

In Search of House Effects: A Comparison of Responses to Various Questions by Different Survey Organizations

TOM W. SMITH

DISCREPANCIES between Harris and the General Social Survey (GSS) on the confidence in leaders questions have raised the more general question of whether trend analysis can be carried out when the surveys being used were done not only at different times but also by different survey organizations ("houses").¹ If interhouse differences are common in occurrence and large in magnitude, then overtime analysis employing data from different houses becomes highly suspect. It becomes difficult, if not downright impossible, to separate the time effects from the house effects.

To explore this problem, a search was made for instances of different houses asking the identical question at approximately the same point in time. By thus controlling for both question wording and time, the number of factors possibly causing differences between houses is reduced. There remain two main types of factors influencing dif-

Abstract This paper examines the responses to various questions asked by different survey organizations. It considers the question of whether different survey organizations produce similar measurements of public opinion or whether house effects produce dissimilar measurements of the same population.

Tom W. Smith is an Associate Study Director for the National Opinion Research Center.

¹ On a number of confidence items Harris and GSS questions show large differences in marginals and divergent trends. On differences in question wording and form as well as on results, see Ladd (1976-77), Smith (1977), and Turner and Krauss (1978). The GSS is conducting several methodological experiments on its 1978 survey to examine this matter further and Elizabeth Martin, Institute for Research in Social Science, University of North Carolina; Tom W. Smith, NORC; and D. Garth Taylor, NORC, are engaged in an analysis of the GSS-Harris confidence items.

ferences: survey-specific response effects, such as question order and position,² and general house effects, such as sampling procedures, interview training, and field supervision. The former factor can occur between any surveys which were not exact replications and can occur within houses as well as between houses. The latter factor is not associated with particular surveys, but affects in general all surveys conducted by a house. While it is not possible to separate these factors rigorously with the data at hand, some attention will be devoted to assessing the role of these factors.

The first type of interhouse comparisons examined were split-surveys or cooperative sampling in which two or more houses conducted the field work for a single study. NORC has participated in at least four of these arrangements: the 1954 Stouffer study with AIPO, the 1955 *Academic Mind* study with Roper, the 1960 Steiner television study with Roper, and the 1977 Medical Care Expenditure study with Research Triangle Institute.³ No interhouse comparisons were made on the *Academic Mind* survey and the 1977 data are not yet available. Cross-house comparisons on the Stouffer and Steiner studies showed very similar results. Generally speaking, there were no apparent house effects on these two studies.

Of course, it can be legitimately argued that these studies represent special cases. With two houses working together on a study, it is usual that special steps are taken to coordinate matters and insure comparability. Thus, while the close correspondence between the houses on these surveys is encouraging, it has limited generalizability. The next type of interhouse comparisons examined comprised those instances when two houses independently asked the same question at the same time. (At the "same time" means that the surveys were either conducted within several months of each other or that a survey by one house was bracketed by prior and subsequent surveys by another house.) Seventeen examples of this type were located before 1950. Since these points had little relevance to the present or to most trend analysis—which can only rarely reach back before the late 1940s—and since marginals were presently available for only 2 of these 17 comparisons, these data were ignored. No examples of this type of comparisons were found in the 1950s or 1960s.⁴ Although

² On these response effects, see Sudman and Bradburn (1974:33-35) and Schuman (1974:10-14).

³ See Lazarsfeld and Thielens (1958); Stouffer (1963); and Steiner (1963).

⁴ A number of examples were, however, found of two houses asking similar, but not identical questions. These included NORC-SRC on job satisfaction, AIPO-Harris on capital punishment, AIPO-Harris on gun regulation, AIPO-Harris on the admission of

some examples almost certainly exist, none were discovered from available archival sources.

In the 1970s, 38 examples were found. All are cases in which a question selected from Gallup-American Institute of Public Opinion (AIPO), Michigan's Survey Research Center (SRC), or Roper for the GSS was repeated by the original house, thereby overlapping with the GSS series. Table 1 gives the proportions for these questions (see Appendix: Question Wording for the exact usages). Two types of analysis were conducted on these items. For the 33 instances when surveys were fielded within five months a direct comparison was made between the proportions. As Table 1 reveals, the differences were distributed as follows:

Differences in Proportion	Number of Instances
.0 to .01	3
.01 to .02	8
.02 to .03	8
.03 to .04	1
.04 to .05	7
.05 to .06	3
.06 to .07	2
.07 to .08	1
.09+	1
	33

There were significant differences in 10 instances and nonsignificant differences in the other 23 cases. The 10 differing cases were not randomly scattered among the cases but came from 2 clusters, 5 national spending items compared between a Roper survey in December 1973 and GSS74, and 2 misanthropy items compared between the 1974 SRC election survey and GSS75, plus 3 other cases—presidential vote in 1968 asked by the 1972 SRC election survey and GSS73, party identification on the 1976 SRC election survey and GSS77, and voting for a woman for president asked by AIPO in August 1975 and GSS75.

Close inspection of the national spending questions indicated a strong likelihood of a house difference in the proportion replying "don't know." Table 2 gives the proportion answering "don't know" on the two Roper surveys and GSS73 and GSS74. On every single item the GSS proportions are lower than Roper. While it is clear that the proportion "don't know" can change over time (Roper73 is lower

China to the UN, AIPO-Harris on votes for eighteen-year-olds, SRC-AIPO on party identification, and NORC-AIPO on abortions. All appear to show the same marginals, similar trends, or both. See Manpower Administration (1974); Erskine (1971a, 1971b); Social Change Archives, GSS; and P. E. Converse (1976:31, 168).

Table 1. Marginals^a

Item/GSS Mnemonic ^b : Response	Time	House	Proportion	N	Probability <.05	Difference (T ₁ -T ₂)
Capital punishment (CAPPUN): Execute	2/72	AIPO	.510	1,507	No	-.020
	3/72	GSS	.530	1,609		
	11/72	AIPO	.570	NA ^c	No	-.032
	3/73	GSS	.602	1,492		
	3/74	GSS	.630	1,480	No	-.010
	3/75	GSS	.601	1,483		
	3/76	GSS	.655	1,496		
	4/76	AIPO	.666	1,540		
	3/77	GSS	.672	1,520		
Ideal number of children (CHLDIDEL): 4 or more	3/72	GSS	.246	1,613	No	-.014
	1/73	AIPO	.205	1,549		
	2/74	AIPO	.177	1,562		
	3/74	GSS	.191	1,484		
	3/75	GSS	.169	1,488		
Attitudes toward countries BRAZIL: +3, +4, or +5	5/72	AIPO	.254	1,540		
	3/74	GSS	.325	1,474		
	3/75	GSS	.262	1,479		
	6/76	AIPO	.222	1,544		
	3/77	GSS	.250	1,517		
CANADA: +3, +4, or +5	5/72	AIPO	.781	1,540		
	4/73	AIPO	.764	1,528		
	3/74	GSS	.788	1,474		
	3/75	GSS	.774	1,481		
	6/76	AIPO	.788	1,544		
	3/77	GSS	.759	1,517		
CHINA: +3, +4, or +5	5/72(C)	AIPO	.071	1,540		
	5/72(W)	AIPO	.201	1,540		
	4/73	AIPO	.213	1,528		
	3/74	GSS	.142	1,474		
	3/75	GSS	.126	1,480		
	6/76(C)	AIPO	.068	1,544		
	6/76(W)	AIPO	.222	1,544		
	3/77	GSS	.129	1,517		
EGYPT: +3, +4, or +5	4/73	AIPO	.140	1,528		
	3/74	GSS	.188	1,474		
	3/75	GSS	.137	1,481		
	6/76	AIPO	.140	1,544		
	3/77	GSS	.187	1,516		
ENGLAND: +3, +4, or +5	5/72	AIPO	.658	1,540		
	4/73 ^e	AIPO	.595	1,528		
	3/74	GSS	.612	1,474		
	3/75	GSS	.575	1,481		
	6/76	AIPO	.633	1,544		
	3/77	GSS	.569	1,516		
ISRAEL: +3, +4, or +5	3/74	GSS	.387	1,474		
	3/75	GSS	.320	1,480		
	6/76	AIPO	.323	1,544		
	3/77	GSS	.353	1,517		

JAPAN: +3, +4, or +5	5/72	AIPO	.362	1,540		
	4/73	AIPO	.396	1,528		
	3/74	GSS	.389	1,474		
	3/75	GSS	.335	1,483		
	6/76	AIPO	.419	1,544		
	3/77	GSS	.323	1,518		
RUSSIA: +3, +4, or +5	5/72	AIPO	.177	1,540		
	4/73 ^f	AIPO	.143	1,528		
	7/73	AIPO	.174	1,544		
	3/74	GSS	.182	1,474		
	3/75	GSS	.175	1,481		
	6/76	AIPO	.075	1,544		
	3/77	GSS	.115	1,519		
Judicial punishment (COURTS): Harsher	12/72	AIPO	.744	1,504	No	.013
	3/73	GSS	.731	1,494		
Afraid to walk alone (FEAR): Afraid	3/74	GSS	.448	1,480		
	6/75	AIPO	.443	1,558		
	3/76	GSS	.439	1,497		
Women for president (FEPRES): Vote for	7/71	AIPO	.658	1,531		
	3/72	GSS	.701	1,611		
	3/74	GSS	.778	1,479		
	3/75	GSS	.778	1,489		
	8/75	AIPO	.735	1,515		
	3/77	GSS	.771	1,526		
Suitability for politics (FEPOL): Agree	3/74	GSS	.436	752		
	9/74	SRC	.435	1,012		
	3/75	GSS	.477	1,488		
Suitability for politics (FEPOLY): Men	3/74	GSS	.326	730	No	-.061
	9/74	SRC	.387	488		
Marijuana laws (GRASS): Legalize	2/72	AIPO	.152	1,513		
	1/73	AIPO	.157	1,508		
	3/73	GSS	.183	1,501		
Gun permit (GUNLAW): Opposes	10/71	AIPO	.719	1,502		
	3/72	GSS	.702	1,610		
	5/72	AIPO	.716	1,540		
	3/73	GSS	.735	1,495		
	3/74	GSS	.753	1,477		
	2/75	SRC	.706	452		
	3/75	GSS	.737	1,488		
	2/76	SRC	.726	638		
3/76	GSS	.715	1,493			
	3/77	GSS	.716	1,528	No	-.021
Misanthropy—evaluation of people FAIR: Fair	3/72	GSS	.592	1,611		
	11/72	SRC	.589	2,179		
	3/73	GSS	.573	1,503		
	11/74	SRC	.576	1,543		
	3/75	GSS	.616	1,488		
	3/76	GSS	.592	1,499	No	-.040
	11/76	SRC	.599	1,873		

Table 1.—Continued

Item/GSS Mnemonic ^b : Response	Time	House	Proportion	N	Prob- ability <.05	Differ- ence (T ₁ -T ₂)
HELPFUL: Helpful	3/72	GSS	.465	1,612	No	.001
	11/72	SRC	.469	2,174		
	3/73	GSS	.468	1,501		
	11/74	SRC	.507	1,528		
	3/75	GSS	.562	1,488		
	3/76	GSS	.431	1,498		
	11/76	SRC	.519	1,877	.030	-.055
TRUST: Trusts people	3/72	GSS	.458	1,612	No	-.001
	11/72	SRC	.458	2,179		
	3/73	GSS	.459	1,502		
	11/74	SRC	.466	1,551		
	3/75	GSS	.393	1,485		
	3/76	GSS	.444	1,497		
	11/76	SRC	.513	1,882	.004	-.073
Spending for foreign aid (NATAID): Too little	7/71	Roper	.038	1,487	No	-.009
	3/73	GSS	.042	1,503		
	12/73	Roper	.021	1,766		
	3/74	GSS	.030	1,481		
	3/75	GSS	.054	1,489		
	3/76	GSS	.029	1,494		
	3/77	GSS	.034	1,527		
Spending for military (NATARMS): Too little	7/71	Roper	.150	1,488	No	-.019
	3/73	GSS	.112	1,496		
	12/73	Roper	.150	1,764		
	3/74	GSS	.169	1,479		
	3/75	GSS	.166	1,484		
	3/76	GSS	.241	1,492		
	3/77	GSS	.232	1,553		
Spending for cities (NATCITY): Too little	7/71	Roper	.415	1,488	.007	-.067
	3/73	GSS	.482	1,499		
	12/73	Roper	.432	1,757		
	3/74	GSS	.499	1,474		
	3/75	GSS	.471	1,479		
	3/76	GSS	.426	1,492		
	3/77	GSS	.403	1,525		
Spending for crime prevention (NATCRIME): Too little	7/71	Roper	.611	1,490	No	-.026
	3/73	GSS	.646	1,497		
	12/73	Roper	.640	1,757		
	3/74	GSS	.666	1,481		
	3/75	GSS	.656	1,484		
	3/76	GSS	.657	1,489		
	3/77	GSS	.657	1,524		
Spending for drug prevention (NATDRUG): Too little	7/71	Roper	.618	1,493	No	-.021
	3/73	GSS	.659	1,483		
	12/73	Roper	.579	1,759		
	3/74	GSS	.600	1,478		
	3/75	GSS	.551	1,482		
	3/76	GSS	.587	1,493		
	3/77	GSS	.552	1,520		

Spending for education (NATEDUC): Too little	7/71	Roper	.439	1,488	.047	-.049
	3/73	GSS	.490	1,499		
	12/73	Roper	.458	1,752		
	3/74	GSS	.507	1,474		
	3/75	GSS	.490	1,487		
	3/76	GSS	.502	1,495		
	3/77	GSS	.476	1,527		
Spending for environment (NATENVIR): Too little	7/71	Roper	.559	1,495	<.001	-.131
	3/73	GSS	.611	1,498		
	12/73	Roper	.459	1,766		
	3/74	GSS	.590	1,476		
	3/75	GSS	.534	1,490		
	3/76	GSS	.548	1,494		
	3/77	GSS	.475	1,524		
Spending for welfare (NATFARE): Too little	7/71	Roper	.178	1,474	.010	-.051
	3/73	GSS	.198	1,497		
	12/73	Roper	.170	1,762		
	3/74	GSS	.221	1,481		
	3/75	GSS	.234	1,484		
	3/76	GSS	.133	1,493		
	3/77	GSS	.123	1,524		
Spending for medical care (NATHEAL): Too little	7/71	Roper	.552	1,491	No	-.040
	3/73	GSS	.608	1,497		
	12/73	Roper	.599	1,761		
	3/74	GSS	.639	1,477		
	3/75	GSS	.626	1,485		
	3/76	GSS	.605	1,491		
	3/77	GSS	.558	1,526		
Spending for space exploration (NATSPAC): Too little	7/71	Roper	.063	1,497	<.001	-.042
	3/73	GSS	.075	1,503		
	12/73	Roper	.035	1,768		
	3/74	GSS	.077	1,480		
	3/75	GSS	.074	1,490		
	3/76	GSS	.092	1,496		
	3/77	GSS	.101	1,530		
Gun ownership (OWNGUN): Owns	5/72	AIPO	.435	1,513		
	3/73	GSS	.473	1,495		
	3/74	GSS	.462	1,480		
	3/75	AIPO	.464	1,512		
	3/75	AIPO	.453	1,536		
	10/75	AIPO	.474	1,558		
	3/76	GSS	.466	1,493		
Political identification (PARTYID): Democratic	3/72	GSS	.474	1,607	No	-.018
	11/72	SRC	.403	2,702		
	3/73	GSS	.411	1,493		
	3/74	GSS	.423	1,483		
	11/74	SRC	.385	1,570		
	3/75	GSS	.404	1,486		
	3/76	GSS	.421	1,497		
	11/76	SRC	.396	2,244		
3/77	GSS	.444	1,520	.036	-.049	

Table 1.—Continued

Item/GSS Mnemonic ^b : Response	Time	House	Proportion	N	Probability <.05	Difference (T ₁ -T ₂)
Presidential vote, 1968 (PRES68): Nixon	3/72	GSS	.453	1,056	.022	.057
	11/72	SRC	.504	1,038		
	3/73	GSS	.439	1,005		
Presidential vote, 1972 (PRES72): Nixon	3/76	GSS	.577	943	No	.025
	11/76	SRC	.622	1,440		
	3/77	GSS	.597	919		
School integration if a few blacks (RACFEW): Object	3/72	GSS	.050	535		
	7/73	AIPO	.076	449		
	3/74	GSS	.045	492		
	3/75	GSS	.047	473		
	9/75	AIPO	.065	539		
	3/77	GSS	.073	481		
School integration if half black (RACHAF): Object	3/72	GSS	.232	535		
	7/73	AIPO	.305	449		
	3/74	GSS	.299	492		
	3/75	GSS	.262	473		
	9/75	AIPO	.297	539		
	3/77	GSS	.255	481		
School integration if mostly black (RACMOST): Object	3/72	GSS	.546	535		
	7/73	AIPO	.670	449		
	3/74	GSS	.646	492		
	3/75	GSS	.638	473		
	9/75	AIPO	.583	539		
	3/77	GSS	.620	481		
Work if rich (RICHWORK) ^c : Continue working	Winter 1969/70	SRC	.674	1,523	No	-.023
	Winter 1972/73	SRC	.658	2,148		
	3/73	GSS	.681	831		
	3/74	GSS	.636	837		

^a Missing values excluded from analysis but "don't knows" retained. All significance tests adjust for multistage sampling by multiplying the standard deviations by 1.414.

^b Here and elsewhere items are referred to by their standard GSS mnemonic. See Davis et al. (1977).

^c 1,500 used in calculations.

^d AIPO (5/72-C) is "Red China"; (5/72-W) is "Nationalist China (Taiwan)." AIPO (6/76-C) is "Communist China"; AIPO (6/76-W) is "Nationalist China (Taiwan)." All others are "China."

^e "Great Britain" used in AIPO (4/73). "England" used in all others.

^f "Soviet Union" used in AIPO (4/73); "Russia" used in all others.

^g Universes differ slightly between houses.

Table 2. Proportion "Don't Know" on National Spending Items, 1971-1974

Item (GSS Mnemonic)	Surveys			
	Roper71	GSS73	Roper73	GSS74
NATCRIME	.163	.061	.089	.051
NATEDUC	.099	.043	.059	.038
NATSPAC	.054	.047	.054	.036
NATFARE	.105	.043	.081	.040
NATAID	.129	.055	.072	.040
NATARMS	.170	.059	.118	.067
NATCITY	.262	.120	.226	.147
NATHEAL	.107	.035	.065	.035
NATENVIR	.138	.057	.086	.066
NATDRUG	.129	.063	.096	.055
Mean	.136	.058	.095	.058

than Roper71 in all but one case), it is probable that Roper generates a higher level of "don't knows" than GSS does.⁵

With the "don't knows" excluded from analysis (see Table 3) the differences between Roper73 and GSS74 are reduced in eight instances, unchanged once, and increased once. For 2 of the 5 significant differences (NATEDUC and NATCITY), the reduction was sufficient to make the differences fall within sampling error while three items remained significant (NATSPAC, NATFARE, and NATENVIR.)

Table 3. Change in Difference Between Roper73 and GSS74 on National Spending Items with "Don't Knows" Excluded

Item (GSS Mnemonic)	"Don't Knows" In	"Don't Knows" Out	Change
NATHEAL	-.040	-.021	-.019
NATCITY	-.067	-.027	-.040
NATARMS	-.019	-.011	-.008
NATAID	-.009	-.009	.000
NATFARE	-.051	-.046	-.005
NATSPAC	-.042	-.043	+.001
NATEDUC	-.049	-.040	-.009
NATCRIME	-.026	.000	-.026
NATENVIR	-.131	-.130	-.001
NATDRUG	-.021	+.005	-.026
	-.046	-.032	-.013
		(absolute .033)	

⁵ In a Roper survey on the United Nations conducted in 1977 for the League of Women Voters Education Fund, questions previously asked by NORC, AIPO, and Potomac Associates all received substantially higher levels of "don't knows" than they had earlier. In these cases, however, there is no temporal overlap between houses to help control for time (League of Women Voters Education Fund, 1977). On other differences between houses, see J.M. Converse (1976-77:515-30).

A second factor of note is that the differences have a definite direction. Even after the "don't knows" have been corrected for, the proportion answering "too little" on Roper73 is lower than on GSS in 8 out of 10 cases (an average difference of $-.032$). This could represent a house effect but alternative explanations are equally plausible. It could represent a seasonal effect since the lower spending support in Roper was registered in December, the highest month for consumer spending. Having just made or being about to make high personal expenditures, people might be in a fiscally more conservative mood. Or, there could be real shifts because of changing historical events and conditions over the three months between the surveys. Such a likely event was the first energy crisis (the oil embargo crunch). In December 1973, energy was one of the top domestic stories on 15 days out of 31, but in March 1974 it made the headlines only once.⁶ While this may have had a general impact on spending, it is clear that it had an impact on the environment question. Support for the environment was much lower in December ($-.131$) and this was two to three times greater than the other two significant differences ($-.043$ and $-.046$). Clearly, many people saw an unfavorable connection between environmental spending and energy (the long delays in the Alaskan pipeline, East Coast offshore drilling, western coal mining, etc.). Whether there is a similar effect on other items is less obvious, but quite possible. Even without accepting a generalized energy crisis effect on other items, we see that it has an impact since the average difference excluding the environment item drops from $-.032$ to $-.021$. In brief, the differences on the national spending items seem to indicate that there may be a house effect involving "don't knows." For the three items significantly different once the "don't knows" are removed, one difference is clearly caused by nonhouse effects and the others may be as plausibly explained by seasonal or historical effects as by house effects.

Turning to the next cluster, we see that 2 of the 3 misanthropy items asked on the GSS and the SRC election series differ in 1974-75. What makes these differences so interesting is that two years previously these same items showed virtually no differences at all. From 1972 through 1974 these items appear to be very constant, both within and across houses, but since then they have shown considerable fluctuation. The proportion considering people helpful rose .055 between SRC74 and GSS75 and then fell .131 from GSS75 to GSS76—the largest annual change recorded on any GSS item. The proportion trusting people fell .073 from SRC74 to GSS75, and then rose by .051

⁶ Based on top stories listed in *The Official Associated Press Almanac 1975* (1975).

to GSS76. The lack of differences between houses in 1972-73 and the largest fluctuations between GSS75 and GSS76 indicate that the 1974-75 differences may well be due to other reasons besides house effects. Clearly the behavior of the misanthropy items merits closer inspection.

With the national spending and misanthropy clusters examined, there remain three other questions that show significant differences. Voting for a woman for president differs by .043 between GSS75 and AIPO in August 1975, five months later. The five-month interval was the maximum time difference used here for direct survey-to-survey comparisons, and when "don't knows" are excluded, the difference drops to .040 and becomes insignificant. Clearly this is a borderline case. The proportion Democratic differs by $-.049$ between the 1976 Michigan election survey and GSS77. While no significant differences appear between the 1972 election survey and GSS73 or between the Michigan 1974 election survey and GSS75, the small differences are in the same direction as in 1976-77, a point discussed below. The last case, presidential choice in 1968, has a small but important difference in question wording. On the GSS, Humphrey is the first candidate mentioned, while on SRC, Nixon is named first (there are other differences as well; see the Appendix: Question Wording). This ordering has a known effect on responses, increasing the proportion choosing the first mentioned candidate.⁷ As predicted, the GSS point shows .065 less for Nixon than SRC does (GSS72 shows .051 less than SRC).

In brief, it appears that of the nine possible instances of house effects, several can be credited to other causes, the presidential voting differences to an order effect, and the environmental spending difference to a historical effect. Other differences, such as on the remaining national spending items and the misanthropy items, may be due to house effects, but alternative explanations are at least equally persuasive. One fairly substantial example of a house effect appears to be the proportion of "don't knows" on the GSS and Roper national spending questions.

To carry the analysis of house effects further, an analysis was made of the trends shown by 32 of the 38 items, because it was possible to compare trends in these 32 instances. To ascertain the comparability of trends, no-change or constant models were first fitted to the GSS and non-GSS series. If the constant model proved inadequate to explain

⁷ A candidate ordering effect occurs on both the actual ballot as well as in surveys and is greater when intensity of support for candidates is less. (James Rabjohn, University of Chicago/NORC, personal communication, 9/7/77.) On the presidential voting in 1972 there are signs of a similar but smaller difference between Michigan and GSS.

the series, a linear change model was fitted to the series. Three results could come from this second test. The data could show (1) a linear trend with no significant variation, (2) a significant linear component with a significant amount of unexplained variation, or (3) no significant linear trend.⁸

Next, the house series were compared to see if the GSS and non-GSS series were similar to each other. Often the comparison of interhouse trends was quite difficult. The two series rarely started or ended at the same point in time, so that they only approximately covered the same time span. To match the time spans as closely as possible, it was often necessary to use only part of one series (see Table 4 for the selection of time points). Also, there were often hardly enough points to give a solid measure of time trends. This was especially true for non-GSS series, where only a single point was available in 5 instances, only two points in 18 cases (a minimum of three data points being necessary to detect a nonlinear trend), and three or more points in 10 cases. This created problems when comparing trends, since a one-point "series" is a contradiction and a two-point "series" can only be constant or linear, while a series with three or more points can be constant, linear, linear-component, or nonlinear. In the case of single-point "series," this point was compared to the two points that bracketed it. If it was not bracketed by points from the other house, no trend analysis was done. To handle the two-point cases, it was necessary to consider whether a nonlinear trend on one series was really different from a linear or constant trend produced by a two-point series from the other house.

The house series were judged to be similar if (1) they both tested out as constant and their pooled proportions were not significantly different from each other; (2) a point bracketed by others had a constant fit with these points; (3) both trends tested as linear or linear-component and there was no significant difference between their slopes; or (4) a bracketed point fit in a linear model with the preceding and following points. Trends were judged different when (1) the same type of model applied but the pooled proportions or slopes differed significantly, or (2) different models applied to the separate

⁸ For the details of the statistical tests applied here, see Taylor (1976). In brief, the first hypothesis tested is that the sample proportions are from a constant universe value, which is estimated to be the pooled average of the proportions. The criterion for the goodness-of-fit is the chi-square statistic that divides the squared deviation of the observed value from the predicted value by the variance of the observed value. This is referred to as the "test for homogeneity." The next hypothesis tested is that the sample proportions are from a linear universe trend. The chi-square goodness-of-fit test is used to compare the actual proportions with their linear estimates. This is referred to as the "test for linearity."

series and this did not appear to be due to the artifact of when or how often the item appeared. Series that fit different models but which showed evidence that this might be due to a shortage of data points and/or differing time spans were classed as "intermediate" (see Table 4).

Inspection of Table 4 indicates that in 21 instances the series were similar; in three instances, intermediate; and in eight instances, different. As in the case of the marginal comparisons, the differences were clustered. Three of the eight disagreements were on country items (EGYPT, ENGLAND, JAPAN), two were from the national spending variables (NATENVIR, NATSPAC), two were misanthropy items (HELPFUL, TRUST), and the last was party identification. The observed differences in these countries could be due to several factors. First, the time series spanned by the two houses were different, with AIPO covering 1972/73-1976 and GSS 1974-1977. To look at this possible effect some more, the subseries for 1975-1977 (GSS-AIPO-GSS) was examined. In each case, significant differences remained. Second, we are not observing items with clear directional trends but, like the national spending items above or the expectation of war question, an item subject to large short-term fluctuations. Current events and/or shifts in foreign policy could well have such an effect on the ranking of countries.⁹ A final factor that may contribute to the differences is an order effect, like the one noted in the case of presidential vote above. The Appendix shows that AIPO has asked a different mix of countries in differing orders. While there is no proof of an order effect in this case, such an effect may exist. The national spending and misanthropy variables have been discussed above. The environmental spending difference comes from an episodic effect and the space spending difference may come from this source as well. On party preference, the proportion Democratic is constant for both series. The pooled proportion Democratic estimated from the GSS's (.430) is, however, significantly greater than the Michigan election estimate (.396). Although this -.034 difference is not stable, showing up as significant in only one of the three individual comparisons analyzed earlier, there is a consistent direction to the differences. Part of the difference apparently results from a greater tendency to code respondents "other," "no preference," or some other unread response on the Michigan election surveys than on the GSS's.¹⁰ Among the three mentioned responses (Republican, Democratic, Independen-

⁹ If all these were from one house, this is probably the conclusion most researchers would come to about trends for these countries.

¹⁰ An inspection on GSS and Michigan interview specification, however, revealed no obvious reason for this.

Table 4. Trend Comparisons

GSS Mnemonic	Combined Model	Separate Models		Difference in Models Significant	Trends Compared
		GSS	Other		
CAPPUN	Linear	Linear	Linear	No	Similar
CHLDIDEL	Linear	Linear	Linear	No	Similar
COUNTRIES					
BRAZIL	Nonlinear	Nonlinear	Constant	Yes	Intermediate
CANADA	Constant	Constant	Constant	No	Similar
CHINA	NTAPP				NTAPP
EGYPT	Nonlinear	Nonlinear	Constant	Yes	Different
ENGLAND	Linear component	Constant	Constant	Yes	Different
ISRAEL	Nonlinear	Nonlinear	NTAPP	NTAPP	Similar ^a
JAPAN	Nonlinear	Linear	Constant	Yes	Different
RUSSIA	Linear component	Linear	Linear	No	Similar ^b
COURTS	NTAPP				NTAPP
FEAR	Constant	Constant	NTAPP	No	Similar
FEPRES	Linear component	Linear component	Linear	No	Similar
FEPOL	Constant	Constant	NTAPP	No	Similar
FEPOLY	NTAPP				NTAPP
GRASS	NTAPP				NTAPP
GUNLAW	Constant	Constant	Constant	No	Similar
MISANTHROPY					
FAIR	Constant	Constant	Constant	No	Similar
HELPFUL	Nonlinear	Nonlinear	Linear	Yes	Different ^c
TRUST	Nonlinear	Nonlinear	Linear	No	Different ^d
NATAID	Nonlinear				Similar ^e
NATARMS	Linear component				Intermediate
NATCITY	Nonlinear				Similar

TOM W. SMITH

Table 4.—Continued

GSS Mnemonic	Combined Model	Separate Models		Difference in Models Significant	Trends Compared
		GSS	Other		
NATCRIME	Constant				Similar
NATDRUG	Linear component				Similar
NATEDUC	Constant				Similar
NATENVIR	Linear component				Different ^f
NATFARE	Linear component				Similar
NATHEAL	Nonlinear				Similar
NATSPAC	Linear component				Different
OWNGUN	Constant	Constant	Constant	No	Similar
PARTYID	Constant ^g	Constant	Constant	Yes	Different
PRES68	NTAPP				NTAPP
PRES72	NTAPP				NTAPP
RACFEW	Constant	Constant	Constant	No	Similar
RACHAF	Constant	Constant	Constant	No	Similar
RACMOST	Constant	Constant	Linear	No	Intermediate ^h
RICHWORK	Constant	Constant	NTAPP	No	Similar

^a 1975-1977 trend constant.^b Excludes AIPO (4/73) "Soviet Union."^c Various subseries were examined 3/72-3/73 and 11/72-11/74 were constant and 3/73-3/75 was linear.^d Constant for early subseries (3/72-3/73; 11/72-11/74) and for all points excluding GSS75.^e Since the GSS series runs from 1973 to 1977 while the AIPO series runs only from 1971 to 1973, it was not appropriate to compare them. Instead the 7/71-12/73 and 3/73-3/74 subseries were checked (3 points in each). Agreement on both lead to a ranking of "similar," disagreement on both a "different" rank, a split decision led to an "intermediate" rank. "Don't knows" were excluded from analysis.^f See discussion of this variable in preceding section.^g Omits GSS72.^h For the two AIPO points the difference is significant at the .042 level making it a "linear" change.

IN SEARCH OF HOUSE EFFECTS

dent), the difference between Michigan and GSS falls to $-.022$ and becomes insignificant.

Of the three intermediate cases there is one additional item from both the country and national spending questions (BRAZIL and NATARMS) and a school integration item (RACMOST). Possible factors involved in the country and national spending clusters are discussed above. The school integration question is the double-filtered part of the integration question and the two preceding parts show constant and similar trends (see Appendix for filters and wording). On this part, however, the two AIPO data points show a significant difference at the .042 level while the GSS is constant. Obviously, this is a borderline case.

The preceding search for house effects among proportional differences and trends revealed a number of possible candidates. In 10 out of 33 instances response proportions were significantly different. Consideration of nonhouse effects indicated that at least two of the differences (PRES68 and NATENVIR) were due to other factors and the remaining might also be due to temporal or other unisolated factors (e.g., order or context). The analysis also pinpointed the "don't know" response level as a possible example of house effects. The trend analysis showed 8 instances of disagreement, 3 intermediate cases, and 21 nonconflicting series. As with the one-to-one comparison of proportions, nonhouse effects account for at least some of these differences.

It can be argued that because the differing cases are clustered primarily among three questions—national spending, countries, and misanthropy—house effects are not a general or random occurrence but concentrated among particular questions. It might even be argued that since differences are largely restricted to these questions, time, placement, or other factors rather than true house effects are responsible. In sum, while the available data are much less complete than would be desired, what does exist suggests that house effects are not a large and systemic problem. It is clear, however, that both general house effects and survey-specific response effects do occur. To deal with this problem three courses should be followed. First, house and other response effects should be routinely checked for whenever analysis compares two surveys. Second, methodological research is needed in order to (1) document procedural differences between houses and then measure the effect of these differences on results, and (2) assess and calibrate other response effects.¹¹ Third, in plan-

¹¹ For a pioneering evaluation of procedural differences between surveys, see Bailar and Lamphier (1977). The literature on response effects is large; see, for example, the following two special issues: Ferber (1977) and Alwin (1977).

ning replication studies, close attention should be given to minimizing such possible effects by duplicating as far as possible, not just question wording but interviewer specifications, question placement, coding rules, and other features.

Appendix: Question Wording

CAPPUN

(a) AIPO and GSS 1972-73: Are you in favor of the death penalty for persons convicted of murder?

(b) GSS 1974-76: Do you favor or oppose the death penalty for persons convicted of murder?

CHLDIDEL

(a) AIPO and GSS: What do you think is the ideal number of children for a family to have?

(b) AIPO has response "No opinion" while GSS has responses "As many as you want" and "Don't know."

COUNTRIES

(a) AIPO 5/72: Here is an interesting experiment. You will notice that the boxes on the card go from the HIGHEST POSITION OF PLUS 5, or a country which you like very much, to the LOWEST POSITION OF MINUS 5, or a country you dislike very much. How far up the scale or how far down the scale would you rate the following countries?

Russia, Brazil, Red China, Japan, England, Nationalist China (Taiwan), Canada

(b) AIPO 4/73: Has same wording with following list:

[China, Canada, West Germany, Great Britain, Japan, Italy, France, Chile, Sweden, Soviet Union, Egypt]

(c) AIPO 7/73: Here is an interesting experiment. You notice that the boxes on this card do [*sic*] from the HIGHEST POSITION OF PLUS 5—or something you like very much—all the way down to the LOWEST POSITION OF MINUS 5—or something you dislike very much. How far up the scale or how far down the scale would you rate the following organizations:

CORE (Congress of Racial Equality), FBI (Federal Bureau of Investigation), Ku Klux Klan, AMA (American Medical Association), John Birch Society, NAACP (National Association for Advancement of Colored People), AFL-CIO (labor unions), NAM (National Association of Manufacturers), CIA (Central Intelligence Agency), ACLU (American Civil Liberties Union), your local police department, U.S. Supreme Court, Congress, the press, the United States, Russia

(d) AIPO 6/76: Here is an interesting experiment. You notice that the ten boxes on this card go from the highest position of plus five—for something you have a very favorable opinion of—all the way down to the lowest position of minus five—or something you have a very unfavorable opinion of. Please tell me how far up the scale or how far down the scale you rate the following nations.

England, Communist China, Russia, Sweden, Cuba, France, W. Germany, Italy, Japan, Egypt, Israel, Brazil, Argentina, Australia, Chile, Nationalist

China (Taiwan), Canada, India, Iran, Holland, Switzerland, Poland, Mexico, United States, S. Africa, Philippines

(e) GSS: You will notice that the boxes on this card go from the highest position of "plus 5" for a country which you *like* very much, to the lowest position of "minus 5" for a country you *dislike* very much. How far up the scale or how far down the scale would you rate the following countries?
READ EACH ITEM:

Russia, Japan, England, Canada, Brazil, China, Israel, Egypt

COURTS

(a) AIPO and GSS: In general, do you think the courts in this area deal too harshly or not harshly enough with criminals?

FEAR

(a) AIPO and GSS: Is there any area right around here—that is, within a mile—where you would be afraid to walk alone at night?

FEPRES

(a) AIPO omits "were" from GSS question below.

(b) GSS: If you party nominated a woman for President, would you vote for her if she were qualified for the job?

FEPOL

(a) SRC and GSS: Tell me if you agree or disagree with this statement: Most men are better suited emotionally for politics than are most women.

FEPOLY

(a) SRC and GSS: Would you say that most men are better suited for politics than are most women, that men and women are equally suited, or that women are better suited than men in this area.

GRASS

(a) AIPO and GSS: Do you think the use of marijuana should be made legal, or not?

GUNLAW

(a) AIPO and GSS: Would you favor or oppose a law which would require a person to obtain a police permit before he or she could buy a gun?

(b) SRC omits "or she."

MISANTHROPY

(a) SRC and GSS:

(1) FAIR: Do you think most people would try to take advantage of you if they got a chance, or would they try to be fair?

(2) HELPFUL: Would you say that most of the time people try to be helpful, or that they are mostly just looking out for themselves?

(3) TRUST

Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?

(b) GSS accepts responses of "depends" and "don't know." SRC uses "don't know" only.

NATIONAL PROBLEMS

(a) GSS and Roper: 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. First (READ ITEM A). . . are we spending too much, too little, or about the right amount on (ITEM)?

- A. Space exploration program
- B. Improving and protecting the environment
- C. Improving and protecting the nation's health
- D. Solving the problems of the big cities
- E. Halting the rising crime rate
- F. Dealing with drug addiction
- G. Improving the nation's education system
- H. Improving the conditions of Blacks
- Roper omits "H. Improving the conditions of Blacks."
- I. The military, armaments and defense
- J. Foreign aid
- K. Welfare

OWNGUN

(a) GSS: Do you happen to have in your home (IF HOUSE: or garage) any guns or revolvers?

(b) AIPO 5/72 omits "IF HOUSE: or garage."

(c) AIPO 1975: Do you have any guns in your home?

PARTYID

(a) SRC and GSS: Generally speaking, do you usually think of yourself as a Republican, [a] Democrat, [an] Independent, or what?

(b) GSS had a precoded response of "other" each year and added the response "no preference" in 1975-1977. SRC includes both categories each year.

(c) SRC includes the bracketed articles.

PRES68

(a) SRC—IF RESPONDENT HAS EVER VOTED IN A PRESIDENTIAL ELECTION: Now, in 1968 you remember that Mr. Nixon ran on the Republican ticket against Mr. Humphrey for the Democrats and Mr. Wallace on an independent ticket. Do you remember for sure whether or not you voted in that election?

—IF RESPONDENT VOTED IN 1968 ELECTION: Which one [presidential candidate] did you vote for?

(b) GSS: Now in 1968, you remember that Humphrey ran for President on the Democratic ticket against Nixon for the Republicans, and Wallace as an Independent. Do you remember for sure whether or not you voted in that election?

A. IF VOTED: Did you vote for Humphrey, Nixon or Wallace?

PRES72

(a) SRC—IF RESPONDENT HAS EVER VOTED IN A PRESIDENTIAL ELECTION: Now, in 1972 you remember that Mr. Nixon ran on the Republican ticket against Mr. McGovern for the Democrats. Do you remember for sure whether or not you voted in that election?

—IF RESPONDENT VOTED IN 1972 ELECTION: Which one [presidential candidate] did you vote for?

(b) GSS: In 1972, you remember that McGovern ran for President on the Democratic ticket against Nixon for the Republicans. Do you remember for sure whether or not you voted in that election?

A. IF VOTED: Did you vote for McGovern or Nixon?

RACIAL INTEGRATION OF SCHOOLS

(a) GSS: Would you yourself have any objection to sending your children to a school where a few of the children are (Negroes/Blacks)?

IF NO OR DON'T KNOW TO A: Where half of the children are (Negroes/Blacks)?

IF NO OR DON'T KNOW TO B: Where more than half of the children are (Negroes/Blacks)?

(b) AIPO 7/73: Do you have any children now in grade or high school?

IF YES, ASK:

B. Would you, yourself, have any objection to sending your children to a school where a few of the children are black?

IF NO, ASK:

C. Where half are black?

D. Where more than half of the children are black?

(c) AIPO 3/75: As in 1973 except that those answering "No" to part C were not asked part D and that last word was "blacks."

RICHWORK

(a) GSS: If you were to get enough money to live as comfortably as you would like for the rest of your life, would you continue to work or would you stop working?

(b) SRC uses: "you'd" instead of "you would" and omits "or would you stop working."

References

Alwin, Dwaine (ed.)

1977 "Survey design and analysis: current issues." *Sociological Methods and Research* 6 (entire issue).

Bailer, Barbara, and C. Michael Lamphier

1978 *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.

Converse, Jean M.

1976- "Predicting no opinion in the polls." *Public Opinion Quarterly*

1977 "40:515-30.

Converse, Philip E.

1976 *The Dynamics of Party Support: Cohort-Analyzing Party Identification*. Sage Library of Social Research, Vol. 35. Beverly Hills: Sage.

Davis, James A., Tom W. Smith, and C. Bruce Stephenson

1977 *Cumulative Codebook for the 1972-1977 General Social Surveys*. Chicago: National Opinion Research Center.

Erskine, Hazel

1971a "The polls: Red China and the U.N." *Public Opinion Quarterly*. 35:125-37.

1971b "The polls: the politics of age." *Public Opinion Quarterly*. 35:482-95.

Ferber, Robert (ed.)

1977 "Recent developments in survey research." *Journal of Marketing Research* 14:(entire issue).

Ladd, Everett Carll, Jr.

1976- "The polls: the question of confidence." *Public Opinion Quarterly*. 1977 40:544-52.

Lazarsfeld, Paul F., and Wagner Thielens, Jr.

1958 *The Academic Mind: Social Scientists in a Time of Crisis*. Glencoe, Ill.: The Free Press.

League of Women Voters Education Fund

1977 "Public Opinion on the UN: What Pollsters Forget to Ask."

Manpower Administration,

1974 *Job Satisfaction: Is There a Trend? Manpower Research Monograph No. 30*. Washington, D.C.: Government Printing Office.

Official Associated Press Almanac, 1975

1975 *Maplewood, N.J.: Hammond Almanac*.

Schuman, Howard

1974 "Old wine in new bottles: some sources of response error in the use of attitude surveys to study social change." Paper presented to Research Seminar Group in Quantitative Social Science, University of Surrey, England.

Smith, Tom W.

1977 *Can We Have any Confidence in Confidence? GSS Technical Report No. 1*. Chicago: National Opinion Research Center.

Steiner, Gary A.

1963 *The People Look at Television: A Study of Audience Attitudes*. New York: Knopf.

Stouffer, Samuel

1953 *Communism, Conformity, and Civil Liberties: A Cross-Section of the Nation Speaks Its Mind*. Gloucester, Mass.: Peter Smith.

Taylor, D. Garth

1976 "Procedures for evaluating trends in qualitative indicators." In James A. Davis (ed.), *Studies in Social Change Since 1948*. NORC Report 127A. Chicago: National Opinion Research Center.

Turner, Charles N., with Elissa Krauss

1978 "Fallible Indicators of the Subjective State of the Nation." *American Psychologist* (forthcoming).