Low	11	> 16 -	22	> 44 -	75	> 72 +
Total	100%	100%	100%	100%	100%	100%
N	27	150	23	63	65	168

If no regression effects were present, we would expect, from the reversal of the time sequence, all eight signs to now be minus, reflecting the previous all plus pattern. If only the regression effects from selection in terms of pretest scores (and now posttest scores) were present we would have expected all signs to be plus. The plus signs show up in only four of the eight loci. In this case, apparently, there is more than the simple regression effects so far discussed. But lest the vulnerability of the mode of analysis to this regression confusion be doubted, three-variable distributions can easily be generated which show the same perfect plus patterns for both ways of stacking the data. For example, Table 4 shows a frequency distribution, for which the pretest and posttest show identical means and variances. When turned into percentages, either by column or by row (exposure groups kept separate), the same pseudo-confirmation of influence is generated, as shown in Table 5. These effects will be stronger as the correlation between exposure and each of the attitude tests is higher,

Table 4

Hypothetical Data Illustrating Pseudo-Change

	Values on First Measure (Raw Frequencies)					
	High		Medium		Low	
Second Measure	Exposed	Not	Exposed	Not	Exposed	Not
High	30	60	20	30	10	10
Medium	20	30	60	60	30	20
Low	10	10	30	20	60	30

Table 5

Hypothetical Turnover Table Illustrating Pseudo-Change

	Values on First Measure								
	High			Medium			Low		
Second Measure	Exposed		Not	Exposed		Not	Exposed		Not
High	50	<	60 +	18	<	27 +	10	<	17 +
Medium	33	>	30 +	55		55	30	<	33 +
Low	17	>	10 +	27	>	18 +	60	>	50 +
Total	100%		100%	100%		100%	100%		100%

and stronger as the pretest-posttest correlation is lower.

An Alternative Analysis in Terms of Pretest-Posttest Gains

If the turnover table panel-effects analysis is defective because regression is confounded with influence, are there more appropriate modes of analysis which will preserve the panel study's precious difference from one-shot static correlational study and emphasize its greater proximity to true experimentation? Any mode of analysis which enters and classifies in terms of the respondent's scores, or what Underwood [9] calls "subject variables", will be liable to regression effects. But what if we enter the data through exposure to the motion picture rather than through pretest scores? We have argued that going to Gentlemen's Agreement or reporting that one has gone, is in itself a response measure, imperfectly related to the other response measures. "Scores" on attending or non-attending can thus also be expected to regress on other measures. However, if exposure had been ascertained independently of and in between the pretest or posttest interviews, this regression could probably be assumed to be equal in both directions and thus not a possible source of pseudo-effects. While this is not technically so in this study, in that the report of exposure was obtained in the same interview as the retest of attitudes, it could frequently be so in three-wave panel studies, or in studies in which exposure was ascertained independently of the respondent's report of attitudes on the pretest and posttest occasions. Disregarding this qualification for the time being, let us analyse the data as an experiment. Scoring Low, Medium, and High as 0, 1, and 2, the following means result:

	Saw GA	Did Not
Pretest	.783	.892
Posttest	.670	.953
Difference	.113	-.061

Giving each respondent a difference score and comparing the exposed and not-exposed groups, a t ratio of 2.16 is obtained, significant at the .02 level employing a one-tail test. Interpreting this as a quasi-experiment, we might assume that the -.061 change for the "Did not see" groups represents a combination of the main effects of the several extraneous variables called by Campbell [10] "history" (i.e., non-experimental sources of attitude change to which persons in both groups were exposed), "maturation" (the aging of the sample between the two measurements) and "testing" (the effects of repeated measurement upon attitudes). Presumably, the "Saw GA" group would have shown a similar loss had it not been for the influence of the movie. The fact that the two groups were initially different in mean score, and in other unknown but systematic ways, makes it possible that the results are the product of some complex interaction of extraneous variables, but this is not a likely rival hypothesis. The analysis approaches, therefore, the "Non-Equivalent Control-Group Design" (Campbell) [11] frequently found in published social-psychological research, in which two intact natural groups are employed, one being given the experimental treatment and the other being used as the control, but with no random assignment of persons to the two groups to establish true statistical equivalence. (The mode of analysis employed here follows the general line of recommendations made by Thorndike [12] in that the misguided subsampling of pretests to achieve pretest equality has been avoided, and instead gains from the respective starting points have been examined. A covariance analysis would provide a more powerful test.)

However, one somehow judges that the panel-effects design is not quite as controlled an experiment as this makes it out to be. One judges, for example, that it is not as clearly interpretable as evidence of effect as would be a quasi-experiment in which an intact psychology class was shown the movie Gentlemen's Agreement, while an engineering class (probably averaging more antisemitic) was used as a control group. (This hypothetical study would, of course, have less external validity or representativeness (Campbell) [13] but greater internal validity.) Two factors contribute. In the natural movie setting, the "experimental group" is for the most part self-selected, having sought out attendance at the movie. In addition, we have had to take the respondent's say-so as to whether he belongs to the experimental or control group. Stouffer [14] has warned of the errors of interpretation which this can produce, through the correlation of selective memory for exposure with the attitude itself. But if exposure had been ascertained independently, as by a separate intermediate wave of interviewing, such biases could be expected to accentuate equally both the correlations of exposure with pretest and with posttest. It would not in such a case account for a greater apparent pretest-posttest gain on the part of the supposedly exposed groups.

However, in this particular study, the exposure was reported on the same interview as the posttest, some eight months after the pretest. One of the very general findings on questionnaires is that any two correlated items administered on the same day, and especially if a part of the same instrument, correlate more highly than if the occasions are separate. (This effect of temporal and spatial contiguity on correlation is so strong that adjacent items on a multiple-ratings form correlate higher than non-adjacent ones, providing a shift in average correlation level from .46 to .66 in the study of Stockford and Bissell [15].) In addition to the spatial and temporal proximity factors, other aspects of the panel method work to produce a higher correlation between two correlated items occurring in the same interview than between two items from two different interviews. For one thing, each interview is apt to be made by a different interviewer, and therefore the sources of variance associated with interviewer differences augment relationships within one interview and lower relationships between interviews made at different times. And in addition, reinterview studies can scarcely avoid an occasional interviewer mistake or respondent mis-statement in re-identifying former respondents, so that some of the paired pretest-posttest interviews actually come from different persons, and thus seriously attenuate the correlations across interviews.

For these several reasons, then, we would certainly have to expect a higher correlation between the respondent's report of exposure and the posttest than between exposure and pretest, even if reporting seeing the movie were only symptom and had had no causal effect upon attitudes. This is indeed what is found. When these relationships are expressed as product moment correlations (or as point biserial rs) the values are: pretest x exposure -.052, and posttest x exposure -.132. (As biserial rs, the values become -.072 and -.183.) This difference is significant at the .03 level (one-tailed test) when tested by the formula for the significance of the difference between two correlations sharing one array in common (e.g., Peters and Van Voorhis [16], the correlation between pretest and posttest is .561). If this higher posttest correlation were present because of the factors we have suggested, how would this affect the regression on pretest and posttest scores of respondents selected in terms of reported exposure? There would be more regression in the direction of the lower correlation: thus starting with the group showing nonantisemitism by going to G. A., we would expect them to be nonantisemitic on both pretest and posttest, but, because of the lower correlation with pretest, to have regressed further toward the mean on the pretest than on the posttest, and thus when for them pretest and posttest are compared, a pseudo-gain would be shown, in that the posttest mean would be more nonantisemitic than the pretest. There is a further feature that supports this interpretation. If the total group showed no change (and it does not, as shown below), and if the differential-regression explanation were correct, then the non-exposed group should show a complementary loss (or increased antisemitism). This is indeed what is found. Furthermore, if we were to try to transform the "Saw G. A." - "Did not see" dichotomy into points on a continuous scale, the "Saw G. A." group, representing only one fourth of the group, would be more extreme, and hence would show larger regression and differential regression effects. (The "Saw G. A." group as 23% of the population would have a median $\underline{z}$ score value of -1.20, while the 77% "Did not see" would contain members on both sides of the mean, with a median of +.30.) Thus, the larger pretest-posttest change found for the exposed group is in line with the explanation. In particular, the hypothesis of genuine effect would have been stronger had the unexposed group shown no change.

What this analysis attempts to say is this: the difference between the "exposed" and the "not exposed" groups in their attitude shifts between pretest and posttest might have nothing to do with the temporal order of pretest-exposure-posttest at all, and thus is not a legitimate basis for any interpretation of cause and effect. The apparent effect would occur with other temporal orderings. Consider a three-wave study: the first wave measures antisemitism, the second exposure and antisemitism, the third wave antisemitism. The explanatory hypothesis in terms of regression due to differential correlation now alleges that if we treat the third wave as a pretest and the second as posttest, and analyze the "gains" from "pretest" to "posttest", the "exposed" group will show the larger "gains" (or, if true time order be considered, the greater losses). In multiwave panel studies such time-reversing comparison analyses could easily be run.

The differential-correlation-regression hypothesis cannot be decisive, however, for if a genuine effect had occurred, the correlation between exposure and attitude would likewise have increased. In the typical laboratory experiment in attitude change, for example, the correlation moves from zero in the pretest to some positive value, as the experimental group draws away from the control group. The interpretation of the correlation gain (and with it the significant difference in gain scores and our initial unsymmetrical reverse regression presentation) is thus equivocal. Even a tentative choice between the two interpretations of the empirical finding of correlation shift depends upon a judgment of which of two very plausible expectations is given stronger weight: the plausibility that the motion picture will change attitudes or the plausibility that the higher correlation is caused by joint occurrence of report and posttest on the same interview. But while the hypothesis of motion-picture effect is not ruled out, and may even be judged equally plausible to the rival hypothesis of differential correlation and regression, the data cannot be interpreted as adding confirmation to that already plausible expectation.

Note again that the rival regression explanation is plausible only because of the joint occurrence of report of exposure and posttest on the same interview. Panel-effect studies are not limited to this combination. A three-wave study (pretest, exposure-ascertainment, and posttest) would presumably be immune from it, unless it were to be shown that later waves in general showed higher cross-wave correlations than did earlier waves. Exposure might also be recorded independently of the respondent's report, as through names collected in mass sign-ups for a door-prize drawing at the theaters showing the film. Such panel studies would still not be "true" experiments. The experimenter would still have no power over who was exposed to the stimulus and who not, but even so, the interpretation of effect would lack this plausible rival hypothesis, and something much closer to experimentation would have been achieved through the use of measures repeated in time.

Further Alternative Analyses of Effect

Thus not only the original turnover-table form of analysis, but also the examination of pretest and posttest means for an "exposed" and "not exposed" group have been shown liable to pseudo-change effects resulting from statistical regression. Are there available other analyses of effect which do not involve the selection of respondents in terms of attitude symptoms and which are therefore immune to regression effects? Two relevant, if not definitive, analyses remain.

If some have seen the motion picture, and if some have therefore changed in attitude, this implies that the mean attitude of the group as a whole, the unexposed plus the exposed, has been changed. Of course, we would expect such changes in mean to be very small, diluted as the effect must be by the dead weight of the unaffected. In the present data, the total group pretest mean is .867, the posttest mean, .887, the difference between them being but .020, totally insignificant and in the wrong direction besides. It is, of course, possible that other sources of change covered over a genuine effect of the film, but to make this plausible necessitates hypothesizing a source of increased antisemitism over the eight month period.

In this analysis, we have descended to a very unsatisfactory experimental design, the "One-Group Pretest-Posttest Design", the inadequacies of which are well known [17]. We lack a control group to help rule out the rival hypotheses that the change if found is due to the repeated interviewing of the respondents, or to their increase in age and education, or to extra-experimental sources of change provided by the historical events of the intervening months. For this particular study, the effort to provide such a control group from within the community, in which all could potentially have seen the movie, has been undermined by the necessity of classifying respondents by their responses and thus introducing regression effects that could not be disentangled from exposure effects. A control study seems essential, to be made at the same historical time in a separate but comparable city in which Gentlemen's Agreement was not available. Although such a control city would never be identical in the sense that a laboratory control group and experimental group are identical before the introduction of the experimental variable, it nonetheless could serve to help rule out, or to make very implausible, the rival hypotheses that retesting caused the change, that maturation caused it, or that nationwide historical events caused it. The probability of effects from local extraneous events could be evaluated by a study of the newspaper records for the intervening time.

One other analysis avoiding selection of respondents by their response scores seems worth exploring. Again as in the overall mean gain analysis, we accept the model that the natural situation contains an experimental population exposed to the movie and a control population not exposed, and these cannot be adequately separated. Starting from the model of laboratory experiments, we can look at the variance of test scores of the pooled experimental plus control groups for the pretest and for the posttest. If the experimental variable has been effective then the variance of this pooled population is larger on the posttest. Such variance shifts thus become a symptom of effects appropriate when one has reason to believe that not all of the group has been exposed to the experimental variable but when one cannot adequately separate out the exposed and the non-exposed. If this symptom is lacking, the hypothesis of effect is unlikely. In the Gentlemen's Agreement study the pretest σ is .882, the posttest σ is .902, the t ratio of the difference is .63, for which $p < .27$ by a one-tail test. While in the correct direction, this does not reach an acceptable level of significance in spite of the large number of cases involved. And even if it did, we would need a control study comparable to that suggested for the overall mean gain, to be sure that the effect was not due to retesting, extraneous events, etc.

The Kitt and Gleicher Study

To further explore extant applications of the panel-effects design, the study by Kitt and Gleicher [18] on the effect of party contact upon political interest was subject to the same reanalyses. The original turnover table is as shown in Table 6.

Table 6

Kitt and Gleicher's Turnover Table

	Level of Interest in Politics in August					
	High		Medium		Low	
Level in October	Contacted	Not	Contacted	Not	Contacted	Not
High	67	> 58 +	42	> 25 +	5	> 8 -
Medium	26	< 29 +	44	52	26	> 18 +
Low	7	< 13 +	14	< 23 +	69	< 74 +
Total	100%	100%	100%	100%	100%	100%
N	81	192	69	207	43	158

When the raw frequencies are computed, the table flipped in time, and the per cents computed, the inverted turnover table becomes Table 7.

Table 7

Kitt & Gleicher's Data After Temporal Inversion

	Level of Interest in Politics in October					
	High		Medium		Low	
Level in August	Contacted	Not	Contacted	Not	Contacted	Not
High	64	> 63 +	34	> 29 +	13	> 13 ?
Medium	34	< 30 -	48	56	22	> 25 -
Low	2	< 7 +	18	< 15 -	65	< 62 -
Total	100%	100%	100%	100%	100%	100%

Flipping gives a mixed picture of pluses and minuses, indicating some simple regression effects as well as other effects. When these are avoided by analysis of pretest and posttest means by exposure groups, these results are obtained:

	Pretest	Posttest	Gain in Interest
Contacted	1.197	1.202	.005
Not	1.061	.977	-.084

While the contacted group gains scarcely at all, the not-contacted group loses. The difference in gains produces a t ratio of 1.44, $p < .08$ by one-tail test. But again the interpretation of this is equivocal because the contact information and posttest come from the same interview. The correlations involved are .077 between pretest and contact, .121 between posttest and contact, the t for this difference being 1.27, $p < .10$. (The pretest-posttest correlation is .548.) The pretest and posttest means for the pooled contacted and noncontacted group are 1.096 and 1.035, showing a net loss in interest, quite the opposite effect from that hypothesized. The t ratio for this loss is 2.19, $p < .03$ by a two-tail test. For the increased variability analysis, we find that the pretest σ is .789, the posttest σ is .812, the t for the difference being .94, $p < .17$ by a one-tail test.

It is difficult to find in these mixed results convincing evidence of effect. In addition, some general comments on the strategy of social research seem in order: one should always use the best experimental procedure available at a given cost level. Glock's panel analysis of Gentlemen's Agreement was well worth doing because no one had achieved, or even proposed, a better method of evaluating the effects of a motion picture upon the natural audiences that expose themselves to it. For the political contact problem, however, this is not the case, and true experiments are possible with no greater cost, and have indeed been illustrated by Gosnell's [19] classic study and more recent explorations [20]. As in the case of any persuasive effort involving contacts with specific individuals in a setting in which they are not directly aware of the nature of the contacts being received by others, excellent control over random assignment to conditions can be achieved through a "Posttest-only Control-Group Design" in which the apparent artificiality or reactive effects of experimental arrangements are absent (Campbell [21]). Experimentally-introduced political contacts can be provided by essentially the same sampling procedures and staff as a first wave of interviews, to be followed later by an interview of the sampling-equivalent contacted and not-contacted groups. Or party workers, whose time is naturally so limited that they do not get around to contacting everyone, may be induced to allocate their limited number of possible contacts according to an experimental schedule. Or, retreating to a panel study, the determination of who was contacted may be ascertained independently of the respondents, as by interviews with the party workers.

In the natural setting motion picture case, on the other hand, neither the experimenter nor any other external agency such as the party worker, has the initiative in introducing the contact. Exposure to the experimental variable is thus much more inextricably a symptom of attitude. The short-run decision and the group presentation raise further problems. Probably for this problem there is no better approach than the panel study of the three-wave variety or with exposure to the motion picture ascertained independently of the posttest wave as well as of the pretest.

Some Comments on the More General Problem

While this paper will not attempt to analyze in detail actual or possible regression effects in other types of panel analysis, a few general comments can be made. Regression effects are a serious hazard in all the analyses introducing "qualifiers", and in all studies selecting for examination persons with atypical combinations of qualifiers. Maccoby [22] has called attention effectively to the problem in certain types of panel analysis other than those treated here, using primarily random error considerations to explain the effects. It would however be unfortunate if her presentation led to the belief that the mistaken interpretation is limited to instances of unreliable measures or the use of broad categories of measurement. Regression effects will be present wherever the correlation is less than unity, no matter how reliable or with what degree of refinement the variables in question are measured, or what the underlying sources of correlation, or the lack thereof, may be. The effects here discussed are tautological restatements of the fact of imperfect relationships and their degree. In the initial example of the stature of fathers and sons, errors of measurement are certainly a trivial part of the effect.

The sixteen-fold table [23] may well be immune to regression effects such as have been studied here. In the party preference and Willkie attitude table [24], the correlation between the two variables increased markedly, and causal inferences should be restricted to such cases (although we have no assurance on this point for the other causation indices reported). As we have noted, increased correlation between dependent and independent variables is one symptom of experimental effect, and since both variables are represented on both waves, the plausible rival explanation for the correlation increase found in the panel-impact analyses does not hold. The marginals in this case are also so balanced that differential regression seems ruled out. On the other hand, the effect inferred is tantamount to saying that these two conditions hold: 1) The correlation between the two variables (party preference and Willkie attitude) is increasing; and 2) The test-retest reliability of party preference is higher than that for Willkie attitude. If there are plausible reasons to expect that both of these conditions exist, independent of the hypothesis of mutual effect, than a serious rival hypothesis is at hand. For example, it might be generally found that party preference had the higher reliability, whether or not the correlation with attitudes toward candidates was increasing. It might also be found that a test-retest effect occurred in panel studies leading to higher correlations among all correlated variables in the later waves. Were these conditions found to hold for studies in general, then inferences as to causal effects between the two variables in this specific instance would be gratuitous.

Summary

The evaluation of the panel-impact design may be summarized as follows: Potentially, the employment of measures extended in time makes the panel study come closer to experimental analysis than does the one-wave survey. However, the recurrent problem of statistical regression whenever respondents are grouped for analysis in terms of their own responses renders the turnover-table mode of analysis unacceptable. Moreover, when exposure and posttest come from the same interview, the artifactually-resulting higher correlation of exposure with posttest than with pretest introduces a more subtle regression effect, producing a pseudo-difference in pretest-posttest gain scores for the exposed group and a pseudo-loss for the non-exposed group. For a three-wave panel study, or one in which exposure was ascertained independently of the pretest and posttest attitude measures, the comparison of mean gains for exposed and not exposed groups is judged relatively valid as quasi-experiments go, and well worth doing in the absence of better designs.

NOTES

1. A.S. Kitt and D. B. Gleicher, "Determinants of Voting Behavior", Public Opinion Quarterly, 14 (1950), pp. 393-412.

2. C. Y. Glock, "Some Applications of the Panel Method to the Study of Change", American Society for Testing Materials, Symposium on Measurement of Consumer Wants, Special Technical Publication No. 117, 1951, pp. 46-54. Reprinted in P. F. Lazarsfeld and M. Rosenberg (editors), The Language of Social Research: A Reader in the Methodology of the Social Sciences, Glencoe: The Free Press, 1955.

3. F. Galton, "Typical Laws of Heredity", Proceedings of the Royal Institute of Great Britain, 1879 (Proceedings of 1875-1878), pp. 282-301.

4. F. Galton, "Regression Toward Mediocrity in Hereditary Stature", Journal of the Anthropological Institute of Great Britain and Ireland, 15 (1886), pp. 246-263.

5. C.I. Hovland, A.A. Lumsdaine, and F. D. Sheffield, " 'Regression' in the Analysis of Effects of Films", Appendix D in Experiments

in Mass Communication, Princeton: Princeton University Press, 1949.

6. Q. McNemar, "A Critical Examination of the University of Iowa Studies of Environmental Influences Upon the I. Q.", Psychological Bulletin, 37 (1940), pp. 63-92.

7. R. L. Thorndike, "Regression Fallacies in the Matched Groups Experiment", Psychometrika, 7 (1942), pp. 85-102.

8. C. Y. Glock, op. cit..

9. B. J. Underwood, Psychological Research, New York: Appleton-Century-Crofts, 1957.

10. D. T. Campbell, "Factors Relevant to the Validity of Experiments in Social Settings", Psychological Bulletin, 54 (1957), pp. 297-312.

11. D. T. Campbell, "Quasi-Experimental Designs For Use in Natural Social Settings", in D. T. Campbell, Experimenting, Validating, Knowing: Problems of Method in the Social Sciences, New York: McGraw-Hill, in press.

12. R. L. Thorndike, op. cit..

13. D. T. Campbell, "Factors Relevant to the Validity of Experiments in Social Settings", op. cit..

14. S. A. Stouffer, "Some Observations on Study Design", American Journal of Sociology, 55 (1949-50), pp. 355-361; p. 356.

15. R. E. Stockford and H. W. Bissell, "Factors Involved in Establishing a Merit-Rating Scale", Personnel, 26 (1949), pp. 94-110.

16. C. C. Peters and W. R. Van Voorhis, Statistical Procedures and Their Mathematical Bases, New York: McGraw-Hill, 1940, p. 185.

17. E. g., see D. T. Campbell, "Factors Relevant to the Validity of Experiments in Social Settings", op. cit..

18. A. S. Kitt and D. B. Gleicher, op. cit..

19. H. F. Gosnell, Getting Out the Vote: An Experiment in the Stimulation of Voting, Chicago: University of Chicago Press, 1927.

20. S. J. Eldersveld, "Experimental Propaganda Techniques and Voting Behavior", American Political Science Review, 50 (1956), pp. 154-165. Reprinted in H. Eulau, S. J. Eldersveld, and M. Janowitz (editors), Political Behavior: A Reader in Theory and Research, Glencoe: The Free Press, 1956.

21. D. T. Campbell, "Factors Relevant to the Validity of Experiments in Social Settings", op. cit., pp. 304-309.

22. Eleanor E. Maccoby, "Pitfalls in the Analysis of Panel Data: A Research Note on Some Technical Aspects of Voting", American Journal of Sociology, 61 (1956), pp. 359-362. Also Eleanor E. Maccoby and R. Hyman, "Measurement Problems in Voting Studies", in

E. Burdick and A. J. Brodbeck (editors), American Voting Behavior, Glencoe: The Free Press, 1959.

23. E.g., S. M. Lipset, P. F. Lazarsfeld, A. H. Barton, and J. Linz, "The Psychology of Voting: An Analysis of Political Behavior", in G. Lindzey (editor), Handbook of Social Psychology, Cambridge: Addison-Wesley, 1954, Volume II, Chapter 30; p. 1161.

24. Loc. cit., Table 13.