The Laws of Studying Societal Change

Tom W. Smith

NORC/University of Chicago

Introduction

In 1950 Isaac Asimov formulated his Laws of Robotics:

> First Law: A robot may not injure a human being, or, through inaction, allow a human being to come to harm.
>
> Second Law: A robot must obey orders given by human beings, except where such orders would conflict with First Law.
>
> Third Law: A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

Following his model, the Laws of Studying Societal Change might go as follows:

> First Law: The way to measure change is not to change the measure.
>
> Second Law: When constant measures produce non-constant measurement, change the measure to measure change.

First Law

The First Law is easy to understand. Across time a constant stimulus is desired so one is presumably measuring the same thing in the same way at all points in time. If wording, interviewer instructions, response options, context, or other elements of the measure are altered, then the change in time and the change in measures are totally confounded and it is impossible to sort out the trends from measurement effects.

The First Law should also be easy to implement. What could be easier than just doing things the same, just repeating procedures? But surveys are complex endeavors with many elements such as sampling, interviewer composition, training and supervision, data-collection mode, question wording, context and order, coding, and so forth (Schuman, Steeh, Bobo, and Krysan, 1997).[1] Keeping all the parts well-documented and consistently applied is no small task. For both good and bad reasons implementation is often imperfect.

Bad Reasons for Violating First Law

The two main bad reasons are 1) unintentional measurement changes due to slip-ups and

---

[1] This essay discusses surveys, but much of what it considers would also apply to censuses, administrative records, and other data sources.

misunderstandings and 2) intentional changes due to either the urge to improve or changes in style. Slip-ups occur when words are unintentionally inserted or deleted, response options are altered, items omitted from scales, etc. For example, in an International Social Survey Program (ISSP) module one country dropped the last item on a scale when a list on the longer A4 paper was copied to 8.5 by 11 paper and the final item literally fell-off the page and out of the survey. In the case of the General Social Survey's (GSS) group membership battery, it replicated an item from the 1967 Verba-Nie Political Participation in America study. Their book indicated that one group was "church-affiliated groups" and this was included. However, in actuality, as the original questionnaire made clear, this was not an explicitly asked about group, but had been created from the other-specify item after data collection. Making it an explicitly asked about group notably increased mentions (Smith, 1990).

Sometimes the changes are more subtle. An initial analysis of an open-ended question asked on the 1955 and 1971 Detroit Area Study showed some major shifts. But when the 1971 coders went back and recoded the 1955 raw data, the differences disappeared. Changes in coding practices, not time, had created the shift (Duncan, Schuman, and Duncan, 1973).

Intentional changes occur because questionnaire crafters like to tinker with items to improve them. Sometimes this merely reflects the idiosyncratic preferences of different questionnaire designers. For example, between 1936 when an item on voting for a woman for president was first asked by Gallup and 1972 when the GSS adopted the item, the items was asked 10 different times with five different wordings (Smith, 1976).

Other times it may reflect more general changes in house style or research preferences. For example, from the 1950s to the 1970s, the wordings and response scales used on many policy-preference items on the National Election Studies were changed, often twice, to reflect the latest thinking on what type of scales worked best. This led to much debate and some special experimental studies to try and sort out whether certain variation in attitude constraint and other matters were the result of true change or were merely measurement artifacts (Nie, Verba, and Petrocik, 1976; Niemi and Weisberg, 1976).

Good Reasons for Violating the First Law

Good reasons for changes often involve trade-offs with other competing priorities. For example, the GSS has used the same set of open-ended questions to capture information on the occupation and industry of respondent, spouse, and father from 1972 to the present. However, the Census adopted a fundamentally new system of classifying occupation and industry around 1980. Eventually to be comparable to the Census and the many other studies that followed Census practices and because the new scheme better handled emerging occupations and industries, the GSS also adopted the new classification scheme. Thus, for respondent (and likewise for spouse and father) there is an occupation series using the 1970 Census codes for 1972-1990 and a second series for the 1980 Census codes for 1988-2004 with dual coding in 1988-1990.

Likewise, after the 2000 Census adopted a new, multiple-mentions format for racial identification, the GSS implemented a similar item for respondent racial self-classification. This switch was made following several years of experiments about racial and ethnic measurement (Smith, 1995; 1997; 2001).

Of course even when there are good reasons to make changes, there is a clear cost involved

with trend analysis either prevented or at least complicated by the changes. Studying societal change will be easier and more reliable if one follows the First Law and maintains constant measurement

Second Law

But sometimes it is necessary to change the measure to maintain constant measurement. In general, the Second Law needs to be invoked when either the measure remains constant in form, but not in substance, or when responses change in a fundamental way. First, the meaning of words can change over time (Smith, 1987). For example, inflation makes items referring to fixed dollars amounts non-comparable over time. Thus, the highly generous Townsend plan of the 1930s that proposed giving $200 per month to each couple over 65 would now sound miserly if the same dollar figure was used. For similar reasons, the income groups in which respondent's income is reported on the GSS have been shifted upwards from a top code of $25,000+ in 1972 to $110,000+ in 2004. Or consider an item asked by Gallup (Which American city do you think has the gayest night life?). Today presumably San Francisco would finish well above its 5th place position in 1954.

Second, terms may become passe and need to be retooled as usage shifts. Until the late 1960s questions on race relations almost universally referred to one racial group as "Negroes". Due to a strong shift in popular usage this term was rapidly replaced in the late 1960s with "Blacks". Then in the late 1980s there was a less sweeping switch from "Blacks" to "African Americans" (Smith, 1992; Smith, Lavrakas, and Schejbal, 1994). Similarly, terms used in surveys in the 1930s for "relief", "war-munitions industry," and "compulsory old-age insurance plan" would have be altered to something like "welfare", "defense industry", and "social security" (Smith, 1987). Sometimes the changes in meaning are essentially linguistic, but other times the meaning of words interacts with history. The item, "Do you expect the United States to fight in another war within the next ten years?", was naturally understood to refer to a conflict like World War II when it was asked in 1946-1950, but after the outbreak of the Korean War it was modified to refer to a "world war" since the type of war referred to was no longer clear (Niemi, Mueller, and Smith, 1989).

But these examples do not mean that changes in meaning are common or typically occur over a short period. If one is careful in selecting widely-used, standard terms and avoiding the latest slang and trendy words, the necessity for revising wordings will be greatly reduced. One example would be questions about marijuana. As the National Institute of Drug Abuse has noted, there are many terms for this drug and terms from the past such as pot, herb, grass, Mary Jane, and reefer mix with more contemporary terms like Aunt Mary, skunk, boom, gangster, kif, and ganja. The last thing one wants to do in asking questions is to utilize the latest (and ever changing) fad words.

Third, due to changing conditions items may become outdated due to the disappearance of old societal features or appearance of new ones. Thus, a NORC item from the 1940s asked about integrated street cars. This question is no longer meaningful because street cars no longer exist. Similarly, a Gallup item in 1937 that asked, "Would you prefer to get national news from a daily newspaper or from a radio?" was outdated by the 1950s with the rise of television and if television had then been added, it would have again become antiquated by the 1990s with the rise of the Internet (Smith, 1987). Obviously asking the original question consistently across time would not

3

provide constant measurement since it originally covered the main news media, but later did not.[2]

Fourth, questions may include statements that are no longer factual over time. The original GSS item on government spending on crime asked about spending to "halt the rising crime rate." But the crime rate did not rise in most years during the last 15 years. As part of a broader reform of the spending battery as a whole, this was handled by having split-ballot wording experiments on the spending items from 1984 onwards (Smith, 1987; Rasinski, 1989). Another example would be items that referred to specific individuals holding political office. So the long Gallup series on presidential job approval must revise its wording when a new leader takes office.

Finally, in other cases the individual measures may remain consistent, but no longer meaningfully measure the underlying domain adequately. There are two examples from intergroup relations that serve as examples. First, on the GSS the original intergroup relations items (or "race relations" items as they were known in the 1970s) only mention Whites and Blacks. With Hispanics now surpassing Blacks in population and Asians also having greatly increased their numbers, the original framework is clearly inadequate and the GSS launched a "Beyond Black and White" initiative in the 1990s that covered other groups both in topical modules and in a revised core. Second, due to growing support for integration and ethno-racial equality, many items became too skewed to differentiate meaningfully. To deal with this old items were replaced with newer items. As part of this transition, in a number of cases new items were asked along side existing items so they could be calibrated before the old items were phased out.

GSS Demographics, 1972-2004

To judge the frequency of changes in measures, consider 10 basic, background variables on the GSS (sex, age, race, marital status, education, occupation, income, labor-force status, religion, and household size). No changes have been made in the measures for six. Sex, age, marital status, education, labor-force status, and household size have followed the First Law and not changed the measure. Four demographics have been changed (race, occupation, income, and religion). Race was fundamentally revised (from RACE to RACECEN1,2,3 + HISPANIC) to reflect revisions in Census practices, a new conceptualization of race, and the changing racial composition of America. Likewise, occupation switched from using the 1970 Census codes to the completely revamped 1980 codes and occupational prestige was remeasured in 1989 to both match up with the new codes and reflect changes in societal evaluations of the standing of occupations (Nakao and Treas, 1994)(e.g. from OCC, INDUSTRY, and PRESTIGE to OCC80, INDUS80, and PRESTGE80). For religion there was no basic shift, but it was refined by additional codes for major religions, Protestant denominations, and Jews. However, this was done without altering the wording of the items and the

---

[2]It is not by happenstance that these two examples involve technology. Technology often changes rapidly and often a new technology will almost completely replace older ones very quickly.

Technological changes also affects survey, data collection itself. For example, telephone surveys largely replaced in-person surveys in the US from the early 1970s to the 1980s. This change is technology, both reflected various societal changes (e.g. spreading telephone coverage, increased use of computers) and complicated the measurement of same by varying the mode.

new codes could be collapsed back into the original similar codes. Finally, one of the unaltered demographics, marital status, changed its social role. It is no longer deemed adequate in and of itself to capture people's living arrangements. The rise of cohabitation means that marital status must be used along with the detailed information from the household enumeration form and a direct item on cohabitation to fully delineate living arrangements.

Summary

But while there are good and necessary reasons to follow the Second Law, departures from the First Law should not be taken lightly and should be done only in a methodologically sound manner. First, changes should be introduced so that the needed refinements are achieved, but the historical trends are also preserved. For example, by the addition of follow-up questions such as cohabitation to marital status, the use of more detailed codes that can be collapsed into simpler original codes as in the case of religion, or having parallel old and new series with split-ballots as for the spending-priorities items. Second, when a new measure replaces an old measure so that a simple and consistent time series can not be maintained, then calibration is desirable. For example, the 1980 occupation and industry codes were switched to after three years of dual coding and the multiple-mentions racial item was adopted after a split-ballot test of the old and new measures. Even with calibration, such switching complicates trend analysis. But it provides a sound methodological basis for splicing trends and extending over time analysis across measurement alterations. Similarly, multiple items making scales of gender roles or intergroup relations have been altered over time with the introduction of new items to the initial battery and then later the dropping of some of the skewed or substantively outdated original items. Of course this only allows scales to be compared over time, for the dropped items there is obviously no continuing time series. In brief, through these and other careful methodological designs, the First and Second Laws can both be respected.[3]

To reliably measure societal change, one must be serious about measurement and careful in how measurement is carried out. One starts by following the First Law. That will be all one needs to do in the vast majority of circumstances. Then one needs to continually consider whether for particular circumstances one needs instead to obey the Second Law. Some criteria for following the Second Law are clear like skewed marginals or certain changes in factual conditions that make old items out-of-date or no longer accurate. But other reasons, such shifts in language use and meaning, are harder to judge. Both because it is often less certain when one should switch from the First to the Second Law and because even a well-founded switch needs to be carefully introduced, one should use split ballots, the adding of additional items, refined codes that can collapse back into original codes, and other techniques to introduce the new. One should never simply delete the old and add the new measures. In sum, the way to study societal change is to follow the First and Second Laws.

---

[3]Multiple-items scales are of course always a good idea to improve reliability and validity. This is especially true of scales designed to measure trends over long periods as individual measures may cease to function properly due to changes factual circumstances, language, skewness, or other factors. Ideally with multiple items there would still be enough items to construct valid and reliable scales even if some items ceased to work well.

<center>References</center>

Duncan, Otis Dudley; Schuman, Howard; and Duncan, Beverly, Social Change in a Metropolitan Community. New York: Russell Sage Foundation, 1973.

Nakao, Keiko and Treas, Judith, "Updating Occupational Prestige and Socioeconomic Scores: How the New Measures Measure Up," Sociological Methodology, 24 (1994), 1-72.

Nie, Norman H.; Verba, Sidney; and Petrocik, John R., The Changing American Voter. Cambridge: Harvard University Press, 1976.

Niemi, Richard G.; Mueller, John; and Smith, Tom W., Trends in Public Opinion: A Compendium of Survey Data. Westport, CT: Greenwood, 1989.

Niemi, Richard G. and Weisberg, Herbert F., eds., Controversies in American Voting Behavior. San Francisco: W.H. Freeman, 1976.

Rasinski, Kenneth A., "The Effect of Question Wording on Public Support for Government Spending," Public Opinion Quarterly, 53 (1989), 388-394.

Schuman, Howard; Steeh, Charlotte; Bobo, Lawrence; and Krysan, Maria, Racial Attitudes in America: Trends and Interpretations. Revised edition. Cambridge: Harvard University Press, 1997.

Smith, Tom W., "The Art of Asking Questions, 1936-1985," Public Opinion Quarterly, 51 (1987a), S95-S108.

Smith, Tom W., "Aspects of Measuring Race: Race by Observation vs. Self-Reports and Multiple Mentions of Ethnicity and Race," GSS Methodological Report No. 93. Chicago: NORC, 2001.

Smith, Tom W., "Changing Racial Labels: From Colored to Negro to Black to African American," Public Opinion Quarterly, 56 (1992), 496-514.

Smith, Tom W., "Is There Real Opinion Change?" International Journal of Public Opinion Research, 6 (1994), 187-203.

Smith, Tom W., "Measuring Race by Observation and Self-Identification," GSS Methodological Report No. 89. Chicago: NORC, 1997.

Smith, Tom W., "Ms. President? A Study of Trends in the Political Role of Women, 1936-1974," in Studies of Social Change Since 1948, Volume II, edited by James A. Davis. Chicago: NORC, 1976.

<center>6</center>

Smith, Tom W., "A Review of the Ethno-Racial Measures of the General Social Survey," GSS Methodological Report No. 85. Chicago: NORC, 1995.

Smith, Tom W., "That Which We Call Welfare by Any Other Name Would Smell Sweeter: An Analysis of the Impact of Question Wording on Response Patterns," Public Opinion Quarterly, 51 (1987b), 75-83.

Smith, Tom W., "Trends in Voluntary Group Membership: Comments on Baumgartner and Walker," American Journal of Political Science, 34 (1990), 646-661.

Smith, Tom W.; Lavrakas, Paul J.; and Schejbal, Judith A., "The Use and Perception of Ethno-Racial Labels: 'African-American' and/or 'Black'," Proceedings of the Annual Research Conference. Washington: Bureau of the Census, 1994.