An Experimental Assessment of Extreme Response Style

Abstract

Response style has been a measurement concern for over half a century - the systematic tendency for respondents to choose categories on Likert scales independent of the intended object of measure. Nonetheless, it is not a unanimously held view that response style is meaningful in surveys, or, if it is, what the potential impact of response style bias might be.

This paper first derives measures of extreme response style (ERS) using patterns of extreme response in feeling thermometer items over selected American National Election Studies (ANES) administrations. Using these measures, this study leverages existing random assignment of respondents to identical branching and scaling questions to obtain the experimental effects of item format on ERS in issue opinions. I find that these measures of ERS do exhibit an effect on endpoint selection independent of meaningful content and are strongly associated with political awareness. ERS is found across a number of issues and administrations, with heterogeneous effects by issue area and with inconsistent effects on issue correlations.

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant 144-NL14 to the University of Wisconsin-Madison. The opinions expressed are those of the author and do not represent views of the U.S. Department of Education.

Pär Jason Engle – University of Wisconsin-Madison – pjengle@wisc.edu
Prepared for the American Politics Workshop
May 7, 2012
Introduction

Anyone who has spent any length of time with likert scales has likely noticed, and even been bothered by, a perceived path-dependence in how respondents answer questions. Some seem to gravitate to the middle, the ends, or choose a near endpoint. The observation may seem frustratingly obvious to the casual observer, yet can be difficult to confront analytically. In this paper, I focus on this intuitive threat to both reliable and valid measurement in surveys as response style – how we react to the form of the question itself and not its content. More pointedly, I look at extreme response style (ERS) – the tendency to choose endpoints specifically. In our guild, response set may often be used interchangeably with style. The distinction often made may have some practical meaning: response sets might be helped by good question design, where response styles are tied to the likert scale structure itself.

Response style speaks to both facets of how we might assess the quality of the questions we ask. This evokes the standard, elementary concerns - the quality of the information we obtain from survey questions has been traditionally separated into two very broad concerns of reliability and validity (Linn 2006, Shadish et. al. 2002). Political science has been perennially concerned with the fundamentals of reliability, how it might be accounted for, and its substantive implications when we measure constructs such as beliefs and values (Aachen 1975, Ansolabehere 2008). However, the why of reliable measures has not been of primary concern. Yet, as I contend, if one way of asking a question is more difficult to cognitively navigate, the increased precision can be attributable to a specific cause - a difference in ability, knowledge, or motivation to complete a task. Consequently, to the extent that we might be concerned with the difference in cognitive resource between persons, response style may have important implications for substantive political research. Where there is explanation for measurement error,
then what we assumed to be an otherwise correctable problem injects difficulties into a question’s interpretation, notably: “when measures subject to response style bias are correlated, positive relationships are likely to appear where none exist, genuine positive relationships are likely to be inflated, and genuine negative relationships are likely to be obscured.” (Wells 1961)

This point is not lost in a formal depiction of response, such that, by using simple association, we “…cannot know whether high interitem correlations are attributable to reliable indicators or to unreliable indicators propped up by systematic response error” (Green and Citrin 1994). The bias in question, to confirm intuition, is of particular note when examining relationships between scales.

This two-pronged dilemma is well-established. Yet, despite the thorny implications, the subtle, but meaningful argument is not routinely confronted in studies in public opinion. Public opinion is messy, messages are unclear, and our response to them often muddled. In this environment, it is understandable if the public might create cognitive rules to adjust behavior to match the nature of the ambiguous task at hand: “Response sets become more and more influential to the degree that a respondent is at a loss to answer in terms of specific content… when he or she lacks pertinent knowledge or self-knowledge, is unsure or inaccurate in self-perception, or finds the item ambiguous.” (Messick 1991) By implication, if one were to look for response style, public opinion is as ideal a place to look as we might ask.

The analytical strategy in this paper uses two themes: one of measurement, and another using experimentation to disentangle style from the scales it inhabits. To assess the nature and importance of ERS in particular, I use a more recent measurement strategy that has shown to be effective in isolating ERS and focus on ANES branching and scaling experiments to test for response style and provide indications of response style’s true impact. These two analytical
devices together show that ERS (a) can be isolated, (b) does play a meaningful role in how political issues are interpreted by respondents under strict conditions of random assignment, and (c) can exhibit an impact on valid inference in political issues, particularly through political awareness, the impact of which may be difficult to assess as regularly as we might hope.

Response Style

To start, I borrow a formal definition of general response style that has developed and adopted widely: the tendency to select specific categories on a likert scale independent of the object of measure (Paulhus 1991) – whether these are midpoints, endpoints, or responses that signal agreement with a question (acquiescence).¹ Survey response styles have been theoretical and statistical issue dating to Cronbach (1946, 1954), and have most pressingly concerned marketing researchers who have attempted to account for irregularities that appear in different populations that stem from a common cultural survey approach (Marin et al 1992, Clarke 2001, Chen et al 1995). The observation among these researchers was that a culture’s social norms may not only drive attitudes toward an object, but also come out in which survey options appear to be most attractive (Hui and Triandis 1989). Because of this threat, it is now standard to consider the effect of response style when equating survey responses across cultural context (Cheung and Rensvold 2000). Yet, even within countries, individuals have importantly different approaches to surveys with findings that have been empirically linked to variables of gender, education, income, age of the respondent, and race (Greenleaf 1992a, Greenleaf1992b, Bachman and O’Malley 1984, DeJong et al 2008). And with these associations, potential explanations for response styles have also proliferated. They have included general cultural modesty, personality traits (intolerance of ambiguity or a desire for certainty), anxiety, and attempts to reduce the

¹ We might also view these strategies as habitual or automatic response (Bargh 1997), specifically induced by a question’s structure and not its content.
cognitive demand for scales with many options – which effectively limit responses to a few select options (Baumbartner and Steenkamp 2001, Berg and Collier 1953, Lewis 1955). For many of these explanations, strong associations have been documented, but these have not all translated to consistent rationales that link response style to the social and psychological traits that analysts evoke to describe it.

Of the rationales for response style that have emerged from these previous studies, one compelling cognitive argument for its existence and impact is emphasized here – the mental simplification of larger scales into smaller components, or optimization behavior. Survey optimizers are more thorough in their approach to evaluation, whereby they are less likely to mentally “give up” and choose a scale option that is more cognitively convenient, but less reflective of true preference (Krosnick 1999). Incomplete optimizers, conversely, have been called satisficers - respondents who, at some stage of question interpretation, consideration, and decision, choose a more expedient path to response, bypassing fully reasoned evaluation. It is this mechanism that leads to optimization as a probable, and leading, candidate behind response style. When evaluating issues, we might expect those with less political information will have less ability or motivation to respond fully, will tend to use simplification strategies, and consequently exhibit systematically less reliable response.

In another way, the same process has been described in terms of one or more secondary question answering strategies. A respondent may either immediately know a correct answer, or revert to a close approximation if a correct answer is not immediately found, creating branches of possible response that pit alternate answers against each other, a tradeoff that is colored (driven) by a question’s format. Some recent work in political science has given credence to guessing as a

---

2 This may also relate to differing tendencies toward meta-judgment (Bassili 1996), engaging the likert scales of attitude in two parts, both in direction and in strength.
contaminant in knowledge scales (Mondak and Anderson 2008) and some analytical models have shown that there is benefit to viewing multiple choice response guessing as “nested” in certain knowledge problems (Suh and Bolt 2010). How an attitude question is cognitively processed, via optimization, may be highly susceptible to the presentation of alternatives. As a result, we might expect the least aware and the least knowledgeable to be drawn to secondary strategies, an observation that is not incompatible with findings that associate, for example, low socio-economic status variables with response style.

Previous psychometric work notwithstanding, researchers may not be convinced that attitudes might be somehow “incorrect.” It may also be valid to question what the likely impact of this incorrectness might be, if any. Lingering concerns remain about whether response style even strictly exists (Messick and Jackson 1991) or is best considered purely random. When we talk about survey response in substantive political science, it is often assumed that it contains a naturally unreliable component (competing considerations) and/or a random “sampling” component.

By its nature and previous research indicates that public opinion continues to be a difficult area in which to adjudicate between response precision and response meaning, whether due to an uncertain process of mental sampling or a competitive messaging environment (Zaller 1992). Issue evaluation is complex. Rarely, if ever, is there a correct answer, and correct opinion is difficult to define and evaluate. Added to this, we may not give issues the attention they deserve and know precious little about politics (Delli-Carpini and Keeter 1996). In the issue realm, candidates may also have some disincentive to clarify the public’s information (Franklin 1991). The implicit argument, one that uses complex issue evaluation as its evidence, leads to the conclusion that an answer to a question – as asked – is important in isolation, not necessarily
as part of an underlying latent construct (Weijters 2006). For the unconvinced, *The Great Response Style Myth* (Rorer 1965) serves as one flagship of this argument\(^3\) – that is, unreliable response is a product of natural variation in our response process that genuinely reflects the opinion of the respondent and how the respondent wishes to be presented. The certainty and strength of an attitude as evidenced in the survey are all of true attitude itself and inseparable from the response, where a style of response actually conveys substantive meaning. That there might be true, and consequently untrue, attitude is not a standard in view, even if it is useful to consider threats of stylistic response as untrue (Schaeffer and Presser 2003). Response style, by its nature, is a different and often troubling concept for this view of survey response. Simply by being both systematic and independent of attitude (non-substantive and non-random), response style does not fit well with our classical interpretation of measurement error in opinion. For this analysis, it is necessary to adopt the modal viewpoint where response style is considered response that is not true to the intended construct (Bohrnstedt 1983) - a more reflexive, automatic response disposition which is not attributable to substantive dimensions of interest.

**Measuring Extreme Response Style – The Base Measurement Model and Anchor Items**

Of the several identified response styles, I limit the focus to extreme response style (ERS), or a tendency to select scale endpoints independent of the object of measure. First, I do so because ERS accounts for a larger portion of stylistic variance in past studies than other styles (Baumgartner and Steenkamp 2001). Along with acquiescent response (ARS), ERS accounts for most response style effects, which, on the whole, have been shown to be a non-trivial portion of total response variance, accounting for 5-20% in marketing research scales (DeJong et. al. 2008). Second, unlike acquiescent response, extreme response can be more credibly isolated

\(^3\) In the context of Acquiescent Response Style (ARS)
analytically through psychometric methods developed and fine-tuned over the last several decades. That is, what we know about ERS might be known with greater confidence.

To measure ERS, I appeal to some more recent modeling work which shares a common, intuitive logic. One question is not enough to identify a tendency for ERS, but multiple questions might establish an underlying, latent endpoint response pattern. A thermometer rating (0-100) of “0” and “100” for the sitting president in one question cannot be interpreted alone as anything other than intense presidential reaction. For the same respondent, it is plausible that they may also feel strongly toward the sitting vice president, so another “0” or “100” rating is not a pattern that is particularly informative about response style, but probably indicative of a highly correlated set of positive or negative attitudes. Ideally, a battery of random questions would be optimal to identify extreme response independently (Greenleaf 1992b), yet in practice, these questions are prohibitively costly and therefore absent in nearly all large scale surveys.

Absent an ideal, a different form of response accounting is necessary. If a simple tally of endpoint selections contains bias due to content (partisanship or ideology), one way to circumvent this problem is to explicitly account for content and analyze any “residual” endpoint selection that is present. Such is the measurement strategy here. While other formal methods can be devised differently to similar ends (as in structural equations, see Moors 2003, Moors 2004, Kieruj and Moors 2010), this study uses an alternate approach with a multi-dimensional nominal item response model (Bock 1972) in order to simultaneously identify and separate a confounding meaningful, political dimension from extreme response (Bolt and Johnson 2009).

To accomplish both measures (extreme response and political evaluation), ANES feeling thermometer items are used to form a measurement baseline. Embedded in these baseline items
are attitudes toward political figures or explicitly political groups\(^4\) which together tap a latent pattern of attitude from left to right on the political spectrum (Brady 1990, Brady and Sniderman 1985).\(^5\) Because thermometer items have a uniquely extended format, they are also uniquely suited to detect stylistic variance, being ambiguous and focusing the respondent on what is likely an unfamiliar type of subjective evaluation, exacerbated by an open scale of 101 possible points from 0-100 (or more aptly - nine verbal prompts which I use in these analyses\(^6\)). Respondents are allowed a great deal of latitude to communicate information about their opinion, drawing out what we might view as style, a particularly strong format for investigating survey approach from what is an otherwise dubious measurement construct. In order to isolate response from political attitude, each thermometer item is allowed to associate with each other both politically and through endpoint selection. By separating these latent traits in the same model, the dominant portion of political variance might be held constant, allowing a clearer picture of ERS as a latent trait of interest.

Formally then, the specification is as follows, where \( P(X_{ikj} | \theta_{1j}, \theta_{2j}) \) is the probability of selecting a response category given a substantive latent variable, \( \theta_{1j} \), and an extreme response latent variable, \( \theta_{2j} \). The probability of category response \((k)\) is an unordered logistic function in items \((i)\) for respondent \((j)\), with a slope parameter for category selection in each dimension \((a_{ik1}, a_{ik2})\) and a common intercept \((c_{ik})\). For identification, the sum of the slope and intercept

\(^4\) Only thermometer items that carry a referendum on political objects are considered (e.g. Democrats, Republicans, The President, etc.). Purely demographic group thermometers are not included.
\(^5\) To account for the ideal point characteristic of these data, the nominal response model (Bock 1972) will allow categories at the extreme to be selected less often as respondents become more liberal or conservative. For the dimensional characteristics of these data, some evidence suggests that a single dimension may adequately capture common variance in attitude due to institutional arrangement, a simplifying assumption that is made here (Jackman and Sniderman 2002).
\(^6\) Ranges are collapsed into categories representing nine verbal prompts: 0=Very cold or unfavorable feeling, 15=Quite cold or unfavorable feeling, 30=Fairly cold or unfavorable feeling, 40=A bit more cold or unfavorable feeling, 50=No feeling at all, 60=A bit more warm and favorable than cold feeling, 70=Fairly warm or favorable feeling, 85=Quite warm or favorable feeling, 100=Very warm or favorable feeling.
parameters over each item is constrained to zero. Non-responses, no opinion categories, and refusals are also constrained to exhibit no influence in their relationship to political attitude and ERS \( a_{i(NR)} = 0, a_{i(NR)} = 0 \), which provides the interpretation of ERS conditioned on selecting a point in the scale.

\[
P(X_{ik} | \theta_{1j}, \theta_{2j}) = \frac{\exp(a_{jk1} \theta_{1j} + a_{jk1} \theta_{2j} + c_{ik})}{\exp(a_{ih1} \theta_{1j} + a_{ih1} \theta_{2j} + c_{ih})}
\]

This measurement model for each survey administration rests on an anchor set of thermometer items \((i)\), which vary by availability in each administration year. Again, thermometer items that ask about political figures and groups should define, for each category, a relatively warm or cool interpretation – from left to right, warmer toward conservative political objects and colder toward liberal ones. This would be the dimension of interest in a substantive scaling application, an estimate of continuous political attitude across the survey population \((\theta_{1j})\).

Added to the primary partisan political dimension, and separate from it, is a measure of some residual endpoint selection propensity which is used to estimate ERS, a second continuous latent variable that measures the tendency toward endpoint selection \((\theta_{2j})\). Respondents who have a greater tendency to choose endpoints, from right to left, will be increasingly drawn to the ends of the thermometer scale – separate from their political attitudes. In order to be most explicit, further confirmatory restrictions on ERS identify what we might expect of extreme

\(^7\text{LATENT GOLD (Vermunt and Magdison 2004) is used in model estimation using a Marginal Maximum Likelihood procedure (Vermunt and Magdison 2005). While Bayesian inference is often a standard in political science, there is some question whether the classic Bayesian estimation strategy is appropriate for experimental situations (Kaplan 2010).}\)

\(^8\text{These latent dimensions, like oblique factor rotation, are also allowed to correlate with one another.}\)
response and isolate its meaning further. These restrictions should also lessen the potential impact of other associations that might be present between these thermometer items and any residual substantive meaning. The first set of constraints forces the endpoints of the thermometer ratings (“0” and “100”) to an equal, strictly increasing propensity toward ERS \((a_{t1} = a_{t9} = 1.0)\). The probability of selecting an endpoint will be symmetric and not dependent on the direction of response. That is, it should not matter if a response is a warm or cold extreme, only that it is an extreme selection. The second set of constraints conditions the response between thermometer items. Where extreme response exists, we might initially expect it to vary in the same way, identically between the thermometer items \((a_{l1} = a_{-l1}, a_{l2} = a_{-l2}, \ldots, a_{lk} = a_{-lk})\).9

The net result, formally, is an estimate a set of workable constructs – one explicitly partisan/political \((\theta_{1j})\), and the other defining a residual endpoint selection tendency \((\theta_{2j})\), narrowed in definition by constraint, to reflect endpoint selection. Of these, the primary latent variable of interest that is carried forward is the ERS score \((\theta_{2j})\).

**Random Assignment**

With this measurement model in hand, what is named ERS is simply something residual that correlates with endpoint selection. If the preceding measure of ERS is true to the concept and practically meaningful, it will first predict propensity of endpoint selection in other likert scales and second, be a product of format alone. The expectation is that, among similar items that should induce ERS (five categories or more, Kieruj and Moors 2010), the preceding measure of

---

9 This is partly set up to accomplish what Weijters (2006) finds to be an invariance of response style across items, even though the eventual results challenge this assumption.
ERS should increase as residual endpoint selection increases. For respondents who are higher on the ERS scale, any new items these respondents might encounter should similarly draw out more frequent endpoint selection.

While helpful, an external validation that shows respondents pick endpoints in other scales is insufficient to identify the ERS variable as response style. The residual ‘something’ in this measure must also be present both in an extended likert scale format and absent when this response structure is taken away. The critique of response style research by non-experimental methods highlights the particular weakness to non-experimental inference - that an anchor scale of ERS may be related erroneously to any target item. Response style studies may often fail to adequately isolate pure response style, and they may do so because we are unable to measure ERS the way we intend, purging residual meaning. Notably, we might never be completely sure we have done so adequately. The main doubt is a classic and valid one: are we measuring what we think we are, or is real response getting in the way? If not, the conclusion might be that response style is less important than we suppose, if it exists at all independently of measurement artifact. Previous successes in detecting response style may lead the skeptic to consider response style illusory, meaningless, or more uncommon than our measures suggest. The higher the common variance between base items that are not explicitly accounted for, the greater is the potential for bias in the ERS measure. This measurement critique is bolstered by how ERS has not behaved completely consistently in many respects, by covariate association, across ERS research, with both the impact and character of response style varying, sometimes considerably, from study to study (Baumgartner and Steenkamp 2001).

To combat this concern, I use existing random assignment in ANES administrations using branching and scaling formats. Because the contamination we anticipate is based on
content, if we are able to hold this content constant by design, a clearer, more convincing picture of response style should emerge as a result.

Fortunately, the ANES has a history of conducting randomized methodological experiments that use branching and scaling items. Specifically, I chose different administrations of experimental branching versus scaling which randomly assign identical questions to branching and scaling conditions, following the reliability study of Krosnick and Berent (1993). Where equivalent issue items are available by time period and the character of study differs in structure - target questions, sample size, differing base thermometer items, and the variable lag between the base and target questions. In these studies, seven point scaled items are randomly assigned to respondents in either the scaling or equivalent branching form.

- The first is the 1988-89 ANES panel re-interview (pilot) data with 614 respondents on the pilot (of n=2,040 for the 1988 base year) and three issue questions, matched with responses from the 1988 survey.
- The second is the 1990-92 panel re-interview data with two issue questions (n=1,980).
- The third is the 1982 ANES Methods Comparison Project with 2,461 respondents with three focal issue questions.

\[\text{10} \text{ Where branching and scaling items are equivalent in category number and category meaning. Notably, partisan identification differs in how conditions are labeled and more recent experiments in branching and scaling (ANES 2000 and ANES 2008) utilize different numbers of categories and different questions respectively.}\]

\[\text{11} \text{ The administration profiles, base, and focal questions are presented in A1-A3 in the Appendix.}\]
To illustrate the comparability of the questions, an example of a branching and scaling is considered in the following (Table 1) and in the accompanying diagram (Figure 1). The scaled defense question is provided on the left and the branching equivalent is provided on the right.

### Table 1. Randomly Assigned Branching and Scaling Example

<table>
<thead>
<tr>
<th>Condition #1: Scaling Defense (1989 ANES Pilot) Question 7331 (E1)</th>
<th>Condition #2: Branching Defense (1989 ANES Pilot) Question Summary 7341 (E5x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>There has been a lot of debate recently about defense spending. Some people believe that the U.S. should spend a lot less money on defense. Suppose these people are at one end of a seven-point scale, at point number 1. Others feel that the U.S. should spend a lot more on defense. Suppose these people are at the other end of the scale -- at point number 7. And, of course, other people have opinions somewhere in between, at points 2, 3, 4, 5, and 6. Where would you place yourself on this scale, (remembering that point number 1 is a lot less spending on defense and point number 7 is a lot more spending on defense)?</td>
<td>E5. There has been a lot of debate recently about defense spending. Do you think the U.S. should spend less money on defense, more money on defense, or continue spending about the same amount on defense?</td>
</tr>
<tr>
<td>E5a. Would you say the U.S. should spend a lot less or a little less on defense?</td>
<td>--</td>
</tr>
<tr>
<td>E5b. Would you lean toward spending less on defense or more on defense?</td>
<td>--</td>
</tr>
<tr>
<td>E5c. Would you say the U.S. should spend a lot more or a little more on defense?</td>
<td>--</td>
</tr>
</tbody>
</table>

In the scaled item, the question allows a single, straightforward selection ranging from “a lot less” defense spending to “a lot more” (1-7) with the intermediate categories left for the respondent to communicate some gradation of opinion. In the branching, the same question is divided into two parts. The first part signals the direction of preference from three options, and the follow-up question subdivides based on the initial selection that looks to ascertain how much defense should be increased or decreased, or if, by selecting the midpoint, if a small preference exists one way or the other. When put together, the responses to the branching items are assembled into a summary that contains the same two verbal anchors at the ends, which is largely comparable to the scaling, yet more strictly defined at the intermediate options (Figure 1).

---

12 Question text for the other scaling items is provided in the Appendix.
Random assignment between questions similar to the exemplar in Table 1 and Figure 1 will insure that more or less equivalent groups will receive both formats and that only a single facet is being manipulated – the format of the item. Additionally, randomly assigning these branching and scaling items, conditional on response, the estimates will not have the theoretical problems of self-selection. This reduces the comparison to a very manageable one: ERS, by definition, should be present in extended likert scale items, and absent in the equivalent branching items.

Putting both the measurement and experimental pieces together, the ERS measure is combined and the branching and scaling items are evaluated in the same model. Figure 2 sketches the measurement model along the top, common variance between a set of base thermometer items (X₁ through Xᵢ). For illustration, we might assume that these items are completely unrelated except through response style. Along the side box are the focal items, where
the experimental effect of question format is targeted, $X_{(Branch)}$ (which should not contain ERS) and scaling items $X_{(Scale)}$ (which should). These items, as experimental targets, are tested in turn, with the statistical test of interest comparing endpoint responses over $X_{(Branch)}$ and $X_{(Scale)}$.

Figure 2 is simplified in order to illustrate the simultaneous measurement and testing strategy based on equation (1) in terms of an idealized path diagram. The anchor thermometer items are influenced by a latent extreme response. To test whether, and how much, ERS influences the focal items, Figure 2 adds a path to both of the identical experimental and focal items.

**Figure 2. Experimental Design Path Model: Focal and Base (Anchor) Items**
Each of these paths to the focal items contains multiple parameter estimates of category selection. Analytically, this adds both branching and scaling questions to the original model from equation (1), estimated together.

Within each of the paths in Figure 1 from ERS to $X_{(Branch)}$ and $X_{(Scale)}$ are seven separate possible slope coefficients for ERS, one for each scale category. To test the impact of ERS on category selection, the endpoints $[a_{(Branch)1}, a_{(Branch)7}]$ and $[a_{(Scale)1}, a_{(Scale)7}]$ are allowed to vary freely. Additionally, to force the symmetry consistent with the definition of extreme response, category slopes are constrained to equality within $X_{(Branch)}$ and $X_{(Scale)}$ ($a_1 = a_7$), along with each of the categories that are an equivalent distance from extremity ($a_1 = a_7, a_2 = a_6, a_3 = a_5$), identifying four coefficients for a seven point scale - extreme response ($a_1$), near extreme ($a_2$), near midpoint ($a_3$), and midpoint ($a_4$). This design allows intermediate responses to convey information and, more importantly, pins down the essence of what we expect in selecting an extreme response. This structure also lends itself to a directly testable hypothesis for each focal item with the same resulting response categorization and content. Of primary interest, the coefficient tests of extreme response in particular:

$$H_0: a_{(Scale)1} = a_{(Branch)1}$$

(endpoint difference in category selection by ERS)

If there is a difference in endpoint selection, we expect to see a difference on the category selection propensities between scaling and branching. In particular, we might expect the endpoint selection propensity to be significantly higher in the scaling condition. Second, if residual endpoint selection is a concern, we might expect to see significant positive effects or ERS in the
scaling condition \( (a_{Scale})_1 > 0 \), and under ideal measurement conditions, in the branching condition, we might expect to see little or no difference in endpoint selection \( (a_{Branch})_1 = 0 \).

**Experimental Results**

The first question in these items is of average endpoint selection effect. Do endpoint selections’ effect on ERS appear to be significant when scaling is introduced? On nearly all issue questions, the answer is “yes.” The results, comparing the branching condition to the scaling condition are detailed in Table 1 for each class of response from the extreme response, near extreme, near midpoint, to midpoint.

Importantly, all three of these conditions for detecting ERS in these items show some support – difference between items, the positive difference in scaling items, and the insignificance of branching. The strongest indicator is the consistency of ERS in its impact on category selection between branching and scaling conditions. In every case, the endpoint selection propensity in the scaled condition is significantly greater than in the branched condition. Less strong, but still prominent, with the exception of gun control, each endpoint selection is significantly greater than zero. Finally, with the exception of the questions in the 1990-92 Panel measurement, the branching items show no detectable difference from zero. In sum, ERS shows largely what it is supposed to show, an increase in extreme response in scaling versus its branching counterpart, largely positive endpoint selection in scaling items, and branching impact on ERS mostly indistinguishable from zero.
Table 1. Slope Coefficients for Category Selection due to ERS \( a \in \{a_1, a_2, a_3, a_4\} \)

<table>
<thead>
<tr>
<th>Question and Difference Estimates (p-value – scaling endpoint mean selection greater)</th>
<th>Format</th>
<th>( a_1 )</th>
<th>( a_2 )</th>
<th>( a_3 )</th>
<th>( a_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Defense</td>
<td>Branching</td>
<td>-0.10</td>
<td>-0.07</td>
<td>-0.18</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>Scaling</td>
<td>0.36</td>
<td>0.02</td>
<td>-0.29</td>
<td>-0.16</td>
</tr>
<tr>
<td>Central America</td>
<td>Branching</td>
<td>0.09</td>
<td>-0.20</td>
<td>-0.18</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>Scaling</td>
<td>0.39</td>
<td>-0.05</td>
<td>-0.15</td>
<td>-0.37</td>
</tr>
<tr>
<td>Gun Control</td>
<td>Branching</td>
<td>-0.16</td>
<td>-0.17</td>
<td>-0.18</td>
<td>1.02</td>
</tr>
<tr>
<td></td>
<td>Scaling</td>
<td>0.36</td>
<td>-0.19</td>
<td>-0.07</td>
<td>0.23</td>
</tr>
<tr>
<td>Limits on Foreign Imports</td>
<td>Branching</td>
<td>0.18</td>
<td>-0.21</td>
<td>0.00</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>Scaling</td>
<td>0.61</td>
<td>-0.23</td>
<td>-0.36</td>
<td>-0.06</td>
</tr>
<tr>
<td>Sanctions on South Africa</td>
<td>Branching</td>
<td>0.29</td>
<td>-0.03</td>
<td>-0.27</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>Scaling</td>
<td>0.62</td>
<td>-0.11</td>
<td>-0.37</td>
<td>-0.28</td>
</tr>
<tr>
<td>Defense</td>
<td>Branching</td>
<td>0.12</td>
<td>0.13</td>
<td>-0.05</td>
<td>-0.40</td>
</tr>
<tr>
<td></td>
<td>Scaling</td>
<td>0.37</td>
<td>0.06</td>
<td>-0.26</td>
<td>-0.34</td>
</tr>
<tr>
<td>Good job / Standard of Living</td>
<td>Branching</td>
<td>0.08</td>
<td>-0.01</td>
<td>0.12</td>
<td>-0.39</td>
</tr>
<tr>
<td></td>
<td>Scaling</td>
<td>0.56</td>
<td>-0.11</td>
<td>-0.32</td>
<td>-0.56</td>
</tr>
</tbody>
</table>

Bolded coefficients indicate significant difference from zero at 0.05.

The other response categories provide some important nuance to these results. In the branching items, while endpoints are not typically associated with ERS, some response differences deserve some note. Specifically in the 1988-89 ANES Pilot, the three questions appear to replace endpoint selection with midpoint selection, a property which is not present in the other administrations. Several candidate explanations are possible; notably differences in the character of the base measure, the character of the population that may respond to a follow-on pilot wave, or the lack of availability of a non-response option for a random subset of respondents. These differences reverse in the 1982 MCP, which may suggest that midpoint response (possibly as no opinion) or general ambivalence may also be detected in these base measures.
The slope differences in Table 1 serve as the test of endpoint selection tendencies, but may not put the marginal effects of this selection into focus. The practical impact is unclear from these direct tests. Table 2 and Figures 3 through 9 show what these slopes mean for endpoint selection probability among the focal items in each administration. In these charts, the lines indicate the estimated probability of endpoint selection for branching and scaling formats over ERS, a standard normal variable, holding constant the primary dimension of political evaluation. The table and figures show how frequently we might expect endpoints to be selected, accounting for, but ignoring the political dimension, and looking at endpoint selection solely as a function of response style.

<table>
<thead>
<tr>
<th>Extreme Response Scale</th>
<th>1.5</th>
<th>1.0</th>
<th>0.5</th>
<th>0.0</th>
<th>0.5</th>
<th>1.0</th>
<th>1.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1989 Pilot Defense</td>
<td>5.6%</td>
<td>7.3%</td>
<td>9.5%</td>
<td>12.3%</td>
<td>15.6%</td>
<td>19.5%</td>
<td>24.2%</td>
</tr>
<tr>
<td>1989 Pilot Central America</td>
<td>6.7%</td>
<td>8.8%</td>
<td>11.4%</td>
<td>14.6%</td>
<td>18.5%</td>
<td>23.0%</td>
<td>28.2%</td>
</tr>
<tr>
<td>1989 Pilot Central Gun Control</td>
<td>34.6%</td>
<td>37.4%</td>
<td>40.1%</td>
<td>42.9%</td>
<td>45.5%</td>
<td>48.1%</td>
<td>50.5%</td>
</tr>
<tr>
<td>1992 Panel Foreign Imports</td>
<td>6.1%</td>
<td>9.0%</td>
<td>13.1%</td>
<td>18.6%</td>
<td>25.6%</td>
<td>34.0%</td>
<td>43.5%</td>
</tr>
<tr>
<td>1992 Panel South Africa</td>
<td>4.7%</td>
<td>7.2%</td>
<td>10.8%</td>
<td>15.8%</td>
<td>22.6%</td>
<td>31.2%</td>
<td>41.2%</td>
</tr>
<tr>
<td>1982 MCP Defense</td>
<td>6.3%</td>
<td>8.4%</td>
<td>11.0%</td>
<td>14.3%</td>
<td>18.3%</td>
<td>22.9%</td>
<td>28.3%</td>
</tr>
<tr>
<td>1982 MCP Jobs</td>
<td>7.1%</td>
<td>10.3%</td>
<td>14.7%</td>
<td>20.4%</td>
<td>27.7%</td>
<td>36.3%</td>
<td>45.8%</td>
</tr>
</tbody>
</table>

Overall, as ERS increases, as we might expect from the slope coefficients, the probability of selecting an endpoint on the scales increases. For more typical ranges of response style (-0.5 to 0.5), an change in ERS represents a 6-13% increased likelihood of residual endpoint selection, a range of roughly 50-100% greater, moving from a modestly “non-extreme” responder to a respondent who more typically exhibits this behavior.
Two observations are of note in examining marginal effects. First, endpoints in both branching and scaling are not chosen with the same frequency. Interestingly, branching items typically increase overall levels of endpoint selection. Respondents, on the whole, appear liberated to choose options at the end in a branching format that they otherwise would not. This said, the branching questions typically do not move appreciably with ERS, with the exception of the 1990-1992 ANES Panel. Even though it appears that the branching format does better at inducing endpoint selection overall, it is mostly independent of ERS. Those who are extreme in branching questions, then, appear genuinely extreme. Second, the greatest marginal effects typically occur where response style is the greatest. The difference in the probability of selecting an endpoint is small from very low ERS to average levels of ERS, but the difference becomes typically greater moving from average to high levels of ERS.

Put together, the marginal effect plots confirm that the effects observed in this model are occurring, and occurring within normal range of the data. Content held constant, with the exception of Gun Control,¹³ I find significant differences in endpoint selection tendency due to item format and that these differences are meaningful in relevant ranges of response.

The (Non-Random) Character of ERS and its Consequence

The previous sets of results show how the branching and scaling constructs alone account for an average increase in selecting endpoints among scaling items. In itself, this will indicate systematic variance due to ERS. However, there may be little interpretive meaning behind this trait without demonstrating some influential component to this selection. Otherwise, this increased variance may not be indicative of more serious substantive bias, it may merely indicate

¹³ While the endpoints of the branching and scaling items are significantly different from each other, gun control stands out in both the absolute frequency with which the endpoints are selected and the milder effect of extreme response across the relevant range of data.
increased variance that might be adequately dealt with as measurement error. To provide some character to ERS, I have tied ERS to traditional covariates of response, adding the more compelling explanations for stylistic response – political knowledge, the effect of which should indicate optimizing and satisficing behavior.

In this case, how influential are levels of political information on ERS? To look at this specific impact, I allow the respondent’s level of political information to vary in its impact on ERS. To this, I use the ANES interviewer rating of political information – a 1-5 rating from very informed to very uninformed by interviewer designation. The addition to the formal specification is a regression of ERS on levels of political information, performed within the base measurement models in (1) and performed simultaneously with estimation of the measurement model:

\[
\theta_{2j} = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4
\]

In each administration, political information exhibits a strong association with ERS. From the most informed to the least, extreme response increases roughly three-quarters of a standard deviation, but the effects do not appear to be meaningful uniformly throughout the range of information. Generally, from very high to average information, there is only one significant difference (average information on the 1988 ANES), which changes more dramatically as political information decreases below average.

Table 3. Adding Political Information as a Determinant of ERS

<table>
<thead>
<tr>
<th>Question/Coefficient (SE)</th>
<th>Fairly High</th>
<th>Average</th>
<th>Fairly Low</th>
<th>Very Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>1988-89 ANES + Pilot</td>
<td>0.110</td>
<td>0.337</td>
<td>0.459</td>
<td>0.893</td>
</tr>
<tr>
<td></td>
<td>(0.099)</td>
<td>(0.096)</td>
<td>(0.102)</td>
<td>(0.117)</td>
</tr>
<tr>
<td>1990-92 ANES + Panel</td>
<td>-0.171</td>
<td>0.141</td>
<td>0.443</td>
<td>0.663</td>
</tr>
</tbody>
</table>
The point is a modest one, but important. Modest, because given what we know about stylistic variance, an association with political knowledge was a minimal requirement for any face validity. Important, however, because the significance of the result is magnified in conjunction with the previous marginal effects estimates. To argue that ERS makes a difference in how endpoints are selected, question by question, is meaningful to the degree that comparing relationships may overstate positive ones and understate negative ones (where two variables both contain response style artifacts). The addition of political awareness expands the implication. In one interpretation, individuals with low awareness will be particularly susceptible to inferential problems, and in another, the association between awareness and issue response will also contain bias. *Whereas political awareness is a predictor of substantive issue response, low political awareness will be more strongly associated with issue evaluation simply due to response artifact.*

One intuitive way to develop an understanding of this potential impact is by simulating for responders with varying extreme response. In particular, given how an item behaves substantively, when ERS is taken into account, how much do correlations between variables change? Positive substantive relationships between two variables should see inflated association under conditions of ERS.

The basic simulations draw from a component of a hypothetical substantive underlying construct \((\theta_{1j} \sim N(0,1))\) and from an Extreme Response distribution \((\theta_{2j} \sim N(0,1))\) and reconstruct the response \((X_{ik} | \theta_{1j}, \theta_{2j})\) for each respondent according to equation (1) using item parameters from the scaled experimental items. The correlations in Table 4 represent simulated
seven point scales based on how they associate through feeling thermometer items. If we look at these through a general partisan lens for the time periods, the associations roughly conform to what we might expect, namely a modest partisan issue divide on defense, very little association by gun control, and by contrast, a more substantial association by jobs. To augment the range of correlation, the strength of the issues has been increased incrementally in the final simulations. The correlations are depressed somewhat from the true correlations given the choice of display by a straight Pearson correlation, which may water down the effect as opposed to its polychoric alternative measure. As such, this is a demonstration of the naive estimate.

<table>
<thead>
<tr>
<th>Question – ERS Quintile</th>
<th>Mean Correlations of Variables with Differing levels of ERS (n=5,000 – 30 sets of correlations)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bottom</td>
</tr>
<tr>
<td>1989 Pilot Defense</td>
<td>0.067</td>
</tr>
<tr>
<td>1989 Pilot Central America</td>
<td>0.047</td>
</tr>
<tr>
<td>1989 Pilot Central Gun Control</td>
<td>0.004</td>
</tr>
<tr>
<td>1992 Panel Foreign Imports</td>
<td>0.002</td>
</tr>
<tr>
<td>1992 Panel South Africa</td>
<td>0.047</td>
</tr>
<tr>
<td>1982 MCP Defense</td>
<td>0.068</td>
</tr>
<tr>
<td>1982 MCP Jobs</td>
<td>0.125</td>
</tr>
<tr>
<td>Increased r - 1982 MCP Jobs</td>
<td>0.364</td>
</tr>
<tr>
<td>Increased r - 1982 MCP Jobs</td>
<td>0.655</td>
</tr>
</tbody>
</table>

From these simulations, we can see that Pearson correlations increase mostly - and modestly. Where an association exists between the latent feeling thermometer dimension and the specific issue, the association is exacerbated by response artifact. This association, however, is not entirely as consistent as we might expect. Different questions are influenced more or less by ERS and varying item response characteristics (notably item skew) may increase or decrease the effect.
By inspecting the effect of political awareness on response style, consistent with the optimization/satisficing hypothesis, there is a meaningful increase in the propensity for the lowest informed in the population to exhibit extreme response. And this propensity can increase correlations among these populations by as much as 3-4%. Again, this propensity is largely independent of the target of measure – attitudes toward prominent political people and groups, and, using the previous experimental results, appears generally in scaling items while not in branching items.

**Discussion/Conclusion**

There are lingering questions about response style in public opinion, questions that have roots in a description of how we respond to surveys, what our thought processes are as we respond, and how we translate political opinions onto survey questions.

Normatively, some doubt persists about response style, especially when the legitimacy of a response is questioned through this type of validity study. Some resistance remains to the notion that opinions and attitudes might be somehow incorrect. Empirically, results in previous ERS research and their accompanying and varied explanations may reinforce these doubts about the nature of style. Depending on measurement context and content, associations in response style studies may appear or disappear depending on the study. The preceding analysis is first confirmatory, attempting to reconfirm the existence of response style in surveys and examining primary drivers of ERS. By using the most ambiguous construct in the ANES and the strength of experimentation, several implications arise within the context of scale construction and issue opinion.
First, these results consistently find that extreme response style is present and meaningful in the target items - a conclusion is reinforced by measurement construction and experimental design. If allowed by the format of the question, respondents do, in fact, exhibit extreme response tendencies. To exemplify how much this may typically affect response, I have demonstrated the parameters with marginal impact of extreme response in these items. With the exception of the gun control issue, I find that ERS typically increases endpoint selection probabilities by 50-100% over the heart of the distribution from -0.5 to 0.5. The magnitudes are additionally suggestive because ERS is only one of several classes of response style that have been detected which should also contribute to systematic, yet non-substantive variance in scales.

However, ERS alone may not indicate particular bias if ERS is distributed relatively randomly through a population. These measures indicate that ERS is decidedly non-random. Notably, in these administrations, ERS exhibits differential impact by a respondent’s level of political information. In this way, ERS appears to validate the optimization hypothesis and provide a basis from which we might identify, not only measurement error, but how the meaning of measurement error may contribute to inferential problems. The results here indicate, for issue items, this type of bias is not merely possible in any given item, but highly likely, both in prevailing theory of item response and empirically.

Second, we know that branching has proven more reliable than scaling formats on surveys. This study shows one meaningful reason why this might be the case. In particular, this analysis finds not only that branching questions are more effective question formats for yielding precise measurements of a target construct, but one reason for their effectiveness is by effectively eliminating extreme response style. Branching questions are more reliable because they mitigate ERS, and by doing so, they are additionally more accurate representations of the target construct.
The last observation is substantive. How much of a bias does this represent and for whom? The result shows, in scales where ERS is present, the bias can be substantial. Notably, Philip Converse may be slightly more correct than his detractors suppose. As political knowledge decreases, the effect of arbitrary category selection increases in a substantively meaningful way, inducing association that is partially illusory.

With these results come new questions that have yet to be addressed that deserve more thorough treatment. Going forward, the main benefit of stylistic inquiry is to inform measurement practice itself, which leads to an unqualified recommendation to change ANES standard from scaling to branching formats where feasible. And where continuity is desirable, reasonable efforts to equate legacy scaling items would result in better measurement and retention of time series integrity. Second, expanding randomized question format studies may be of benefit. Beyond measurement, however, we might look more explicitly at the substance of ERS and examine the difference of impact between items.

From a practical measurement standpoint, using randomly assigned questions as an arbiter of what works, the possibility exists that the measures themselves might be extended and tested in productive ways. An interesting side-effect of being able to detect ERS in randomly assigned formats is that it might assist researchers in the practical endeavor of discerning which methods best isolate style and when. Several variants of modeling response style exist, from the very simple methods of counting category selection frequency to more involved methods that separate respondents by latent class. The simple contribution of random assignment of item format holds some promise as an independent adjudicator of differing measurement strategies. That contamination does, in fact, exist in some of these measures speaks to the utility of being
able to best model response style beyond a simple tally of extremity – a strategy that is, at the very least, unstable, and potentially highly misleading about the causes and consequences of response style in surveys.

Substantively, assessing the magnitude of ERS is an important, interpretive question that will extend beyond this simple exposition of marginal effects. How might we discern whether these effects are large or small and when? The preliminary, but insufficient, answer is that effects will be highly dependent on context. The results suggest that the impact may be considered small overall, but important if we are concerned with the smaller effects on the margin that may carry larger substantive significance. It is somewhat glib to remark that partisan identification explains the vast majority of vote choice, but it is often the smaller component that determines election outcomes. But the point is still valid: a small bias in context may loom large. This will be especially true where likert scales are used intensively and political information is important in the analysis. With this in mind, if we are to assemble an inventory of political variables that may carry deciding weight, ERS, and response style more generally, is a prime candidate for further scrutiny under the right conditions. This study begins to explore what those conditions might be.

As such, if response style is attributable to political information, it is not unreasonable to expect more extremity as complexity and ambiguity of questions increase. This difference in impact of ERS between questions provides another avenue for further inquiry. The set of focal questions here are limited, but might provide some basis for which to examine how respondents vary in question answering strategy between issues. If one issue stands out as potentially instructive, it is gun control. Respondents on gun control, aside from being particularly extreme on the issue’s substance, also do not appear to use endpoint selection strategies to a large degree. The implication might be that gun control is, on average, not a terribly difficult issue for
respondents to comprehend. In this analysis, it would be speculation to attribute a cause from this one case, but the overall lesson and opportunity for research is inescapable. Not only might we be able to measure and correct for response bias, but potentially find substantive use for it.
References


Figures 3-9. Probability Charts – Endpoint Selection by ERS Scale – Branching v Scaling
Gun Control
Probabilities of Endpoint Response

Probabilities of Endpoint Response

Limits on Foreign Imports
Probabilities of Endpoint Response
Sanctions on South Africa
Probabilities of Endpoint Response

Defense Spending
Probabilities of Endpoint Response
Appendix

Table A1. Administrations Used in Analysis

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Focal Issue Items</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Content</td>
<td>Defense, Central American Aid, Gun Control</td>
<td>Foreign Imports, South Africa Sanctions</td>
<td>Defense, Jobs and Standard of Livings</td>
</tr>
<tr>
<td>Base Items</td>
<td>24 Lagged Thermometer (1988)</td>
<td>16 Concurrent Thermometer (1990)</td>
<td>7 Concurrent Thermometer</td>
</tr>
<tr>
<td>n</td>
<td>2,040 (Base) 614 (Focal)</td>
<td>1,980 (Base and Focal)</td>
<td>2,416 (Base and Focal)</td>
</tr>
</tbody>
</table>
A2. Scaling Question Text Used in the Studies

731 El. There has been a lot of debate recently about defense spending. Some people believe that the U.S. should spend a lot less money on defense. Suppose these people are at one end of a seven-point scale, at point number 1. Others feel that the U.S. should spend a lot more on defense. Suppose these people are at the other end of the scale -- at point number 7. And, of course, other people have opinions somewhere in between, at points 2, 3, 4, 5, and 6. Where would you place yourself on this scale, (remembering that point number 1 is a lot less spending on defense and point number 7 is a lot more spending on defense)?

732 E2. Some people think the United States should become a lot less involved in the internal affairs of Central American countries. Suppose these people are at one end of a seven-point scale, at point number 1. Others believe that the U.S. should become a lot more involved in this part of the world. Suppose these people are at the other end of the scale -- at point number 7. And, of course, other people have opinions somewhere in between, at points 2, 3, 4, 5, and 6. Where would you place yourself on this scale, (remembering that point number 1 is a lot less involved and point number 7 is a lot more involved in the internal affairs of Central American countries.)

733 E3. In recent years, there has been increasing public debate about gun control laws. Some people feel the laws covering the sale of firearms should be a lot less strict. Suppose these people are at one end of a 7-point scale, at point number 1. Others feel that the laws covering the sale of firearms should be a lot more strict. Suppose these people are at the other end of the scale -- at point number 7. (Of course, other people have opinions somewhere in between, at points 2, 3, 4, 5, and 6.) Where would You place yourself on this scale (remembering that point number 1 is a lot less strict gun control laws and point number 7 is a lot more strict gun control laws)?

VAR 900432 Q.K1. (RB, P.9) Some people favor increasing limits on foreign imports a lot in order to protect American jobs. Suppose these people are at one end of the scale, at point number 1. Others favor decreasing the limits a lot in order to lower consumer prices and help American exports. Suppose these people are at the other end of the scale, at point number 7. And, of course, other people have opinions in between at points 2, 3, 4, 5, or 6. Where would you place yourself on this scale, or haven't you thought much about this?

VAR 900433 Q.K2. (RB, P.10) Some people feel that the economic sanctions against South Africa should be decreased a lot in light of changes in the treatment of blacks that have taken place there recently. (Suppose these people are at one end of the scale, at point number 1.) Other people feel that sanctions should be increased a lot in order to pressure the government to make further changes. (Suppose these people are at the other end at point number 7. And, of course, other people have opinions somewhere in between at points 2, 3, 4, 5, or 6.) Where would you place yourself on this scale, or haven't you thought much about this?

VAR MC0410 TP J3. Some people believe that we should spend much less money for defense. Suppose these people are at one end of the scale at point number 1. Others feel that defense spending should be greatly increased. Suppose these people are at the other end, at point 7. And,
of course, some other people have opinions somewhere in between at points 2, 3, 4, 5, or 6. Where would you place yourself on this scale, or haven't you thought much about this?

VAR MC0418 TP J4. Some people feel that the government in Washington should make every effort to improve the social and economic position of blacks and other minority groups. (Suppose these people are at one end of the scale at point number 1.) Others feel that the government should not make any special effort to help minorities because they should help themselves. (Suppose these people are at the other end at point 7.) And, of course, some other people have opinions somewhere in between (at points 2, 3, 4, 5, or 6). Where would you place yourself on this scale or haven't you thought much about this?

VAR MC0428 TP Jj. Some people feel the government in Washington should see to it that every person has a job and a good standard of living. Others think the government should just let each person get ahead on his own. Where would you place yourself on this scale or haven't you thought much about this?

1. GOVERNMENT SEE TO A JOB AND GOOD STANDARD OF LIVING
2.
3.
4.
5.
6.
7. GOVERNMENT LET EACH PERSON GET AHEAD ON OWN
Table A3. List of Base Thermometer Items for ERS

<table>
<thead>
<tr>
<th>1988-1989 ANES + Pilot (Base $i=24$)</th>
<th>1990-1992 ANES Panel (Base $i=16$)</th>
<th>1982 ANES Methods Comparison (Base $i=7$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Democratic Party</td>
<td>George HW Bush</td>
<td>Jimmy Carter</td>
</tr>
<tr>
<td>Republican Party</td>
<td>Mario Cuomo</td>
<td>Ronald Reagan</td>
</tr>
<tr>
<td>Big Business</td>
<td>Dan Quayle</td>
<td>Edward Kennedy</td>
</tr>
<tr>
<td>Labor Unions</td>
<td>Ronald Reagan</td>
<td>John Anderson</td>
</tr>
<tr>
<td>Liberals</td>
<td>Jesse Jackson</td>
<td>George HW Bush</td>
</tr>
<tr>
<td>Conservatives</td>
<td>Dem Party</td>
<td>Walter Mondale</td>
</tr>
<tr>
<td>Civil Rights Leaders</td>
<td>Rep Party</td>
<td>Dem Party</td>
</tr>
<tr>
<td>People On Welfare</td>
<td>Supporters Abortion</td>
<td></td>
</tr>
<tr>
<td>Congress</td>
<td>Conservatives</td>
<td></td>
</tr>
<tr>
<td>Environmentalists</td>
<td>Labor Unions</td>
<td></td>
</tr>
<tr>
<td>Anti-abortionists</td>
<td>Women’s Movement</td>
<td></td>
</tr>
<tr>
<td>Federal Govt</td>
<td>People On Welfare</td>
<td></td>
</tr>
<tr>
<td>Fundamentalists</td>
<td>People Pro Environment</td>
<td></td>
</tr>
<tr>
<td>Feminists</td>
<td>Liberals</td>
<td></td>
</tr>
<tr>
<td>Michael Dukakis</td>
<td>Poor People</td>
<td></td>
</tr>
<tr>
<td>Lloyd Bentsen</td>
<td></td>
<td></td>
</tr>
<tr>
<td>George HW Bush</td>
<td>Abortion Opponents</td>
<td></td>
</tr>
<tr>
<td>Dan Quayle</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ronald Reagan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kennedy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dole</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evangelical Groups</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jesse Jackson</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A4. Latent GOLD Syntax Summary, Repeated for all Focal Items

variables
dependent
// base items
  vmc0092o nominal, vmc0093o nominal,
  vmc0094o nominal, vmc0095o nominal, vmc0117o nominal,
  vmc0118o nominal, vmc0120o nominal,
// focal item
  vmc0418o nominal, vmc0552o nominal;
// option independent for political awareness codes 2-5 (pa2…pa5)
  latent
    CFactor1 continuous,
    CFactor2 continuous,
    Cluster nominal 1;
equations
  (1) CFactor1 ; // ERS
  (1) CFactor2 ; // Political Attitude
  CFactor1 <-> CFactor2 ; // Allows factors to correlate
  Cluster <- 1;
// Base Items (vary by administration availability)
  vmc0092o <- 1 + Cluster + (b1)CFactor1 + (a1)CFactor2;
  vmc0093o <- 1 + Cluster + (b1)CFactor1 + (a2)CFactor2;
  …
// Focal Items
  vmc0418o <- 1 + Cluster + (b3)CFactor1 + (a3)CFactor2;
  vmc0552o <- 1 + Cluster + (b4)CFactor1 + (a3)CFactor2;
// set non-response equal to 0 for all a and b
// first category is zero (0) [x,1]
  a1[1,1]=0;
  a2[1,1]=0;
  …
  b1[1,1]=0; b3[1,1]=0; b4[1,1]=0;
// fix extremity on ERS base variables
  b1[1,2]=1; b1[1,9]=1;
// For testing focal item, set endpoints to equality and force symmetry from midpoint
  b3[1,2]=b3[1,7];
  b4[1,2]=b4[1,7];
  b3[1,3]=b3[1,6];
  b4[1,3]=b4[1,6];
  b3[1,4]=b3[1,5];
  b4[1,4]=b4[1,5];