John Brehm
13 June 1985

Report on coding of economic condiditons series

The three series of economic conditions questions proved to be quite difficult to code. We went through three completely different versions of the master economic conditions codes, and two codings. The final version of the economic condition code reached reasonable levels of intercoder reliability by splitting the code itself into four 'dimensions': Topic, Direction, Actor and Causality.

A great deal of the difficulty that coders had with these question series has to do with the mix of information which respondents offered. The questions themselves asked a stem question ("Would you say that you and your family/group/ the nation are better off or worse off financially than you were a year ago?") followed by "Why do you think things have gotten worse/stayed the same/gotten better?"

Though apparently a simple question, the latter question elicited a range of wholly different responses. Some respondents responded in causal manner — such and such caused/led to/meant the type of change they reported in the stem. This much probably matched the intent of the question. Some respondents responded in terms of feelings — they felt better or worse about some aspect of the economy. Some respondents identified indicators — pointing to a rise in GNP or inflation as the reason for their answer. Respondents often answered in terms of expectations: "inflation didn't go up as much as it could," "my paycheck hasn't increased." The economic conditions code itself had to encompass these four general types of responses.

The problem with the latter question is not confined to different categories of reasons: often the respondents outlined conflicting reasons, no matter what the stem response. For example, respondents often mentioned "inflation is down, but my paycheck

same, causing other effects to change.[2]

Our final approach to coding the economic conditions series divided the answers by four dimensions: Topic, Direction, Actor and Causal Link. "Topic" coded the general category of effect — changes in the unemployment rate, the general state of the economy, changes in particular industries would all be examples of topics. "Direction" coded the direction of change in the topic — if inflation went up, then the topic would be "inflation" and direction would be "increasing." If the respondent mentioned an "Actor" such as "Reagan reduced the inflation rate" then actor coded "Reagan." If the respondent explicitly linked change in one mention to change in another mention, then "Causal Link" recorded that "one mention was causally related to another mention." All four of these dimensions permitted missing data, even when another dimension had a valid code. In other words, a respondent might identify a topic but not an actor or direction of change in the topic.

This approach to coding the economic conditions series coped with the three general difficulties we had with the code. The varying kinds of reasons could be captured in an extensive list of topics, whether "causal" or not. Conflicting reasons could be coded as "causally related" if such a connection were explicit, but coded without regard to the stem in any case. (Coding with regard to the stem put a procrustean constraint on the "rationality" of the respondents' answers). Complex answers, where one reason caused a change in another reason, could be captured with the causal links code.

For each individual dimension, this version of the code gives quite reasonable levels of intercoder reliability. (These estimates were obtained by comparing a second coder's

---

[2]The question itself was not designed to follow rules for causality, and perhaps rightly. One necessary rule for a causal relationship is "concomitant variation" that if x causes y, if x changes so does y. Strictly speaking, if "my income improved" is the effect then "inflation stayed the same" can not be the "cause." Since the question asked why did something stay the same, any causal answer would have to answer in terms of either countervailing forces or other causes which also "stayed the same." A reason why this strict Kantian rule might not be appropriate is that in terms of expectations, something that stays the same might be improving or worsening. (If the economy did not grow, most viewers would say the economy worsened).

to a mention, the subject of the mention should be readily identifiable. There are only a small number of valid codes for the actor dimension, a smaller number of valid codes than any of the other dimensions. If one regards differences between coders as random, the fewer valid actor codes should mean fewer chances for random difference and higher reliability. Conversely, the topic dimension contains far more valid codes than the other dimensions — if random difference between coders accounts for disagreement, the topic dimension should be expected to have a lower degree of reliability than the other dimensions.

From mention to mention, coding of each dimension does not appear to improve in reliability. This is in some ways surprising: for each successive mention, the proportion of missing data codes for mentions should increase. One might expect greater agreement among coders about missing data codes than between more subtle categories. Actually, missing data codes proved to be the least reliable of the categories in each dimension, and accounting for most of the discrepancies between categories. More on the sources of disagreement between coders appears below.

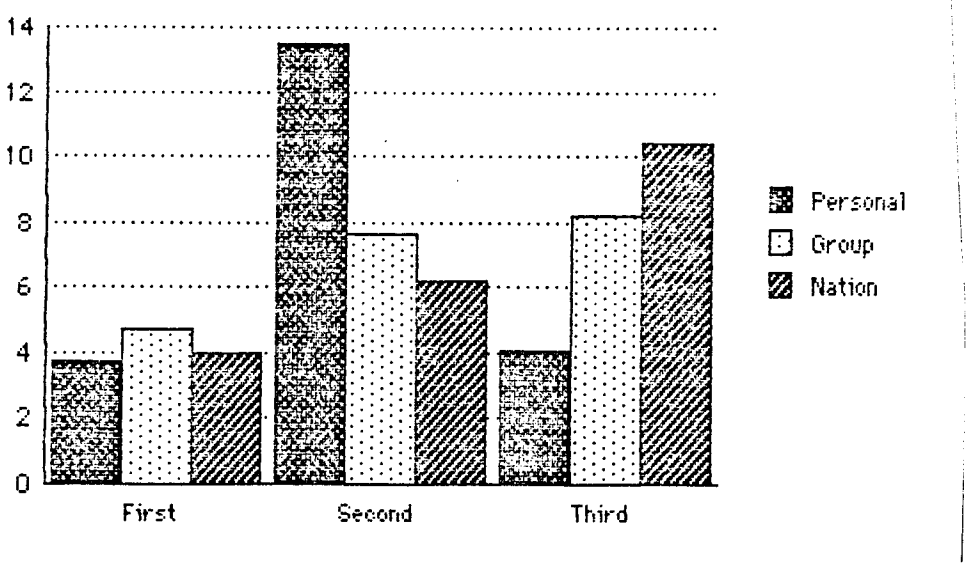
Table 2

Economic conditions code reliability estimates
Full code — five d

Level	Mention	Reliability Estimate
Personal	First	82.2
	Second	80.4
	Third	80.0
Group	First	84.8
	Second	87.8
	Third	93.0
Nation	First	84.8
	Second	87.8
	Third	88.7

Note: Estimates are within +/− 6% (=sqrt(.35/230)*2, 95% confidence level

Figure 3
Histograms of reliability difference



appear to be least independent — gains substantially better reliability in the full five digit code than for the product of the dimensions' reliabilities. Since the topic dimension of the second personal mention has the lowest reliability for any dimension in any code, the source of this difference is apparent.

For the group and nation codes, the dimensions for each subsequent mention appear to be more dependent. (The difference in reliability, Figure 3, rises over the three mentions in both groups and nations codes). The increasing dependence of the dimensions of the groups and nations codes might well be due to discrepancies in the number of mentions different coders attribute to a series of responses. As discussed below, missing data constitute the largest single source of disagreement between coders.

Figure 4 displays pie charts of the source of disagreement between coders. If either coder chose a missing data code, the "source" of disagreement is considered to be "missing data." Otherwise, the "source" of disagreement is "valid" data. The pie charts display the proportion of disagreements for each source, averaged across all mentions and all series. Disagreement in the actor dimension is wholly between missing data and other codes. 89% of the disagreement in the direction dimension is in missing data. Only 29% of the

disagreement in the topic dimension arises from missing data; this proportion is still very large in comparison to the wide number of valid codes in this category.

Figure 5
Source of Topic Disagreement, Percentages

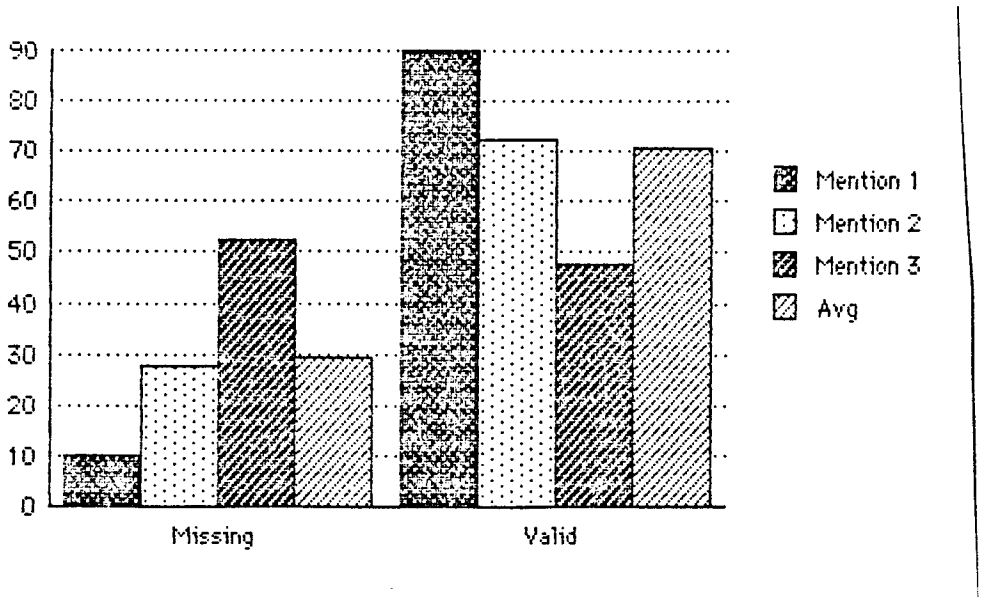


Figure 5 displays the percentage of disagreements in the topic dimension by mentions. It is immediately apparent that disagreement in each subsequent mention is increasingly likely to be about missing data. In the third mentions of the topic dimension (all series), more than half of the disagreements involve missing data codes.

These figures (4 and 5) imply that disagreements between coders arose when the coders disagreed on the number of mentions each response represented. This problem is sensible, although difficult to amend. One reason why coders disagreed on the number of mentions is that the codes (especially in the topic dimension) captured varying degrees of complexity. One code, for example, summarized respondents who observed "no change in real income." Although another respondent may really mean a reference to change in real income, the language that respondent employs may mention "lower income" and "higher taxes." Two coders might well see different number of mentions in the same response. A second reason is that a coder might view a list of mentions as specification of a single