American Psychologist

Journal of the American Psychological Association
Volume 33 May 1978 Number 5

Fallible Indicators of the Subjective State of the Nation

CHARLES F. TURNER *National Research Council*
National Academy of Sciences
Washington, D.C.

ELISSA KRAUSS *National Jury Project*
Berkeley, California

Fallible Indicators of the Subjective State of the Nation

CHARLES F. TURNER *National Research Council*
National Academy of Sciences
Washington, D.C.

ELISSA KRAUSS *National Jury Project*
Berkeley, California

ABSTRACT: *Over the last decade, mounting interest in the social science community has focused upon attempts to monitor the subjective state of the national population. It has been expected that the development of "subjective social indicators" would facilitate social policymaking and forecasting. However, analysis of nine parallel series on public confidence in national institutions suggests that these indicators may be considerably more fallible than past usage would suggest. A number of potential explanations for this fallibility are considered and rejected; it is suggested that the fallibility of these indicators may arise from the varying content of the surveys used to derive indicator estimates. The implications of this fallibility are discussed.*

Social indicators are a class of statistics designed to measure societal conditions. Common social indicators include statistics on objective phenomena such as fertility, crime, social mobility, and life expectancy, as well as attempts to measure subjective phenomena such as the perceived "quality" of national life, tolerance of nonconformity, and public confidence in national institutions. In the last decade, rapidly mounting interest in the social science and policymaking communities has focused upon the role such indicators might play in monitoring changes in the state of the nation (e.g., Duncan, 1974; Sheldon & Parke, 1975; U.S. Department of Health, Education and Welfare, 1969). Apparent parallels between these statistics and the more familiar economic indicators have led to an expectation that the development of reliable social indicator time-series (i.e., a sequence of measurements of the same indicator) would facilitate social reporting and social forecasting. These interests and expectations are reflected in (a) wide distribution by the Office of Management and Budget (OMB) of the volume *Social Indicators, 1973* (Executive Office of the President, 1973) and subsequent efforts to evaluate how federal executives use such data in formulating social policy; (b) establishment by the Social Science Research Council of a Washington-based Center for the Coordination of Research on Social Indicators; (c) programmatic funding by the National Science Foundation for a wide variety of social indicator projects including annual national surveys that generate time-series of subjective social indicators by repeating attitude questions contained in earlier national surveys; (d) the growing use of such data in academic and government research; and (e) the development of new methodologies that employ such data to study how social forces and government policies alter the social state of the nation (see Campbell, 1976; Land & Felson, 1976; Land & Spilerman, 1975).

We wish to acknowledge the generous assistance and advice of Andrew Beveridge, O. D. Duncan, Edwin Goldfield, Norma Graham, Sara Kiesler, and William Madow, and the helpful comments of Donald Buzzelli, Angus Campbell, Richard Christie, Philip Converse, James Davis, Elizabeth Fischer, Norval Glenn, David Goslin, Morris Hansen, Denis Johnston, Kenneth Land, Robert Parke, and Lawrence Santi. We have also benefited from discussions with the Social Science Research Council's Advisory and Planning Committee on Social Indicators and from a long memorandum prepared by Tom Smith of the National Opinion Research Center (NORC) in response to an earlier draft of this article (March 31, 1977).

Special thanks are due the staff of the General Social Survey at NORC, the Louis Harris Data Archive at the University of North Carolina, and Michael Kulikowski at the Harris organization in New York for their cooperative sharing of materials relevant to these issues and to Benita Anderson for her research assistance.

The views expressed in this article are the sole responsibility of the authors and do not necessarily reflect the position of any organizations with which the authors are affiliated.

Requests for reprints should be sent to Charles F. Turner, National Academy of Sciences, 2101 Constitution Avenue, Washington, D.C. 20418.

0003-066X/78/3305-0456$00.75

Although the above examples are mainly American, the "social indicators movement" is not confined to the United States. In the last decade, England, France, Germany, Japan, Canada, Sweden, and Norway have all initiated active social indicator programs, and the Organization for Economic Cooperation and Development has established a "Social Indicators Development Program." [1]

Given the growing use of social indicator data for theory construction and policymaking, it is of crucial importance that the fallibility of these indicators be accurately assessed. Since previous methodological investigations have generated a substantial literature on the error potential of objective social indicators,[2] we have focused our attention upon the subjective domain. In the present article we assess the fallibility of a set of subjective social indicators designed to measure public confidence in the leadership of cultural, political, and commercial institutions. We take as evidence of fallibility disagreement in the level and behavior (across time) of indicators derived from different sources.

The confidence indicators we have studied were derived from surveys conducted between 1973 and 1977 by the National Opinion Research Center (NORC) and by Louis Harris and Associates. Time-series of these indicators have been frequently used in the popular media to document phenomena such as the post-Watergate "crisis in confidence." Use of these social indicators has not, however, been restricted to the popular media; these statistics on public confidence have also been incorporated in recent reports of the National Science Board (1975, 1977) and in the new statistical compendium *Social Indicators, 1976* (Executive Office of the President, 1977). In addition, scholarly research in the social sciences has begun to use these and similar indicators to study change across time in the subjective state of the national population (e.g., Converse, 1972; Davis, 1975a; Etzioni & Nunn, 1975; Mason, Czaijka, & Arber, 1976). This research uses time-series constructed by combining data from two or more surveys and commonly treats fluctuations of 5% to 10% as indicative of trends in the state of the nation. Such interpretations are consistent with standards for sampling error recommended by the survey organizations themselves.[3]

The Level of Public Confidence in National Institutions

Table 1 presents parallel time-series of social indicators representing the level of public confidence in each of nine national institutions. For each institution we present two series, one estimated by NORC and a parallel series estimated by Harris. Except for two instances noted in the legend, each pair of entries in Table 1 reflects estimates of the public's evaluation of institutions described in the same manner by both surveys. We have excluded from comparison all series in which the wording of

---

[1] Discussions of work being done in other countries can be found in Romsoy (1974) and Zapf (1974); examples of this work include the following:

French Republic: Institut National de la Statistique et des Études Économiques. *Données sociales.* Paris: Imprimerie National, 1973.

German Federal Republic: Bundesministerium für Arbeit und Sozialordnung. *Gesellshaftliche Daten, 1973.* Bonn: Presse- und Informationsamt der Bundesregierung, 1973.

Great Britain: Central Statistical Office. *Social trends.* London: Her Majesty's Stationery Office, 1970–.

Japan: Economic Planning Agency. *White paper on national life—1973: The life and its quality in Japan.* Tokyo: Overseas Data Services, 1973.

Norway: Statistisk Sentralbyra. *Sosialt Utsyn.* Oslo: Central Bureau of Statistics, 1974.

Sweden: Statistiska Centralbyra. *Social ultveckling.* Stockholm: 1974.

Trinidad and Tobago: Central Statistical Office. *Social indicators.* Port of Spain: Central Statistical Office, 1975.

Organization for Economic Cooperation and Development. *List of social concerns common to most OECD countries.* Paris: OECD, 1973.

Organization for Economic Cooperation and Development. *Subjective elements of well-being.* Paris: OECD, 1974.

Organization for Economic Cooperation and Development. *Measuring social well-being: A progress report on the development of social indicators.* Paris: OECD, 1977.

United Nations. *Social indicators for housing and urban development.* New York: United Nations, 1973.

United Nations. *Toward a system of social and demographic statistics.* New York: United Nations, 1975.

[2] It should be realized that indicators of putatively objective phenomena, such as age or income, can have substantial subjective components introduced in the process of measurement (e.g., when measured by survey: random and nonrandom errors of recall and response, interviewer effects, etc.). Examples of work being done on the error potential of such indicators include the publications of the U.S. Bureau of the Census' evaluation and research program (e.g., U.S. Bureau of the Census, 1970) and the recent review of national crime statistics (National Academy of Sciences, 1976).

[3] Harris suggests that when two results from their surveys (base $Ns = 1{,}500$ respondents) differ by more than 4%, the difference should be considered "significant," that is, 95 times out of 100 it will reflect true differences in the measured population characteristic rather than random sampling fluctuations. NORC, on the other hand, observes that while mean squared error estimates cannot be directly computed for their surveys, "past experience would suggest that for most purposes (a NORC) sample of 1500 could be considered as having the same sampling efficiency as a simple random sample of 1000" (National Opinion Research Center, 1976, p. 93). Using this standard, differences of 4.5% would be "significant" from a sampling perspective.

institutional descriptions varied across surveys (e.g., we do not compare the Harris series on "Banks" with the NORC series on "Banks and Financial Institutions").[4]

Each entry in Table 1 represents the responses of 1,540 (±60) American adults. The question they were asked derives from an item originally used in 1966 (cf. *Current Opinion,* 1972), which read, "As far as the people running [institution] are concerned, would you say you have a great deal of confidence, only some confidence, or hardly any confidence at all in them?"

Table 1 contains 45 instances in which confidence indicators for the same institution are available from both Harris and NORC. In 27 instances, the two estimates of public confidence differ by 5% or more, and in 10 of 45 instances the two estimates differ by 10% or more.

A cursory examination of Table 1 indicates that the distribution of these discrepancies does not conform to any simple pattern. In 13 of 45 instances, the Harris series show higher levels of public confidence, while the NORC series give higher confidence levels in 27 instances. Furthermore, the apparent stability of public confidence over time varies between series. So, for example, confidence in the Supreme Court (see Figure 1a) appears to be quite stable if we look at the NORC series (year-to-year changes: +2%, −2%, +4%, +0%), while the Harris series shows considerable volatility (+7%, −12%, −6%, +7%). We also find that the direction of year-to-year trends is frequently reversed between the two series. For example, indicators of public confidence in organized religion (see Figure 1b) show consistently divergent trends between 1973 and 1976 (year-to-year changes: Harris vs. NORC = −4% vs. +9%; 0% vs. −20%; −8% vs. +6%).

Clearly, the discrepancies apparent in Table 1 are of sufficient magnitude to warrant concern. The growing use of these and similar subjective social indicators in public and scientific forums suggests that the issues raised by these discrepancies are of sufficient importance to merit detailed investigation.

Some Preliminary Hypotheses

We first noted these discrepancies in October of 1976. Although we had to abandon the idea of using these data for the substantive purposes originally intended, we have been able over the intervening months to consider a number of potential interpretations for the aberrant behavior of these indicators. Our initial hypotheses centered around the possible effects of short-term variations in public attitudes and the impact of survey methods that do not draw simple random samples from the national population. Below are reported in summary fashion our conclusions regarding these explanations.

SEASONAL VARIATION

What is the possibility that seasonal variation (i.e., changes *within* a single year) might account for the discrepancies we have observed?

Since the 1973 and 1974 data were collected 6 months apart in time, it is possible that these data are reliable indicators of rapidly changing public attitudes. This is a particularly plausible possibility for 1974. In that year, the NORC measurements were made during the last portion of the Nixon presidency, while the Harris survey was done *after* Nixon's resignation from office. This variation in timing might cause the confidence ratings for several institutions to differ substantially between the two series.

Since both the NORC and Harris surveys for 1975, 1976, and 1977 were conducted during the period from February 1st to April 30th, it is pos-

---

[4] Several surveys conducted for private clients of the Harris organization also used the confidence question or variants of it. We have chosen to restrict our analysis to the regularly published Harris series on public confidence because (a) unpublished surveys do not always sample from the same universe of the population nor use identical descriptions of all institutions; (b) data from other surveys are available in "raw form" and, unlike the published data, they do not reflect the weighting procedures used by Harris to generate projections to the national population; and (c) the published Harris series appears to be more comparable to the NORC series in other respects; for example, from 1975 on, the Harris estimates are based on surveys whose average cluster size was 8, while other Harris surveys commonly had cluster sizes of 16 (NORC clusters were of size 5).

It should also be noted that the data presented in Table 1 dichotomize the response distribution into "great deal of confidence" versus the other two possible responses ("some" and "hardly any") and all nonresponses (e.g., "don't know," "unsure," no answer, refusal, etc.). This dichotomization is consistent with published sources and past usage of these data. We have investigated the possibility that differences in nonresponse rates might account for the observed discrepancies. Examining unweighted data for 1973 through 1976, we found that the two organizations had highly similar incidences of nonresponse and that repercentaging the data after eliminating the nonresponses did not improve—even marginally—the correspondence between the two sets of estimates.

TABLE 1: *Indicators of Public Confidence—Discrepancies (d) Between Estimates of Percentages of Population Having a "Great Deal of Confidence" in Leaders of National Institutions*

Institution	Source	1973 (%)	d	1974 (%)	d	1975 (%)	d	1976 (%)	d	1977 (%)	d
Major companies	Harris	29	0	21	−10	19	0	16	−6	20	−7
	NORC	29		31		19		22		27	
Organized religion	Harris	36	+1	32	−12	32	+8	24	−6	29	−11
	NORC	35		44		24		30		40	
Executive branch of the federal Government	Harris	19	−10	28	+14	13	0	11	−2	23	−5
	NORC	29		14		13		13		28	
Organized labor	Harris	20	+5	18[b]	0	14	+4	10	−2	14	−1
	NORC	15		18		10		12		15	
Press	Harris	30	+7	25	−1	26	+2	20	−8	18	−7
	NORC	23		26		24		28		25	
Medicine	Harris	57	+3	50	−10	43	−7	42	−12	43	−8
	NORC	54		60		50		54		51	
U.S. Supreme Court	Harris	33	+2	40	+7	28	−3	22	−13	29	−6
	NORC	31		33		31		35		35	
Congress	Harris	29[a]	+6	18	+1	13	0	9	−5	17	−2
	NORC	23		17		13		14		19	
Military	Harris	40	+8	33	−7	24	−11	23	−16	27	−9
	NORC	32		40		35		39		36	
Date of survey	Harris	9/73		9/74		4/75		2–3/76		2/77	
	NORC	3/73		3/74		3–4/75		3–4/76		2–3/77	
Sample size	Harris	1,594		1,527		1,579		1,512		1,522	
	NORC	1,504		1,484		1,490		1,499		1,530	

Note. Data are from the National Opinion Research Center (NORC; 1973, 1974, 1975, 1976, 1977) and Harris and Associates (1973, 1974, 1975b, 1976, 1977).

Marginals published in the NORC codebooks are unweighted; we have retabulated the raw NORC data for 1973–1976, weighting the sample in proportion to the number of "eligible" individuals in the respondents' households. Indicator estimates derived from this reanalysis were virtually identical (i.e., ±1%) to the raw marginals found in NORC's publications; use of the weighted estimates does not decrease the mean value of d.

[a] Published Harris series refers to "Congress," but Harris questionnaire in 1973 described U.S. House of Representatives and U.S. Senate. This entry appears to be an average of the responses for Senate (30%) and House (29%).

[b] Published Harris series refers to "Organized Labor," but Harris questionnaire in 1975 described "Labor Unions."

sible to test the adequacy of seasonal variation as a general explanation for the discrepancies found in Table 1. Considering only the 1975–1977 measurements, we find that, on the average, discrepancies during this period are marginally greater than those in the preceding years. In particular, we find that the Harris and NORC estimates during these years differ by 5% or more in 17 (of 27) instances, and in five instances the discrepancies exceed 10%. Finally, we note that the largest discrepancy in any year (16%) occurred in 1976 when both surveys were conducted during the 8-week period from March 1st to April 30th.

Although the possibility of large *week-to-week* fluctuations cannot be completely ruled out, the foregoing results suggest that seasonal variation is an unlikely explanation for the discrepancies we have observed.

DEMOGRAPHIC REPRESENTATIVENESS

The research literature suggests that when substantial differences in attitudes are found between two surveys, a likely cause is variation in coverage of different age, education, racial, and religious groups in the population. To test this possibility, we compared the available demographic information for the two surveys. Table 2 presents these data for the 1976 surveys along with information derived from Census projections for the entire population.

This comparison indicates that *both* samples de-

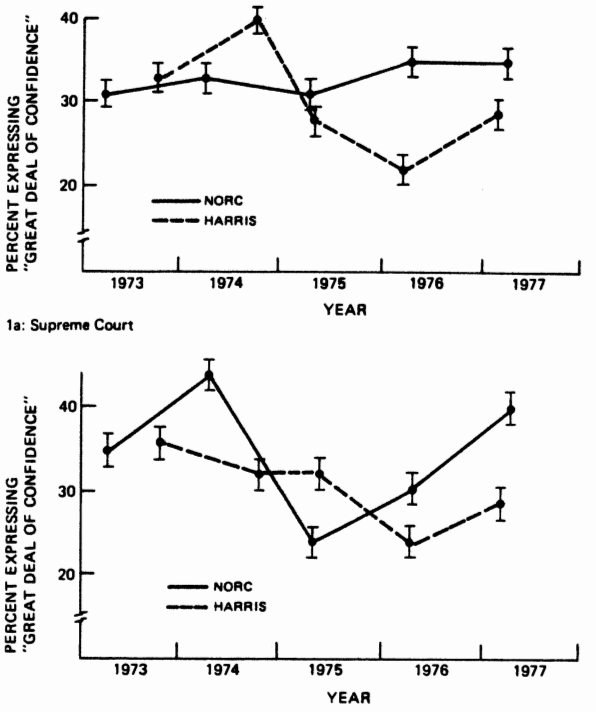

Figure 1. Parallel estimates of the percentage of the American adult population having a "great deal of confidence" in the leadership of (a) the U.S. Supreme Court, and (b) organized religion. Error bars around measurement points demark ± 1 standard error. (As a graphical convention, connecting lines have been drawn between measurements in each series; it should not be assumed, however, that the rate of change in public confidence was necessarily uniform between measurements.)

viate from the national population in having somewhat fewer young and less-educated people and considerably more self-described "voters." The comparison of these samples with U.S. Census data also reveals an apparent underrepresentation of the population earning $15,000 or more; this, however, probably arises from high levels of nonreporting of income by wealthier families in the Harris and NORC samples.[5]

While demographic representativeness is important in deciding whether our social indicator estimates reflect the condition of the national population, deviations from the Census, per se, do not necessarily mean that the two surveys are not comparable. The two sample surveys themselves show great similarity in their coverage of persons of different sex, age, race, religion, voting, and financial status. Education is the only demographic variable whose distribution varies more than 5% between surveys. Harris oversamples the college-educated population by 9%, while NORC's sample comes within 2% of U.S. Bureau of the Census estimates. While this deviation could not possibly eliminate all of the discrepancies present in Table 1, it could reduce these deviations—perhaps to the point of statistical insignificance.

To test this possibility, we weighted the NORC data for 1976 to produce a synthetic sample whose educational distribution would match that of the 1976 Harris sample. Contrary to our expectations, this procedure did not reduce the magnitude of the discrepancies, but rather it *increased* the differences between the Harris and NORC estimates by an average of 1%.

RANDOM SAMPLING ERROR AND THE EFFECTS OF CLUSTERING

Both the Harris organization and NORC drew samples alleged to be representative of all English-speaking, noninstitutionalized residents of the continental United States aged 18 and over. Selection of individual respondents proceeded through two phases. A first series of procedures selected specific geographic locations; [6] blocks or segments of the country were selected with the probabilities of selection being proportional to the area's population.

Within each block or segment, clusters of five (NORC) or eight (Harris) individuals were interviewed. If we assume that the availability and

---

[5] In the NORC and Harris surveys, 8.6% and 7.0% of respondents, respectively, did not report their incomes. U.S. Bureau of the Census (1976a) figures employ income "estimates" whenever people fail to respond to the income question.

[6] Technically, the sampling employed by both organizations used a process known as "multi-stage area probability" sampling. This procedure involved successive selections; first, specific metropolitan areas or counties were selected, and then *within* those metropolitan areas or countries, specific geographic tracts were selected. Finally, from these tracts, specific "blocks" were selected for interviewing. To increase sampling efficiency, stratified random selection procedures were used at each stage; (NORC) stratification was done by region, age, and race in the selection of the tracts, and by race and income during selection of blocks.

We have treated the sample as a simple random sample to the block level; we assume no clustering effects exist at higher levels. Intraclass correlations were computed for responses to the confidence question by region of the country, race, age, sex, and employment status. These correlations were all quite low (median = .01; maximum = .02), and this gives us some confidence in the accuracy of approximations which assume simple random sampling of "blocks."

TABLE 2: *Demographic and Social Characteristics of the National Population and the Harris and NORC Samples*

Characteristic	1976 population aged 18+ (%)	1976 Harris (%)	1976 NORC (%)	Deviation (Harris minus NORC) (%)
Age				
18–29	30.7	28.6	25.9	+2.7
30–49	32.7	33.1	33.2	−.1
50+	36.7	36.3	40.5	−4.2
Education				
0–8 years	17.5	12.4	17.3	−4.9
9–12 years	54.0	49.8	52.1	−2.3
13+ years	28.5	37.2	30.1	+7.1
Sex				
Female	51.2	53.0	55.4	−2.4
Male	48.8	47.0	44.6	+2.4
Race				
Black	11.4	10.0	8.6	+1.4
White	87.1	85.1	90.8	−5.7
Other	1.5	4.9	.4	+5.2
Income[a]				
Less than $5,000	12.0	16.0	17.6	−1.6
$5,000–$9,999	21.1	22.1	22.5	−.4
$10,000–$14,999	22.3	20.6	20.7	−.1
$15,000+	44.4	32.7	32.1	+.6
Missing, refused, etc.	—	8.6	7.0	—
Voting (in 1972)				
Didn't vote	44.5	35.9	38.7	−2.8
Voted for Nixon	33.7	39.0	36.3	+2.7
Voted for McGovern	20.8	22.6	22.9	−.3
Voted for other	1.0	2.5	2.1	+.4
Religion (for whites)[b]				
Protestant	—	60.8	61.7	−.9
Catholic	—	25.3	27.2	−1.9
Jewish	—	3.8	2.0	+1.8

Note. Harris and National Opinion Research Center (NORC) estimates are derived from special analyses of original data. Figures for the national population are taken from the following sources: age, sex, and race—derived from projections for July 1976 (U.S. Bureau of the Census, 1977b, Table 1); education—U.S. Bureau of the Census (1977a); income—U.S. Bureau of the Census (1976a); voting—U.S. Bureau of the Census (1976b) and U.S. Congress, Clerk of the House (1972).

All U.S. Bureau of the Census percentages have been recomputed to reflect only the population aged 18 and over. It should be remembered that the Census data include persons in Hawaii and Alaska (.5% of national population) and institutionalized persons who would not be included in the sampling universe of the Harris and NORC surveys.

Survey data sum to less than 100% owing to the failure of individuals to respond to certain questions; except for income, nonresponse on these questions did not exceed 2% of the total sample.

[a] Income entries are derived from measurements made in 1976 of total family income in the preceding year.

[b] The available Harris analyses do not permit a breakdown of the religious preferences of nonwhite respondents or of those affiliated with other religions or no religion.

selection of respondents introduces inconsequential biases (see below) we can begin to assess the effect of clustering upon the sampling error. In simple random samples of N individuals drawn from a large population, 95 times out of 100 the proportion *of the whole population* which would express a "great deal of confidence" in a given institution would lie in the interval [7]

$$E \pm 1.96\sqrt{\frac{.25}{N}} \tag{1}$$

around our sample estimates (E). Furthermore, when we draw two independent samples of sizes N_1 and N_2 we expect that 95 times out of 100 the difference (d) between the proportions of the samples expressing a "great deal of confidence" would lie within the 95% confidence interval (C_{95}) for the difference between two proportions. That is,

$$|d| \leq C_{95} \tag{2}$$

where

$$C_{95} = 1.96\sqrt{\frac{.25}{N_1} + \frac{.25}{N_2}} \quad . \tag{3}$$

Using Equation 3, we can set limits upon the possible values of the 95% confidence interval for the comparisons we have made in Table 1. An upper-bound estimate is provided by assuming that the sample size equals the total number of units that were chosen *at random*, that is, the number of blocks or segments (NORC ≈ 300; Harris ≈ 187). A lower-bound estimate is provided by assuming that cluster samples have the same efficiency as simple random samples.

Computing C_{95} under these two boundary assumptions, we derive the following limits,

$$3.6\% \leq C_{95} \leq 9.1\%. \tag{4}$$

Given either limit, it is quite clear that the differences shown in Table 1 are unlikely to have arisen through sampling error alone. Indirect procedures (cf. Blalock, 1972, chap. 21; Kish, 1965, chap. 6) using the level of intracluster response correlation as an index of sampling efficiency provide more exact estimates of the magnitude of sampling errors in cluster designs. Applying these procedures, we

[7] To simplify our exposition we use .25 in all formulae. This is a conservative estimate for the term $p(1-p)$, where p would be the proportion expressing a "great deal of confidence." When the proportion in question gets extremely small or large (less than .10 or more than .90), the sampling errors estimated by our formulae will be overly conservative, and the estimate for C_{95} will be somewhat larger than required by theory.

derived an estimate of 4.4% as an approximation [8] for C_{95}. Given this result, we would expect random sampling fluctuations to produce a *single* 10% difference less frequently than 1 in 10,000 times.

NONRANDOM SELECTION OF INDIVIDUAL RESPONDENTS

Since the sampling of individuals within geographic locations was not random, it is theoretically possible for substantial biases to be introduced by the availability and selection of persons to be interviewed. Field workers, for example, may choose to interview particular categories of people because experience suggests that these people are less hostile to surveys. Differences in the nature of these biases may cause the indicator estimates derived from two surveys to vary substantially. Fortunately, it is possible to make some assessment of the likely magnitude of these effects.

In 1975 and 1976, NORC used *both* random sampling and quota sampling at the block level. In the randomly sampled condition, all households in a specified block or segment were enumerated, and specific individuals within the block were preselected *at random* for interviewing. Field workers called at the specified household until an interview was completed or the request for an interview was denied. In 1976 this produced 744 completed interviews from the 991 eligible individuals originally selected for interviewing (i.e., 75% completion rate). The remaining half of the survey was done by the traditional and less costly process of "quota sampling" at the block level. Interviewers were assigned one household as a starting point and a fixed travel pattern; they were instructed to complete interviews with *available* adults so as to fill "interview quotas." These quotas for sex, employment status, and age insured that the block samples contained adequate numbers of the "hard-to-reach" sectors of the population.

To assess the effect of nonrandom selection of respondents, we compared the distribution of responses obtained using these two procedures. We found no significant differences between the estimates of public confidence provided by the quota samples and random samples drawn by NORC. Of course, it is not possible to conclude from these results that quota sampling is *always* equivalent to random sampling. Nonetheless, what evidence there is suggests that divergences of the magnitude shown in Table 1 are not, in practice, attributable to the effect of nonrandom sampling at the block level.

Alternative Approaches to Indicator Fallibility

The foregoing analyses show our preliminary hypotheses to be inadequate explanations for the aberrant behavior of these subjective social indicators. This result has led us to consider other ways of treating these data.

THE RELIABILITY OF RANK ORDERS

There are several possible "fall-back" positions from which we might seek to defend the reliability of these confidence indicators. One of the most attractive involves treating these data as if they were designed only to provide a rank ordering of institutions. So, for example, we might say that the data in Table 1 showed that the public had more confidence in the Supreme Court than it did in Congress, but we would refrain from interpreting the actual percentages.

If the rank orders of institutions were reliable, we might be able to use these data to monitor changes in the standing of specific institutions *relative* to other institutions, although we would have to abandon the idea of measuring changes in the actual level of public confidence in any institution. Initially, this position seems supportable. Computing the correlation between Harris and NORC for each year, we find that the two series are quite consistent in their ranking of the nine institutions. The median rank correlation is .82.

While this result appears encouraging, it is actually somewhat misleading. The presence of a

---

[8] Estimation was done using the 1976 NORC survey; the sampling code incorporated in publicly available data from that survey permits the identification of sets of three contiguous clusters (mean size = 15). Averaging responses to all confidence items, we obtained an estimated intraclass correlation across these cluster sets of .039. Allowing for our cluster sets being an average size of 15, this level of correlation would raise the estimated sampling error from .013 to .016, and it would inflate the value of C_{95} from .036 to .044.

Given the assumptions made here and those noted previously, these figures should be treated as rough approximations. We do note, however, that our approximation agrees quite closely with the NORC suggestion that their sample might be considered to have the same efficiency as a simple random sample of 1,000 respondents.

given level of reliability, unless perfect, tells us nothing about the suitability of an indicator for detecting *change*. The central question in this regard is the reliability of our indicators relative to the magnitude of the changes that are occurring. To address this question, Table 3 compares correlations between rank orders obtained for measurements done in the same year by different survey organizations ("houses"), in different years by the same organization, and in different years by different organizations. From this analysis we see that there is as much variability in the ranking of institutions by different surveys in the *same* year as there is between surveys conducted up to 4 years apart. More concretely, these results reveal that prediction of the rank orderings which were found by Harris in a given year could be done *equally well* by knowing what Harris or NORC found in another year versus knowing what NORC found in the *same* year.

TABLE 3: *Median Correlations Between Confidence Rankings of Institutions, 1973–1977*

Comparisons involving	Number of rank correlations	Median correlation
Different survey houses in same year	5	.82
Same survey house in different years	20	.86
Different houses in different years	20	.90

Note. Measure of correlation is Spearman's rho.

SYSTEMATIC MODELS AND UNSYSTEMATIC ERRORS

Because the data stubbornly resisted our "simple" explanations and because fall-back positions such as analysis of the rank orders were untenable, we have attempted a more rigorous analysis of the dilemma posed by these discrepant indicators.

The indicator data presented in Table 1 have a multivariate structure quite similar to that of many psychological experiments. We have a simple behavior—whether or not the respondent said she or he had a "great deal of confidence"—and we have two independent variables—the year in which the measurement was taken (Y) and the survey "house" (H) that did the measuring. All respondents were run in nine "conditions," which required them to respond to nine different "stimuli," that is the nine institutions (I).

Viewing the data from this perspective, we might consider using standard analysis-of-variance models to disentangle the effects of year-to-year changes from survey-house biases. Technical problems preclude such analyses.[9] However, recent advances in the analysis of multivariate contingency tables (cf. Goodman, 1972; Thiel, 1970) provide a comparable analytical framework that is technically appropriate for these data. Given the purpose of the present article, we do not present a detailed exposition of these methods; excellent discussions of these techniques are available elsewhere (e.g., Bishop, Fienberg, & Holland, 1974; Davis, 1975b).

Using these procedures, we have attempted to identify a minimally complex model for the response patterns summarized in Table 1. (For simplicity, we treat data for each institution separately.) A hierarchy of four models can be formulated; these models incorporate progressively more complex effects of the two independent variables (year and house). The reasoning behind this set of models is analogous to that which motivates hierarchical analysis-of-variance procedures.

Model 1: *No change across time and consistent indicators.* This model posits that within the limits of sampling error, public confidence was *constant* across time and equivalent between the Harris and NORC series. (No main effects or interaction)

Model 2: *Systematic change across time measured by consistent indicators.* This is the model we would prefer to accept. It posits that the year-to-year changes in public confidence are too large to be attributable to sampling error, and that the two indicator series show no significant inconsistencies. (Main effect for year only)

Model 3: *Systematic change across time measured by indicators that have a constant bias.* This model also posits significant changes across time, but it allows the indicators to have a significant but constant bias.[10]

[9] The most important of which is the fact that the dependent variable (whether or not a person expressed a great deal of confidence) has a discrete rather than a continuous distribution.

[10] Our use of the term *bias* is relative. Given the nature of these data, it is not possible to say that one indicator estimate is biased while the other is not. In particular, it is possible for both indicators to be biased estimates of the population characteristic in question but for them to yield

While acceptance of this model compromises simple across-house comparisons of the *level* of public confidence, it would indicate that year-to-year *trends* in the two series were consistent in magnitude and direction. (Significant main effects for both year and house, but no interaction)

Model 4: *Significant and inconsistent biases in indicators.* If we reject Model 3, we are forced to explain the pattern of change reflected in these indicators by positing significant biases that vary from year to year and house to house. Acceptance of this model compromises all comparisons. (Significant Year × House interactions)

We have fitted this hierarchy of models to the indicator series for each of the nine institutions. To allow for the diminished efficiency of cluster samples, we assumed that each measurement represented the responses of 1,000 (rather than 1,500) individuals, and we performed each analysis separately for the entire series (1973–1977) and for the three years (1975–1977) in which the measurements were made at the same time of year.

Table 4 shows the goodness of fit of our models to the time-series for each institution. The measure of fit in this analysis is the chi-square statistic for the difference between the response distribution predicted by the model and that observed in the time-series. Not surprisingly, our results indicate that the "no change" model (Model 1) must be rejected in every instance. However, we also must consistently reject the preferred model (Model 2), which considers the Harris and NORC series to be consistent indicators of the same population characteristic. In every instance but two, we also had to reject Model 3, which posits the two series to be consistent indicators of year-to-year *trends* although inconsistent in their estimates of the absolute level of public confidence. Thus, in 16 of 18 cases we were forced to accept Model 4, which postulates significant "house" biases whose magnitude *varies* from year to year.

identical estimates; this will happen if the biases are of equal magnitude. In the present case we are referring only to the existence of *nonequivalent* biases between measurements made by different organizations (or the same organization at different times). *Constant bias* refers to a condition in which the difference between any such biases is nonzero and constant across time.

An Interpretation

What is the nature of the biases that cause the large and inconsistent discrepancies in Table 1? The foregoing analysis indicates that these biases vary across years as well as across survey houses, which suggests that we should seek our explanations within *specific* surveys rather than across all surveys conducted by one organization. Some relevant patterns emerge upon a careful inspection of the annual estimates presented in Table 1. For example, while we observe no *general* trend for Harris surveys to find lower levels of public confidence than NORC, we do note that in both 1976 and 1977, Harris confidence estimates are lower than the NORC estimates *for every institution.*

This pattern suggests a possible interpretation. All of the "confidence indicators" represent responses to similar questions embedded in dissimilar surveys; the other questions in these surveys varied *both* from house to house and from year to year. If this changing context altered individuals' "response sets" or their interpretations of the meaning of the confidence question, then we might expect estimates of public confidence to show patterns of bias such as we have found in our analyses.

CONTEXT VARIATIONS [11]

The effects of context upon an individual's interpretation and response to stimuli have been widely

[11] Examination of the questionnaires also reveals that the Harris series incorporates a slight alteration in the wording of the preface to the question. After the initial use of this question, Harris, for unknown reasons, changed the preface to read "As far as the people in charge of running—" rather than "As far as the people running—." The Harris organization treats data from both versions as a unitary time-series without noting this variation in their publications.

To test the potential effects of such minor alterations in phrasing upon responses to the confidence item, we compared responses to the second version with those to another preface which read "How much confidence do you have in the people running—?" This version was used in a survey conducted by Harris for a private client in 1975 (Harris, 1975a). Since this survey was done between March 21st and April 3rd, it should be comparable to published Harris data for 1975 that derive from another survey conducted between April 11th and April 17th. Four institutions were identically described in these two surveys; responses to the alternate wordings ("How much confidence" and "As far as the people in charge," respectively) were as follows: military, 25%, 24%; press 24%, 26%; major companies, 18%, 19%; Congress, 11%, 13%.

It can be seen from this comparison that no discrepancy was larger than 2%; this is well within the expected limits for sampling fluctuations in estimates of the same population characteristic.

TABLE 4: *Test of Alternative Models for Behavior of Confidence Indicators*

Indicator series / Model	1973–1977 series: Fit of model (χ^2)	Decision	p	1975–1977 series: Fit of model (χ^2)	Decision	p
Major companies						
1. No change, consistent indicators	137.4	reject	<.001	41.7	reject	<.001
2. Reliably indicated change	51.5	reject	<.001	25.4	reject	<.001
3. Annual change + constant bias	21.6	reject	<.001	8.1	reject	<.05
4. Variable indicator biases	—[a]	accept		—[a]	accept	
Organized religion						
1. No change, consistent indicators	168.8	reject	<.001	83.2	reject	<.001
2. Reliably indicated change	82.8	reject	<.001	51.9	reject	<.001
3. Annual change + constant bias	64.4	reject	<.001	45.4	reject	<.001
4. Variable indicator biases	—[a]	accept		—[a]	accept	
Organized labor						
1. No change, consistent indicators	82.9	reject	<.001	21.8	reject	<.001
2. Reliably indicated change	18.7	reject	<.005	10.1	reject	<.01
3. Annual change + constant bias	15.8[b]	reject	<.05	9.9[b]	reject	<.01
4. Variable indicator biases	—[a]	accept		—[a]	accept	
Executive branch of the federal government						
1. No change, consistent indicators	305.0	reject	<.001	163.0	reject	<.001
2. Reliably indicated change	96.1	reject	<.001	8.5	reject	<.05
3. Annual change + constant bias	95.5[b]	reject	<.001	2.5	accept	
4. Variable indicator biases	—[a]	accept		—	—	
Press						
1. No change, consistent indicators	62.0	reject	<.001	40.5	reject	<.001
2. Reliably indicated change	46.1	reject	<.001	33.2	reject	<.001
3. Annual change + constant bias	43.5[b]	reject	<.001	17.5	reject	<.001
4. Variable indicator biases	—[a]	accept		—[a]	accept	
Medicine						
1. No change, consistent indicators	137.5	reject	<.001	52.6	reject	<.001
2. Reliably indicated change	73.7	reject	<.001	51.6[b]	reject	<.001
3. Annual change + constant bias	27.1	reject	<.001	2.8	accept	
4. Variable indicator biases	—[a]	accept		—	—	
U.S. Supreme Court						
1. No change, consistent indicators	98.8	reject	<.001	58.4	reject	<.001
2. Reliably indicated change	63.7	reject	<.001	52.2	reject	<.001
3. Annual change + constant bias	55.9	reject	<.001	13.7	reject	<.001
4. Variable indicator biases	—[a]	accept		—[a]	accept	
Congress						
1. No change, consistent indicators	198.7	reject	<.001	51.0	reject	<.001
2. Reliably indicated change	23.5	reject	<.001	13.7	reject	<.01
3. Annual change + constant bias	23.5[b]	reject	<.001	7.0	reject	<.05
4. Variable indicator biases	—[a]	accept		—[a]	accept	
Military						
1. No change, consistent indicators	168.8	reject	<.001	110.4	reject	<.001
2. Reliably indicated change	132.9	reject	<.001	108.4[b]	reject	<.001
3. Annual change + constant bias	77.1	reject	<.001	6.3	reject	<.05
4. Variable indicator biases	—[a]	accept		—[a]	accept	

Note. A χ^2 value of zero would indicate a perfect fit between the model and our data. Models were fit using procedures developed by Goodman (1971). Variables are C (confidence response: "great deal" vs. other), H (survey house), and Y (year: 5 categories); C and H are dichotomous variables. In Goodman's notation, our models are constrained to fit the following marginals of the response distributions, with the following degrees of freedom for the test of fit:

Model 1: (YH), (C); 1973–1977 series, $df = 9$; 1975–1977 series, $df = 5$
Model 2: (YH), (CY); 1973–1977 series, $df = 5$; 1975–1977 series, $df = 3$
Model 3: (YH), (CY), (CH); 1973–1977 series, $df = 4$; 1975–1977 series, $df = 2$
Model 4: (YHC); 1973–1977 series, $df = 0$; 1975–1977 series, $df = 0$.

[a] Models allowing for "variable indicator biases" provide a perfect "fit," since they are constrained to reproduce the CYH response distribution.

[b] Instances where additional effects incorporated in a model do not significantly ($p < .05$) improve fit over that of the preceding model (which does not include these effects). Given the purpose of the present analysis, we do not include results for models that posit a constant bias but *no* year-to-year change (marginals fit: YH, CH). A model of this type, rather than Model 3, would provide the most parsimonious description of the data in one instance: the 1975–1977 series on medicine.

documented in the perceptual and social psychological literature. For example, well-known studies by Asch (1956) have shown that knowledge of the opinions of other people can severely bias the public judgments of individuals asked to compare the length of two lines. Research on human emotions (Schachter, 1971) has shown that people use cues embedded in their environmental context to interpret and provide behavioral guidance for the autonomic arousal that characterizes emotional states. Similarly, work (*cf.* Kling & Riggs, 1971, chap. 12; Posner, 1973, chap. 7) on the perception of ambiguous forms and the solution of intellectual problems shows that previous experience on similar tasks can bias subjects' responses by decreasing their "openness" to different ways of viewing forms or solving problems. The latter phenomena are referred to as "response sets" or "perceptual sets" that are created by the context in which stimuli are presented.

Are such context effects a plausible interpretation for the fallibility of our subjective social indicators? We have examined the content of the questionnaires used in each of the 10 surveys. Although our impressions must be treated cautiously, there do seem to be instances where one might anticipate substantial context effects to occur.

A particularly notable example occurred in 1976 when the Harris confidence questions immediately followed a series of six questions on "political alienation." This series included items such as "The people in Washington are out of touch with the rest of the country (agree/disagree)." In contrast, the first item in the NORC survey that year was the confidence question. Although our interpretation is admittedly speculative, it seems possible that this variation in context might contribute to our finding that the confidence levels estimated by Harris were *consistently lower* than the NORC estimates in 1976.

Other examples arise because context effects can result from the order of presentation of the individual elements of a multipart question. Thus we note that the confidence question solicited responses to a series of institutions; the order and content of this series varied both between survey organizations and across time. For example, NORC altered its order of presentation for 1975 and 1976. As a result of changes in those years, organized religion was moved to the eighth position in the series of institutions and it was presented as a partial repetition of the question "How about the people running organized religion?" In other years, NORC presented organized religion as the second institution in their series, and its presentation did not *remind* respondents that they were being asked about their confidence in "the *people running* organized religion" (rather than their confidence in the institution itself). When this alteration in context was first made, the NORC series registered a drop from 44% (1974) to 24% (1975); this was the largest annual change recorded for any institution in Table 1. In contrast, the Harris series on organized religion was perfectly stable during this interval. Conversely, when NORC reverted to its original context, the percentage of people responding that they had a "great deal of confidence" rose from 30% (1976) to 40% (1977). It is reasonable to suspect that changes in the context of NORC's measurements contributed to these large fluctuations.

Unfortunately, there is no experimental evidence to provide a true test of these explanations.

PLAUSIBILITY OF CONTEXT INTERPRETATIONS

While a context interpretation has some face validity, it is not clear that such effects would be of sufficient magnitude to explain the large discrepancies (e.g., 16%) that occur in Table 1. The findings of past research on context effects in surveys are not congenial to such an interpretation. A recent review of this research concluded that the "position of a question [in the survey questionnaire] has by itself little biasing effect for behavioral items and a negligible effect for attitudinal items . . . [and] there do not appear to be any sizeable response effects associated with the placement of questions after related questions" (Sudman & Bradburn, 1974, p. 33). While the literature is generally unencouraging, few published studies have used attitudinal items of the sort we are dealing with.

Let us consider some new evidence on this topic. The 1976 NORC survey incorporated a planned variation in question ordering. In that year, a question about the "burden" of federal income taxes,

> Do you consider the amount of federal income tax which you have to pay as too high, about right, or too low?

was asked either immediately *before* or immediately *after* another question which read,

> We are faced with many problems in this country, none of which can be solved easily or inexpensively. I'm going to name some of these problems and for each one I'd like you to tell me whether you think we're spending too much money on it, too little money, or about the right amount.

Evaluations of spending in 11 areas, such as "space exploration," "improving and protecting the environment," and "welfare," were then solicited.

Since survey procedures were otherwise identical, we can use this "experiment" to assess the magnitude of the effects that may occur from such variations in context. Table 5 tabulates responses to the "tax burden" question for each context. We note from this tabulation that there is a 14.2% discrepancy (between contexts) for the estimated percentage of the population that considers the federal tax burden to be "about right." Apparently, being sensitized beforehand to the large number of programs supported by tax revenues causes a marked attenuation in complaints that federal taxes are "too high."

Although generalizations from this finding are necessarily speculative, it does appear that measurements of the subjective state of the population may be subject to large and variable contextual biases. If context effects in the indicator series on public confidence are as large as those found for the tax question, the puzzling discrepancies of Table 1 have found an explanation—albeit an uncomfortable one.

TABLE 5: *Effect of Survey Context on Public Evaluation of the Federal Tax Burden*

Context	Evaluation of "amount of federal taxes": Too high	About right	Too low	Don't know, no answer, etc.	Sample size
Follows question on spending for federal programs	52.6%	39.9%	.6%	7.0%	721
Precedes question on spending for federal programs	63.2%	25.7%	.6%	10.4%	778

Note. Data for this analysis are from the National Opinion Research Center's 1976 General Social Survey.

Conclusions

Our investigation leads us to conclude that commonly used indicators of the subjective state of the national population can be considerably more fallible than past usage would suggest. Although firm conclusions about the cause of the fallibility of the confidence indicators must await methodological experiments built into future national surveys, we have seen one instance in which another subjective indicator fluctuated substantially with the context in which the question was asked. This result suggests that subjective social indicators may be subject to *non*sampling errors that are considerably larger than the sampling error standards ordinarily used as criteria for inferring change in the state of the nation.

This finding has important implications for a wide range of policymaking and research activities that have begun to use these data to monitor changes in the state of the nation. A number of private and government agencies are presently supporting research programs that aim to construct time-series of subjective indicators by repeating attitude questions asked in past surveys. Our results suggest that comparisons of such indicator estimates may involve substantial and unanticipated errors arising from variations in the contexts in which such measurements are made. At a minimum, knowing only that two measurements were made in different surveys at different points in time, we suspect that the interpretation of apparent "trends" of less than 10% may lead to erroneous inferences about changes in the state of the national population. Conclusions from past studies that "all institutions gained support, some quite substantially . . . in some instances by as much as seven percentage points" (analysis of confidence in science: NORC 3/73 vs. Harris 9/73; Etzioni & Nunn, 1974, p. 194) should be treated cautiously. The results of our investigation suggest that a much higher standard—for example, 15%—may be more appropriate for data such as these.

In the policymaking field, similar caution is advisable. The Institute for Social Research at the University of Michigan has recently completed a study of the use of social indicator data by federal executives. This study concluded that there was a need for government statistical compendia, such as the *Social Indicators* volumes, to "go beyond objective indicators and provide subjective measures of life experience and social well-being" (Caplan & Barton, 1976, p. 26). Clearly, the error potential of such indicators must be an important consideration in determining whether and how this need should be met.

While caution may be a prudent interim measure, we will ultimately need to undertake a fundamental reconsideration of the adequacy of our attempts to measure the subjective state of the population. If the "social indicators movement" is to provide reliable time-series paralleling our national economic indicators, then the errors in these indicators must be reduced substantially below the levels we have observed. After all, we would not tolerate it if the

TABLE 6: *Results from a Fully Replicated 1954 Survey on Attitudes Toward Civil Liberties*

Estimate of percentage of population willing to	Survey conducted by		Discrepancy
	AIPO	NORC	
Permit Communist to speak	27%	27%	0%
Permit "atheist" to speak	37%	37%	0%
Permit "socialist" to speak	57%	60%	−3%
Sample size	2,483	2,450	

Note. AIPO = American Institute for Public Opinion (Gallup poll); NORC = National Opinion Research Center. Data are derived from Stouffer (1955). Questions responded to were: 1. "Now, I should like to ask you some questions about a man who admits he is a Communist .... Suppose this admitted Communist wants to make a speech in your community. Should he be allowed to speak or not?" (The prefatory sentence introduced a *series* of questions; three questions concerning the employment of this person, e.g., in a defense industry, preceded the item on free speech.) 2. "There are some people whose ideas are considered bad or dangerous by other people. For instance, somebody who is against all churches and religion. If such a person wanted to make a speech in your city (town, community) against churches and religion should he be allowed to speak or not?" 3. "Or consider a person who favored government ownership of all the railroads and all big industries. If this person wanted to make a speech in your community favoring government ownership of all the railroads and big industries, should he be allowed to speak or not?"

Department of Labor announced that 48 ± 15% of the population was gainfully employed.

If we are correct in identifying variations in survey context as the source of these large errors, then some straightforward remedies are possible. One strategy would be to replicate surveys *in their entirety*. Past experience with the replication of entire surveys is encouraging. Tables 6 and 7 present examples from two such replications. The data in Table 6 are derived from Samuel Stouffer's (1955) classic study, *Communism, Conformity, and Civil Liberties;* those in Table 7 are from a recent study of public attitudes toward privacy and confidentiality conducted under the auspices of the Committee on National Statistics (National Academy of Sciences). The results of these studies show that items from fully replicated surveys can and do show remarkable consistency in results.

TABLE 7: *Results from a Fully Replicated 1976 Survey on Attitudes Toward Surveys and Government*

Estimate of percentage of population who believe	Survey conducted by		Discrepancy
	Census	SRC	
Surveys serve a "good purpose"	46%	52%	−6%
Surveys are "right" almost always or most of the time	41%	40%	1%
That you can trust government always or most of the time	37%	39%	−2%
That government is run "for a few big interests"	60%	60%	0%
Sample size	599	588	

Note. Census = U.S. Bureau of the Census; SRC = Survey Research Center, University of Michigan. Data are from a study of public attitudes toward privacy and confidentiality conducted under the auspices of the Committee on National Statistics, National Academy of Sciences. Questions responded to were: (1) "In general do you feel that surveys usually serve a good purpose or do you feel that they are usually a waste of time and money?"; (2) "How often do you think that you can trust the results of surveys, do you think they are almost always right, right most of the time, only some of the time, or hardly ever right?"; (3) "How much of the time do you think you can trust the government in Washington to do what is right—just about always, most of the time, or only some of the time?"; (4) "Would you say that the government is pretty much run for a few big interests looking out for themselves or that it is run for the benefit of all the people?"

Nonetheless, since economic considerations ordinarily make the replication of entire surveys economically infeasible, we must improve our understanding of the errors that afflict social indicator estimates derived from different surveys. Recent work (Presser & Schuman, 1975; Duncan & Schuman, Note 1) on context and wording effects in surveys shows some promise of broadening our understanding of these phenomena. However, results from this research also suggest that context and wording artifacts may disrupt the pattern of correlations between subjective indicators and other variables, as well as bias the estimates of the level of an indicator. This research "argues for caution in making the assumption that multivariate patterns of responses will be relatively unaffected [by context] . . . even though the univariate response distributions are affected" (Duncan & Schuman, Note 1, p. 9).

Basic research in this area should be encouraged. Systematic work using archives of past surveys will be useful in defining the generality and magnitude of these problems. Hopefully, such work will eventually point the way (a) to general principles regarding the types of subjective social indicators [12] that are particularly vulnerable to context artifacts,

[12] Preliminary examination of other indicator estimates suggests that the degree of variability in indicator estimates derived from different surveys *may* be a function of an identifiable dimension of the indicator questions themselves. Thus, we know that sample surveys can show remarkable consistency in their estimates of the demographic characteristics of the population and that election predictions have been quite accurate in recent years. Furthermore, attitudes that have a relatively well-defined place in public discussions, such as attitudes toward capital punishment, fertility expectations, and political party identification, also seem to yield rather consistent estimates in different surveys. However, there appears to be emerging some evidence that the most unstable indicators are those involving questions that are the most amorphous in their *meaning* (e.g., What is "confidence" or "trust"?), the most imprecise in their *referents* (e.g., Who are the people "in charge of running organized religion"?), and that involve the most arbitrariness in the selection of a *response* category (e.g., We know what it means to be "for" or "against" the death penalty, but what does it mean to have "a great deal of confidence" vs. "some confidence" or to be "very happy" vs. "pretty happy"?).

and (b) to the particular types of contexts that lead to such biases. Hypotheses derived from such descriptive research might ultimately be tested by experimental work in future national surveys.

Further consideration should also be given to the effects of interviewer training, differential response rates, and variations in fieldwork and sampling methodologies upon the comparability of data collected by different survey organizations. Research on survey practices and data quality being conducted by the American Statistical Association (Bailar & Lanphier, 1977) should provide much useful information in this area. However, we would also suggest that the survey organizations themselves should consider undertaking coordinated methodological research. This work might, for example, provide some small numbers of baseline indicators that would be estimated by each organization. Such data might be useful in deciding when and how data from different sources could be validly integrated, and it would also aid in the detection of errant survey practices. Coordination between survey houses would also facilitate systematic study of the variety of methodological problems that can cause estimates made by different houses to disagree.[13]

There is much that can be done to improve our understanding of the errors that cause our survey data to yield inconsistent estimates of the subjective state of the national population. Our suggestions constitute only a modest and tentative beginning. However, the problem posed by such inconsistencies is an important one. Failure to undertake the fundamental research necessary to improve our understanding of the genesis of such discrepancies may well leave us with time-series that must bear the *caveat,* "Inferences about changes in the state of the nation may be misleading if based upon indicator fluctuations of less than 15% (i.e., fluctuations representing 20 million American adults)." Social indicators that are this insensitive would be of limited usefulness except for confirming obvious social changes after they have happened.

[13] In planning such research it should be kept in mind that discrepancies of the sort we have discussed are not an embarrassment peculiar to the social sciences. Analogous problems arise in the standardization of constants and the coordination of research between laboratories in the physical sciences (see, e.g., Shewhart & Deming, 1939; Steiner, 1975; Youden, 1975).

REFERENCE NOTE

1. Duncan, O. D., & Schuman, H. *An experiment on order and wording of questions.* Unpublished manuscript, Department of Sociology, University of Arizona, 1977.

REFERENCES

Asch, S. Studies of independence and conformity. *Psychological Monographs,* 1956, *70*(9, Whole No. 416).

Bailar, B., & Lanphier, C. *Development of survey methods to assess survey practices: A report of the American Statistical Association pilot project on the assessment of survey practices and data quality in surveys of human populations.* Washington, D.C.: American Statistical Association, 1977.

Bishop, Y., Fienberg, S., & Holland, P. *Discrete multivariate analysis.* Cambridge, Mass.: MIT Press, 1974.

Blalock, H. *Social statistics* (2nd ed.). New York: McGraw Hill, 1972.

Campbell, A. Subjective measures of well-being. *American Psychologist,* 1976, *31,* 117–125.

Caplan, N., & Barton, E. *Social indicators 1973: A study of the relationship between the power of information and utilization by federal executives.* Ann Arbor: University of Michigan, Institute for Social Research, 1976.

Converse, P. Change in the American electorate. In A. Campbell & P. Converse (Eds.), *The human meaning of social change.* New York: Russell Sage, 1972.

Current Opinion, November 1, 1972.

Davis, J. A. Communism, conformity, cohorts, and categories: American tolerance in 1954 and 1972–73. *American Journal of Sociology,* 1975, *81,* 491–513. (a)

Davis, J. A. Hierarchical models for significance tests in multivariate contingency tables: An exegesis on Goodman's recent papers. In H. Costner (Ed.), *Sociological methodology: 1973–4.* San Francisco: Jossey-Bass, 1975. (b)

Duncan, O. D. Developing social indicators. *Proceedings of the National Academy of Sciences,* 1974, *71,* 5096–5102.

Etzioni, A., & Nunn, C. The public appreciation of science in contemporary America. *Daedalus,* 1974, *103,* 191–205.

Executive Office of the President: Office of Management and the Budget. *Social indicators, 1973.* Washington, D.C.: U.S. Government Printing Office, 1973.

Executive Office of the President: Office of Management and the Budget. *Social indicators, 1976.* Washington, D.C.: U.S. Government Printing Office, 1977.

Goodman, L. The analysis of multidimensional contingency tables. *Technometrics,* 1971, *13,* 33–61.

Goodman, L. A general model for the interpretation of surveys. *American Journal of Sociology,* 1972, *77,* 1035–1086.

Harris, L., & Associates. *The Harris survey.* New York: Harris, December 6, 1973.

Harris, L., & Associates. *The Harris survey.* New York: Harris, September 30, 1974.

Harris, L., & Associates. *A survey of public and leadership attitudes toward nuclear power development in the United States* (A survey conducted for Ebasco Services, Inc.). New York: Ebasco Services, 1975. (a)

Harris, L., & Associates. *The Harris survey.* New York: Harris, October 6, 1975. (b)

Harris, L., & Associates. *The Harris survey.* New York: Harris, March 22, 1976.

Harris, L., & Associates. *The Harris survey.* New York: Harris, March 14, 1977.

Kish, L. *Survey sampling* (2nd ed.). New York: Wiley, 1965.

Kling, J., & Riggs, L. *Woodworth and Schlosberg's experimental psychology*. New York: Holt, Rinehart & Winston, 1971.

Land, K., & Felson, M. A general framework for building dynamic macro social indicator models. *American Journal of Sociology*, 1976, *83*, 565–604.

Land, K., & Spilerman, S. *Social indicator models*. New York: Russell Sage, 1975.

Mason, K., Czaijka, J., & Arber, S. Change in U.S. women's sex role attitudes, 1960–1974. *American Sociological Review*, 1976, *41*, 573–598.

National Academy of Sciences (Committee on National Statistics). *Surveying crime*. Washington, D.C.: Author, 1976.

National Opinion Research Center (University of Chicago). *National Data Program for the social sciences: Codebook for the General Social Survey*. Chicago: Author, 1972–1977. (annual)

National Science Board. *Science indicators, 1974*. Washington, D.C.: U.S. Government Printing Office, 1975.

National Science Board. *Science indicators, 1976*. Washington, D.C.: U.S. Government Printing Office, 1977.

Posner, M. *Cognition: An introduction*. Glenview, Ill.: Scott, Foresman, 1973.

Presser, S., & Schuman, H. Question wording as an independent variable in survey analysis—A first report. *Proceedings of the American Statistical Association (Social Statistics Section)*, 1975, 16–25.

Romsoy, N. Social indicators in the United States and Europe. In R. Van Dusen (Ed.), *Social indicators, 1973: A review symposium*. Washington, D. C.: Social Science Research Council, 1974.

Schachter, S. *Emotion, obesity and crime*. New York: Academic Press, 1971.

Sheldon, E., & Parke, R. Social indicators. *Science*, 1975, *188*, 693–699.

Shewhart, W., & Deming, W. *Statistical method from the point of view of quality control*. Washington, D.C.: U.S. Department of Agriculture, Graduate School, 1939.

Steiner, E. *Planning and analysis of results of collaborative tests*. Washington, D.C.: Association of Official Analytical Chemists, 1975.

Stouffer, S. *Communism, conformity, and civil liberties*. New York: Doubleday, 1955.

Sudman, S., & Bradburn, N. *Response effects in surveys*. Chicago: Aldine, 1974.

Thiel, H. On the estimation of relationships involving qualitative variables. *American Journal of Sociology*, 1970, *76*, 103–154.

U.S. Bureau of the Census. *1970 census of population and housing: Reports of evaluation and research program* (Series PHC-E). Washington, D.C.: U.S. Government Printing Office, 1970.

U.S. Bureau of the Census. *Current population reports: Consumer income* (Series P-60, No. 103). Washington, D.C.: U.S. Government Printing Office, September 1976. (a)

U.S. Bureau of the Census. *Current population reports: Voter participation* (Series P-20, No. 304). Washington, D.C.: U.S. Government Printing Office, December 1976. (b)

U.S. Bureau of the Census. *Current population reports: Population characteristics (Educational attainment in the United States, March 1977 and 1976)* (Series P-20, No. 314). Washington, D. C.: U.S. Government Printing Office, December, 1977. (a)

U.S. Bureau of the Census. *Current population reports: Population projections and estimates* (Series P-25, No. 643). Washington, D.C.: U.S. Government Printing Office, January 1977. (b)

U.S. Congress, Clerk of the House. *Statistics of the Presidential and Congressional election of November 7, 1972*. Washington, D.C.: Author, 1972.

U.S. Department of Health, Education and Welfare. *Toward a social report*. Washington, D.C.: U.S. Government Printing Office, 1969.

Youden, W. *Statistical techniques for collaborative tests*. Washington, D. C.: Association of Official Analytical Chemists, 1975.

Zapf, W. Social indicators, 1973: Comparison with social indicator reports of other nations. In R. Van Dusen (Ed.), *Social indicators 1973: A review symposium*. Washington, D. C.: Social Science Research Council, 1974.