Response Style in the Political Survey

By

Pär Jason Engle

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Political Science)

at the

UNIVERSITY OF WISCONSIN-MADISON

2016

Date of final oral examination: 12/14/2015

The dissertation is approved by the following members of the Final Oral Committee:

Byron E. Shafer, Hawkins Chair of Political Science, Political Science
John Witte, Professor, Political Science
David L. Weimer, Professor, Political Science
Alexander Tahk, Assistant Professor, Political Science
Daniel Bolt, Professor, Educational Psychology

Contents

Acknowledgements 1
Introduction 2
Why Response Style? 5
Response Style: A Definition and a Debate 9
Ideology: Responding to Ambiguous Constructs 13
The Role of Reasoning and the Political Survey Response 16
Measuring Response Style 19
Formalizing Intuition: Parameterizing and Identifying Response Style 22
Organization of Thesis 28
Chapter 1. An Experimental Assessment of Extreme Response Style 30
Response Style 32
Measuring ERS 36
Random Assignment 40
Experimental Results 47
The (Non-Random) Character of ERS and its Consequence 55
Discussion and Conclusion 59
Chapter 2. The Effect of Extremity on Behavior: The Bias of Response Style 69
When Responses are Missing: Survey Fidelity 69
Extreme Response, Extreme Ideologues, and Behavior 71
Measuring ERS 74
Strong Opinion and Strong Response 78
Strong Opinion, Strong Response, and Behavior 81
Extreme Response Bias: Strong Partisanship as a Predictor of Voting 85
Discussion/Conclusion 90
Chapter 3. ERS and Political Issues: Extreme Response Style (ERS) Differences in the American National Election Study (1992-2008) 97
The Meaning of Extreme Response Style 99
Issues, Political Evaluation, and Response Style 104
Measuring ERS 108
Measurement Results 114
Validity of ERS 114
Endpoint Selection 117

Exemplar: Democratic Party 1992 118
ERS Issue Item Differences: Over Time and Between Items 125
Spending and Services 125
Defense 129
Abortion 132
Traditional Family and Moral Standards 135
ERS Thermometer Variance: Feelings about Selected Political Objects 138
Liberals, Conservatives, Democrats, and Republicans 141
Cultural Feeling Thermometer Items, Gays/Lesbians and Christian Fundamentalists 144
Democratic and Republican Presidential Candidates 145
Discussion/Conclusion 147
Chapter 4. Acquiescent Response and Institutional Confidence: When Partisanship Informs Political Dissent 158
Acquiescent Response Style 163
Measuring ARS 169
Assessing the Validity of Acquiescent Response 176
Acquiescent Response and Confidence in Political Institutions 180
Discussion/Conclusion 190
Chapter 5. Constraint, Moderation, and Civic Education Instruction 201
Constraint and Moderation 204
Educational Practice and Political Outcomes 208
Measuring Constraint and Moderation with Social Welfare Ideology 210
Results 217
Discussion/Conclusion 224
Chapter 6. Political Discourse and Religion: Belief Constraint and Moderation 226
Using Constraint and Moderation to Evaluate Political Discourse 229
Measuring Constraint and Moderation 230
Formulating Expectations of Religious Constraint and Moderation 237
Denominational Constraint and Moderation on the Ideological Right and Left 244
Devotional Constraint /Moderation on the Ideological Right and Left 247
Discussion/Conclusion 248
Conclusion 255

Acknowledgements

I owe a debt of gratitude to many whom I may not be able to directly repay. The contributions here are incomplete, and indicate rough net probability added of positive outcome without quantification. First, I need to acknowledge the dedication of my advisor, Byron Shafer, the line of support that kept me going when I did not want it and whose advice I often refused to my own detriment. My family deserves an emotional assist: Mark, Sharon, Mary, Sunshine, Eleanor, Kepler, Nichole, and the Hawkins/Bass clan. And, at critical times, what passes for Providence intervened in the night.

The night is dark.
Let our fears of the darkness of the world and of our own lives
rest in you.

The night is quiet.
Let the quietness of your peace enfold us,
all dear to us,
and all who have no peace.

The night heralds the dawn.
Let us look expectantly to a new day,
new joys,
new possibilities.

Many thanks to others at the Interdisciplinary Training Program for the Education Sciences (Mary Tedeschi, Molly Wesling, Adam Gamoran, and the Social Capital Project), my Political Science cohort of excellence, John Witte, Deb MacFarlane, and the support network at the Wisconsin State Department of Public Instruction's Office of Educational Accountability.

Introduction

Anyone who has asked young children about their attitudes expects to hear superlatives. As far as I know, anything having to do with cartoons is really, really good, and liver is still very, very bad. Even more mundane objects in a young child's experience have the characteristic of very good or very bad. And, while this is cute, it would not be a profound observation. Yet research has shown, as we grow up, we do not entirely lose the vestiges of this kind of primitive attitude evaluation. We retain them in ways that are curiously regular, even as we might add experience and attention to the revelation of our own preferences. In response to questions about our attitude, we can search our thoughts and feelings and provide an answer as true to our opinions as we can. However, even simple questions hold ambiguity that we confront the best we are able, and when satisfied with a reasonable response, we unconsciously abandon further reason and proceed to judgment. The cognitive imprints that resist controlled judgment, despite even our best efforts to subvert them, nevertheless find their way onto the closed-end survey, a staple of behavioral political science research in the modern era.

The role of heuristics, judgmental shortcuts in how we evaluate political objects, has attracted more attention in the last two decades in political science. The public lacks adequate information necessary to make decisions about political objects, and therefore relies on shortcuts, symbolic mental markers from which the public might base a decision that is correctly reflective of their preferences. This is hardly a controversial or even recent observation, but the use of heuristics have been more precisely articulated and thoroughly explored recently as phenomena about which we might gather evidence. Finding mental evidence is daunting, however. The American experience of race is an exemplar of how social

context may inhibit the task of obtaining true opinion, where expressing certain racial attitudes are taboo, but under the right circumstances might be drawn out. Respondents do not reliably reveal their thought processes openly, yet it is precisely their thought processes that interest us. This project opens the door to examining a new form of evidence that examines how we think and ostensibly how we think about politics.

It is in this spirit that I examine response style: *a tendency for the respondent to choose categories on a survey rating scale independent of the object of measure*. One response style cases is of primary concern and, with acquiescent response, accounts for the bulk of response style impact: extreme response style (ERS). The parallel to heuristics in decision-making is direct: when confronted with a choice of what rating category to endorse, we rely on shortcuts to tell us how to answer. And we do this, often, without giving any conscious thought to our decision.

A significant portion of variance in widely used market research scales has been identified as an artifact of stylistic response (generally 5 to 20 percent [Baumgartner and Steenkamp 2001]), and there is ample reason to believe that political science metrics will exhibit similar, if not more pronounced stylistic tendencies. The format of a question, in its wording, phrasing or structure may not be clear to all respondents or the complexity or ambiguous nature of the content may be daunting. Any confusion may be compounded by whether a respondent is deeply or only peripherally aware of politics, which is only exacerbated by the way candidates and parties compete, often by redefining political messaging. To measure response style, I make liberal use a factor analytic analogue that is closely related to standard ideal point estimation−item response theory (IRT). The IRT variants I use come closer to those models used in standardized attitudinal and educational measurement, tailored to finding

patterns of response style in multiple items. Traditionally, these models find practical use in constructing standardized tests. As ability, knowledge, or any target trait increases, patterns of response reveal the latent characteristic with which we are concerned. Most commonly this is academic achievement, yet here I use attitudinal latent models to focus on patterns of response that contain latent characteristics of *response style*.

The benefits of this dissertation and this research line come primarily in three ways. First, overtly acknowledging response as a structural characteristic of our understandings of surveys cannot be underestimated; it has been dealt with in isolation, yet we have not demonstrated how response style can, and does, affect results on the modern behavioral survey, and how otherwise innocuous response might be used practically to understand politics. Second, understanding response style contributes to the continuing survey effort to minimize the noise and bias in survey measures and to identify how and when response style may meaningfully impact substantive results. Last, and most substantively, we can learn about what it might mean to have a response style, try to understand what it measures by where it operates and finds meaning, and thereby get a glimpse of how people might think about politics and make choices around political issues.

In this thesis, I summarize a few key questions that relate to response style and form models and constraints appropriate to answering them. They are:

(a) Does response style exist? Is it possible to verify response style experimentally?

(b) What are some of the potential biases we might expect to see?

(c) If there is overt meaning in response style, what might it be, and what might it tell us about political questions, political opinion, and formative learning in the electorate?

Why Response Style?

The content of an individual's survey response is the starting point in survey research: how many citizens are actively engaged in politics, how many support government involvement in health care, how warm do people feel toward Jimmy Carter? What are people saying about what they believe, what are their personal judgments, and how do they react to change? All of these are vital questions, and large bodies of literature are rightly devoted to what messages come out of raw, atomic instances of expressed opinion. These form the questions and answers that are read, interpreted, and form the basis for some understanding of why people think and do the things they do in the political world. Practically, these are the meat and potatoes of political survey research, serving as a vehicle for public political input, public opinion accounting, opinion research, and providing actionable information in the business of shaping public policy. Researchers use surveys to make inferences about political behavior and practitioners often use them to delineate the boundaries of governance according to the image of what can and cannot be done under some constraint in the larger polity.

Yet this thesis ventures to a different and subterranean area of content, measuring not what people say, but how they say it. We can view response behavior in a couple different ways. First, we can see aberrant response foibles as a nuisance to correct in question construction. From this perspective, respondents do not answer questions in the way we expect due to a pre-existing, constant, and personal habit or disposition. Much research in survey construction is devoted to how to write questions that avoid these response problems (Schaeffer and Presser 2003), and as a result, offending items are becoming less frequent. From this perspective, response style is behavior that is content-irrelevant, and by cleansing this impurity from the data, we might get a clearer picture of the substance we really want to measure, which

leads us to question in which circumstances response style is the biggest threat to validity and how might we counter this threat. The answers, many already available, will have implications for survey design and will suggest corrective ideas for response bias in popular questions where it is a concern−the what, where, and how much of response style in surveys that are attributable to respondent traits and question quality. Yet many contours of response style have not been examined directly, or explored at all in commonly used political surveys. Not to say that the science of constructing good survey questions has been neglected overall. On the contrary, work has largely been targeted toward prospectively fixing “bad" questions or assessing which questions might contain potential response bias and why.

The second perspective on response style is as a contextual phenomenon. From this perspective, respondents do not answer questions in the way we expect because of a disposition that is exacerbated by ambiguous or charged content. The pivotal concern is whether response style is a disposition (something that the individual possesses), or applied based on situation (something that is induced by the environment). I expect that both the dispositional and situational characteristics of response operate and that these tendencies live in an individual and are brought out by the question and the content. To the extent that response style lives with the individual, response style will reveal characteristics of the individual and individual reasoning. Research to date has revealed some indication that cognitive ability and engagement of the respondent (Baumgartner and Steenkamp 2001, Krosnick 1999) are key determinants of response style. To the extent response style is situational, however, we can gain specific insight into both the respondent and the political objects they respond to. How much of response is a product of a person’s own disposition, and how much is disposition acting in context-characteristics of the polity or the subject matter itself?

Despite its potential, this line of inquiry has not filtered into political science in a sustained way, whether because of a perceived lack of importance, a problem of measurement, or a prevailing line of theory that respects, maybe to a fault, the ability of the individual to calculate his or her own attitudes. Response style can be perceived, understandably, as a relatively arcane threat to validity, and as a cause or consequence, the measurement of response styles has been generally less than completely robust. The immediate practical benefit lives in prospectively identifying and correcting for response bias and demonstrating both the magnitude and direction of bias, a practical endeavor of response accounting. A less immediate benefit, and the aim of this larger project, is to build a case for its practical and theoretical importance, identify when it may be important, and suggest how extracting style from response might be useful.

The following sections sketch, and make a case for response style as systematic and interactively revealed behavior. Much of this literature lives outside of the discipline, and by necessity reaches into some unfamiliar territory for some. Yet the essential issues veer directly back to core concerns of political science. Formatively, this lives in the more seminal works of Zaller and Krosnick (Zaller 1992, Zaller 1996, Krosnick and Schuman 1988, Holbrooke et al. 2001, Krosnick 1991, Krosnick 1999), where a picture of survey response is painted as a substantive trait that exhibits regularities that meaningfully interact with content. It also includes a swath of research in political psychology that has specifically addressed the dimensionality of surveys or what is identified as components of political reasoning, certainty, "meta-attitude," and "automaticity" (Bassili1996, Bargh 1997). Reconciling the insights from this tradition with that of marketing research, psychology, and testing is a prerequisite discussion.

Again, I look at the survey as *reflecting more than what people say, but how they say it*. We can peer through the survey as a window into political attitude. What we see helps us make conclusions about what is happening in the mind of subjects. However, the window may not provide an entirely accurate picture. The window may not provide clarity, as we have discovered through the study of survey measurement error and reliability. To increase clarity, then, is to filter out statistical noise through multivariate measurement. The hallmark case of ideological constraint in public opinion is, perhaps, the standard bearer in this regard (Aachen 1975, Converse 1964), where fuzzy areas of opinion might be honed by adding variables to a latent measure to increase its clarity of meaning. Yet, in addition to clarity issues due to measurement error, we might be more concerned with bias: the window on our world may actually distort the objects that we see on the other side. Our survey window on public issues may not tell us correctly what the public thinks.

These are the distortions we might expect from habitual response or response style: a tendency for the respondent to endorse categories on a survey rating scale independent of the object of measure. Whether response style bias is meaningful in any specific case will depend on two necessary conditions: the strength of style as a predictor of variance in any question or set of questions, and the direction of response style bias by a confounding context or set of individual attributes. If response style bias is truly random then it will be part and parcel of measurement error in a survey. However, previous research has shown that this is not the case, notably that response style is a meaningful component of ambiguous scales, and shows significant systematic variation between individuals and between countries. *How people say what they say can distort what it is we think they are saying*.

Response style also provides a window for researchers to assess the automatic character of responses more generally. Decisions about issue endorsement are made through surveys in an environment of limited information, ambiguity, a high degree of abstraction, and often in a structured choice environment that invites a certain degree of arbitrariness. We also know that the institutional environment of politics in America combined with competitive messaging does not necessarily help in making nuanced decisions easy for the public, often increasing the ambiguity of issue evaluation. The substantive implication of response style can be a referendum on decision strategies under uncertainty, whether decisions to endorse issues are made consciously or are automatic in nature, and whether it makes a difference. This will involve looking at who responds automatically and on what issues, not just as an indicator of bias, but as a way to assess how people think about politics and what implications this may have for understanding American democratic politics. When people respond extremely or acquiescently, they may not be telling us about their issue preferences, but response behavior does tell us something. *How people say what they say can inform us about individuals, the issues they respond to, and about larger issues of politics.*

Response Style: A Definition and a Debate

Response style refers to any tendency to respond to a survey in a way that does not reflect the target of a question (Jackson 1958), historically called response sets (Cronbach 1946, 1950). Response styles are systematic tendencies for respondents to choose endpoints, midpoints, or socially desirable (acquiescent) responses from an array of fixed survey options. Paulhus (1991) and others have extended the Paulhus definition of response style as a systematic tendency to choose options on a rating scale independent of the construct being measured, a definition that persists in psychometric survey measurement. This accepted

definition highlights some unreconciled disciplinary divisions when it comes to survey response style. Marketing research, cross-cultural inquiries, and psychology have isolated and examined these styles as a source of systematic bias in attitude scales, often reporting the source of variation in bias by culture, gender, income, education, and race among other characteristics of respondents.

Marketing research in particular has developed more disciplinary interest in response style as a scale contaminant. By necessity, marketing studies put a premium on finding "true attitudes" as an effective tool for gauging consumer sentiment. Mainstream public opinion research, on the other hand, has traditionally viewed survey responses as more or less sacrosanct, legitimate response worthy of interest and study; that is, issues of style are not viewed traditionally as separable from survey content. The current lines of research in public opinion and political psychology implicitly harbor a concept of response error−from the perspective of a person choosing categories on a rating scale, for example−that models internal conflict over a decision whose error is random (when asked a question, objects are pulled, or sampled, from memory) or modeled in non-random ways that are more idiosyncratic (when asked a question, objects are sampled in a skewed way based upon context) (Zaller 1992). The idiosyncrasies of the survey are then attributed mainly to dynamics of how we receive information, accept it, and then retrieve it for use. Respondents, as a result, should find their true attitude and faithfully represent what we expect of them: their preferences, preference strength, and learning. The resulting error in attitude translation comes from the process of true deliberation. In contrast to this mental model, in the traditional psychometric point of view, respondents learn how to approach a social situation of the survey, process the aim of the target

questions, and translate their cultural biases, social norms, and learned habitual approaches onto a survey as response.

To make progress, we need to alter some of our assumptions to specifically acknowledge that true attitude is indeed separable from untrue, habitual response (Bohrnstedt 1983) and that this untrue response occurs because the political respondent uses stylistic strategies, unconscious responses to the survey situation and content. The most compelling cognitive rationale for this notion of separability of person and context comes as a breakdown in respondent optimization (Krosnick 1991). Opinion scholars largely assume that when respondents approach survey questions, they faithfully and fully engage in the scientifically proscribed activity. Ideally, when a subject responds to a survey, she optimizes by deriving the intent of the question, searching available memory for attitudes and associations, reaching a judgment based on those attitudes, and translating that judgment onto an available survey response (Krosnick 1999). A respondent uses these shortcuts to decrease the cognitive load of a survey, much like the respondent chooses opinions with insufficient information. When a respondent is motivated to respond truly and has the wherewithal to complete this process, our surveys are minimally affected by measurement error or response bias. However, when response is complicated at any stage of optimization, as the cognitive load of the question increases, subjects will have the tendency to fall back on simplifying constructs. Ambiguity, content that evokes conflicted individual experience, and answer choices that do not match well with a respondent's subjective evaluation will prompt what has come to be called “satisficing” response behavior.

Two of the most influential response cases, extremity and acquiescence, form the basis for the cognitive inquiry in this dissertation. Extremity is the endorsement of endpoints, outer

ranges of a Likert scale, and acquiescence indicates a tendency toward general agreement toward the valence of a survey stimulus. While there may be some distinct tendencies in each of these styles, under Krosnick's optimization hypothesis and in summaries of the response-style literature (Baumgartner and Steenkamp 2001), prevailing research shows that acquiescent and extreme response are correlated (r=0.57) with each other.

Consensus in response style research is strong on the point of separability, but perhaps overstated in the exact mechanism by which it may operate. In multiple cross-cultural studies, the individual tendency toward extreme and acquiescent response has been muddied by the introduction of strong cultural variation in response style, between and within countries (Cheung 2000, Bachman 1984, Marin et al 1992, Johnson 2005). The result is contextual heterogeneity that weakens the argument that the individual respondent specifically is the driving force behind response style (Meisenberg 2008). I take these unreconciled results, not to overemphasize the dichotomy between the individual and the collective in response style, but to turn to the alternate approach to viewing response style as interactive. And by examining the context of response style, it is possible to start to untangle social understanding, question content, and personal attribute.

Out of this effort, alternate hypotheses emerge that emphasize political reasoning and contextual stylistic response. One such hypothesis I entertain is the differential application of controlled reasoning among questions (Chapter 3). One possibility that allows for both context and personal attributes to contribute is justification (Chapter 4). Absent sufficient reason for a response, we might concoct a response that is not strictly true to our attitude, but at least justifiable: “while the statement I endorse in the survey is not my true attitude, it is acceptable enough.”

Ideology: Responding to Ambiguous Constructs

To lend intuition to the effect of response style, consider a survey of 1,066 respondents in the 2004 ANES. The respondents apply themselves and others to the iconic question concerning ideology on a one to seven scale:

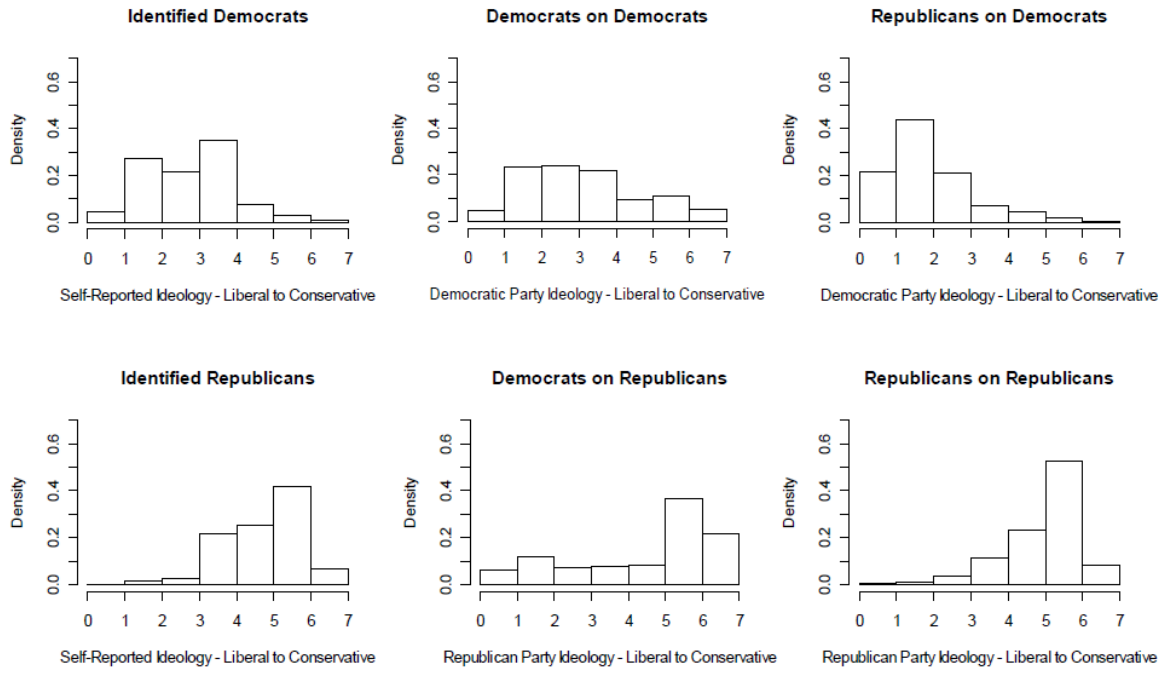
> We hear a lot of talk these days about liberals and conservatives. When it comes to politics, do you usually think of yourself as extremely liberal, liberal, slightly liberal, moderate or middle of the road, slightly conservative, extremely conservative, or haven't you thought much about this? (1-7)

Assume these respondents have a true preference toward answering this question and their preferences are distributed such that some choose extreme values, some choose simply to identify themselves as liberal or conservative, some choose to be slightly liberal or conservative, and the remainder select the middle label. Consider also that these same individuals hold a propensity to respond to a survey in a more or less extreme way, such that, say, half the respondents will be more likely to veer outward, and the other half will be more adherent to the question's target. Comparing those with an outward tendency and adherents to the construct, the outward tendency will have necessarily higher variance. If this response style is completely random, there is no immediate cause for great concern, since a correction for reliability or measurement error should suffice in making the variable useful for inference. However, if response style is associated with some other politically relevant measures (education level, race, or income[1]), the interpretation of variance becomes problematic as the strength of these associations increase.

[1] For extreme response style, the effects of each are significant and relevant in both the response style literature and initial assessments.

To demonstrate some of the empirical difficulty in analyzing the single question, I consider three questions in the 2004 ANES which ask a respondent to rate directly where they stand, and where the parties stand on this liberal/conservative scale (Figure I.1). Each panel in Figure 1 represents a distribution of responses, where respondents place themselves, Republicans, and Democrats from Extremely Liberal (1) to Extremely Conservative (7). The general pattern of response is predictably liberal for Democrats and conservative for Republicans.

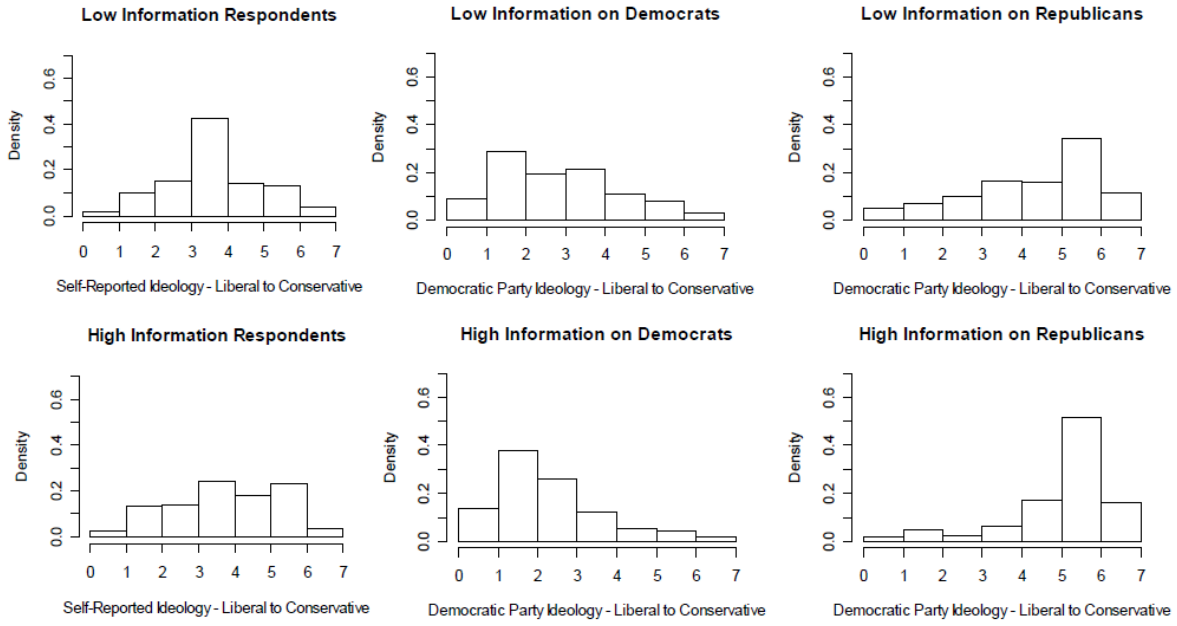
Figure I.1. Overall Ideological Self and Party Placement by Partisan Identifiers 2004 ANES (n=1,066)



The self-placements of Democrats and Republicans may look approximately "correct." However, the results by party present a different picture, and some concerns arise as a result. First, in the top row of charts, Democrats seem less consistent about evaluating their own party's ideological location than they are in characterizing Republicans. This tendency holds

for Republicans as well in the second row. Second, and as a corollary, "in-party" evaluations of Republicans and Democrats vary. Democrats are less precise in evaluating both themselves and Republicans than Republicans appear to be. We may be quick to assess this situation, in that Democrats "see ideology differently" than Republicans. But is this necessarily true or is it an artifact of who Democrats are?

Figure I.2. Overall Ideological Self and Party Placement by High and Low Political Information
2004 ANES (n=1,066)



Looking at Figure I.2 by level of political information, we get another picture – on where the respondent characteristics are pivotal. That is, as political information decreases, the perception of the parties becomes less precise, and predictably reverts to more central response. A cursory conclusion is that, when people who are more knowledgeable about politics are asked about ideology, they are more able to pinpoint where the Democrats are (as compared with their peers). On its face, these measures of party placement are predicated on accurate knowledge of where the parties are and a working knowledge and acceptance of ideology as a

way to organize political choices, and idiosyncratic cracks in that working knowledge is apparent when these figures are placed side-by-side.

Overall, the instructive lessons from this small example are that (a) responses do not necessarily conform to analysts' a priori expectations, at least not those to whom we could ascribe to knowledge of the construct that might be targeted, but (b) responses are formulated nonetheless. And, as a result, we cannot be sure if there is some meaning behind this self-report of ideology or not, and if so, what exactly is being captured. Similarly, if response styles possibly are present here, then what might we be able to say differently about our interpretations, especially because there is strong suggestive evidence that awareness and response style are linked?

The Role of Reasoning and the Political Survey Response

With roots in cognitive science, there are two, seemingly opposite propositions about how response style might be generated. The first contends that response style is a stable, fixed, and individual proposition: individuals possess a propensity to engage in default activity responding to surveys and do not deviate over time, between survey items, or between subject matter domains. The second proposition, however, contends that the default character of response may reveal a selective automaticity in the public mind (Bargh 1997, Bargh 1999, Bargh 2000, Eagly and Chaiken 1993, Banaji 2001). Traditional measures of automaticity have been timed response from the end of the survey question to the beginning or end of a response (Bassili 1996). This approach has been largely effective in ascertaining how the structure of attitudes are found in memory, yet this approach has been limited to cases where timed measurement has been feasible. Response style provides a potential avenue by which we might

assess the automatic evaluative character of the public mind and delve into what is behind automatic response, and inspect its regularities and peculiarities.

People utilize automatic responses because they are compelled to make an incomprehensible number of decisions in any day, internally deliberating, judging, and acting on any number of daily issues. To function normally, we recognize that deliberation is not always possible or desirable. To deal with our everyday life, it is imperative that we construct decision rules that best use the information we have toward functioning usefully personally and in society generally. In this light, it is not surprising that the strategies we use in everyday decisions functions similarly in politics. These rules are referred to as heuristics, shortcuts in how we evaluate political objects stemming from what we might call “low information rationality" (Popkin 1994).

Automatic response has garnered quite a bit of focus in the last two decades in political science. The intuition is fairly straightforward: people lack the complete information necessary to make decisions about political objects, and therefore rely on shortcuts, symbolic mental markers from which they might make a decision that is correctly reflective of their preferences. This is hardly a controversial or even recent line of observation (Downs 1957, Sniderman et al. 1993), but the use of heuristics and quantitative definition of mental uncertainty have been more keenly articulated and thoroughly explored recently as more than pure theory, but as a phenomenon for which we might find and construct evidence (Lau and Redlawsk 2001, Bassili 1996). Often, the target of this inquiry revolves around minimalism (Luskin 1987). Not only aren't we typically equipped to make political decisions, but we are often staggeringly deficient. Typical members of the public do not know much about politics and they know so little that they cannot readily connect issues together (Converse 1964), cannot identify basic facts about

government (Delli-Carpini and Keeter 1996), and are even structurally discouraged when seeking more reliable information (Alvarez 1998, Franklin 1991). There is no shortage of examples, so that when they are collected together, we have a picture of our collective brainpower that is less than flattering.

When confronted with these observations, the literary tone ranges from mournful to disdainful to prophetic and the tone largely depends on whether the author believes it really matters. Even if we do not have much information, we may know enough to make adequate decisions. Some accounts suggest we are 75 percent accurate (Lau and Redlawsk 2001) in discerning our true preferences with very little information, and if inaccuracies are random, then mathematically we are governed correctly and well (Erikson et al. 2003). Some evidence suggests that we are even insulated from grievous errors in public opinion despite our lack of knowledge because we rationally fail to collect information (Zaller 1992), so if we are called upon to deliberate we can do so effectively (Fishkin 1997). We may even create a division of labor by which we give opinion leaders the charge of wrestling with politics so we do not need to (Berelson 1954). The difficulty with this stance is that it is plausible, even probable, that there is systematic bias in representation and issue preference that is related to lack of knowledge (Verba et al. 1995, Althaus 1998).

The fact remains, when we ask people about their political attitudes and positions on surveys, for the vast majority of our samples we get less information than we might like, and even when we get it, we might not know what it means. We ask ourselves about our feelings about political figures, issues, and votes, yet we don't know who the figures are, have never fully formed an opinion on issues, and cannot even remember how we voted. We try very hard to obtain information from people who cannot give it to us. Nominally equal in our democratic

polity, we all can and do have influence. Like voting, analyzing the survey is usually approached in an egalitarian fashion, yet effectively, some opinions are more considered than others. When surveyed, we are all equally influential in our judgments. But, as discussed previously, we are not all equally willing or able to communicate the reasoning we would expect of a participatory democracy. For some research efforts it makes little difference. In polls, we want to know who the likely voters are, and for this purpose, knowledge is less important. If you plan to vote, your vote counts just as much as any other population adjusted likely voter. Yet for other endeavors, it is more appropriate to want to sample to population of knowledge and reason that comes through in individuals' responses. When we talk about issues, we may often care less about public opinion's center of gravity defined by individuals equally, but the center of gravity defined by knowledge, motivation, expertise, and influence. When we talk about ideas, the implicit weight of ideas comes from reasoning, persuasion, and causal stories (Stone 1989, Bromley 2006).

Measuring Response Style

To facilitate response accounting, this project focuses on a reliable and valid way to keep score. Measurement plays a key role in this effort. And the measurement and analytical strategy, while computationally involved at times, relies on a simple intuition. No single item is sufficient to identify any single latent construct, but a constellation of items gives us power to discern how a latent trait behaves. Variance between persons (within items) will give us leverage to estimate how items behave and variance between items (within persons) will give us leverage on how persons behave. The measurement challenge is to marshal the available survey resources to focus in on a consistent topic, which, for these studies is largely public issue evaluation. And using this, we might gain some insight into substantive propositions,

how the public mind connects political ideas and persons, how respondents translate these connections onto a closed-end survey, and what we can learn about respondents and their responses by paying attention to what they say. The natural modeling approach to a problem of item response involves using the methodological tradition designed and furthered to answer just these sorts of questions, item response theory.

To measure "what is said" is a first step, but in doing so, I extend these models to include not only what is said, but how it is said in context. This is accomplished by using what is said to control for a primary latent construct, leaving residual variance that can be distilled and constrained into the properties of behavior that indicate style, but do not indicate association with the construct being measured.

For this measurement enterprise, I assemble several secondary sources, large sample surveys that include political content and ask about issues, primarily the American National Election Study (ANES), an example using the World Values Survey (WVS), and the IEA (The International Association for the Evaluation of Educational Achievement) Civic Education Database, to assess the broad contours of stylistic response in various political settings. This ranges from assessment of individuals, groups of individuals, countries, and question context. These are instruments to assess whether style is uniquely political, uniquely American, or more importantly, unique to issue areas. And, where differences exist, what might be the possible meaning of any deviation and how might it tell us about the nature of the political respondent and how they interact with political issues? Facets of survey response style have been examined, but generally separately, acquiescence or extremity, but not in a sustained or comprehensive way across surveys, types of response, or in a pointedly political context. These

surveys in particular provide the breadth, depth, and consistency required to carry out this kind of research in the context of political science which spans several content areas.

Some of the irony and draw of response style research on secondary data is its reliance on questions that are less than ideal or otherwise analytically weak. Certainly this research can contribute to our knowledge about how to make questions better, but I would like to go a step further. Through this research, I want to interrogate these substandard questions in a new light, taking from them what their structure gives us to learn about the character of the survey respondent, the character of personal issue evaluations, and the character of issues in the public mind.

A good deal of effort thus far in this research agenda has gone into providing an analytical base for studying response style. The very nature of this dissertation's analytic space relies on proper measurement of a construct we know exists, but cannot see from any given response. Given the nature of the latent variable, I believe it is incumbent to make the case for its existence and the conditions for its appearance and stability. I have provided some theory for response style in the previous section. This section concentrates on the development of response style models/measures using patterns of response, and to validate these measures using prior research. This includes parameterizing response style explicitly using multiple items for the two largest components of stylistic response: (a) extremity being the tendency to pick endpoints or end intervals on a rating scale and (b) acquiescence indicating a positive bias toward more socially acceptable language in context.

For extremity, I have used some measurement models for extreme response using partisan feeling thermometer items, an open-ended 0-100 scale that gages feeling toward groups in American society (ostensibly 0-50 is “cold” and 50-100 is “warm”). The particular model I

have preferred uses a latent category selection tendency to identify both positive and negative endpoints as informative of extreme response style (ERS) and controlling for what we might call "true" extremity that is indicated by a respondent's answers to other questions that target a similar target concept.

For acquiescence, identifying and measuring a tendency to agree is slightly more problematic, but might be adequately isolated in surveys with more diverse content as in chapter 4. Ideally, acquiescent response (ARS) would be measured best with questions that are completely unrelated substantively. In practice, this is difficult to do, even by design. The best indicators will be surveys with more varied and unassociated content, such as the World Values Survey, which have been designed to touch lightly on a wider variety of subjects. As with ERS, eliminating the confounding effects of content is essential to the measurement effort.

Formalizing Intuition: Parameterizing and Identifying Response Style

This section provides a bit more analytical detail and defends the choice of item response theory to best model the patterns of choice that characterize response style. The choice to use item response theory is one that comes from the natural structure of response data. First, I model response style as a latent variable that cannot be identified through one question alone, but through patterns of response. The direction of the causal process of these data are not appropriately modeled from question response to response style, but viewing each question as caused by the latent response variables, the behavioral characteristic of response style. This is a deviation from the classical derivation of response styles (Greenleaf 1991) where response tendencies are counted and summed.

The core of the model of response style defines the Likert scaled question as the dependent variable, a question with K categories, typically ordered k in$\{5,7,10\}$, denoting the

points in the scale, sometimes centered upon a neutral category. Though these are ordered, the introduction of response style calls into question how these response categories are mentally ordered. Much like an unordered logistic or probit model, an important departure from the intent of the scale is that each item is allowed to be more or less indicative of extreme or acquiescent response. For extreme response, the outermost options will be more descriptive of an underlying tendency to respond extremely and for acquiescence, any agreement will be just as likely to indicate agreement.

In the classic logistic model, the dependent variables follow a multinomial logistic distribution with respect to a latent response style variable (Bock 1972). For simplicity, the underlying response style is assumed to be a standard normal variable, θ_{RS}. As response style tendencies (θ_{RS}) increase, the probability of selecting a category that agrees with the stylistic mental rule will increase with extremity identifying both endpoints as strictly increasing in (θ_{RS}) and with acquiescence all agreement categories increasing in (θ_{RS}). An important deviation from standard factor analysis here is the introduction of the question as having linear propensity function containing an intercept c_{ik} and a slope a_{ik} for each item category:

$$(I.1) \qquad P_{ik}(X_{ik}|\theta_j) = \frac{\exp\,(\,a_{ik}\theta_j + c_{ik})}{\sum_{\mathrm{i}}^{\mathrm{K}} \exp\,(\,\mathrm{a_{ik}}\theta_j + c_{ik})}$$

Where *X* is an individual question, *i* indexes the question, *k* is the question category, *k* in the denominator indexes category summation, and *j* indexes the individual respondent. The equation is constrained additionally to identify the model: $\sum_k a_{ik} = 0$ for each dimension and $\sum_k c_{ik} = 0$ and maintain the probabilities for all θ_{RS} to equal 1. Estimation of these models might be accomplished through a Marginal Maximum Likelihood, including an EM algorithm and using the Newton-Raphson procedure widely used in off the shelf packages such as Bilog, Multilog, PARSCALE, Latent Gold, or the open source "R" to name a few. More flexible and

advanced modeling of item response models has more recently been enabled through the development of Markov Chain Monte Carlo (MCMC) estimation packages as well, WINBUGS and STAN in particular have become standard for more exploratory IRT research in psychometrics. For these specific models in this thesis, I use of one software package in particular exclusively, Latent Gold, for its capability and flexibility in these particular applications which involve many parameters which require iterative steps for both the item threshold parameters and the underlying scale values (as value distributions in quadratures: Vermunt and Magdison 2005), and to establish some baseline of comparability, ease, and transparency between models.

While the nominal response model is the base model I use, several alterations and constraints are needed in order to identify response style specifically. No political science battery has been designed to measure and correct for response style, which leaves us with an array of questions that are targeted at substantive constructs. To isolate response style, as much as practicable, we need to neutralize content. In many cases, this can be accomplished relatively effectively when stylistic components are second or third factors in an otherwise one-dimensional scale. For example, in a group of questions that measures an underlying trait, ostensibly a political preference, a respondent might be genuinely predisposed to answer strongly as a reflection of a trait of preference. If one holds strong views defending Social Security and strong views defending Medicare, two very closely related programs, we might conclude that an extreme response is rightly attributed to this underlying association, and not attributed to an endpoint selection tendency. Introduce more measures, and more opportunities will exist that can demonstrate the style of the response separate from the substance of the scale. The innovation in measurement here is to find both a common target of interest, set

aside that preference, while developing other constraints in the model to identify systematic use of extreme selections which do not come from systematic connections of "true" preference.

For extreme and acquiescent response styles, this requires two different strategies because the mental model of extremity is bidirectional and acquiescence is unidirectional. For extreme response, this may mean randomly "flipping" the direction of each item (or add equivalent constraints) for a respondent to eliminate associated content effectively. For acquiescence, content needs to be explored more specifically in an attempt to identify items that are content associated and eliminate items, or correctly identify associations of content and account for them analytically.

It is important to note that the technology for inspecting response behavior has undergone tremendous advancement in recent decades. Technical advances in measurement have contributed a great deal toward latent variable modeling in political science, which comes as a result of two primary needs: (a) to alleviate measurement error and (b) to better depict the target of inference. For (a), the stock example is that of issue constraint and ideological connection. The amount of error that is contained in any one issue item gives an incomplete picture of how issues are situated in the public mind. For (b), in political measurement, individual issue items are often used in piecemeal as independent or dependent variables and are subsequently used to make inferences about larger theoretical concepts.

For better or worse, my background has been colored by a perspective on measurement that comes directly from my experience with educational testing. Sometimes it is necessary to have an air of confession when professing this, but I think this need not be the case given the role of measurement, surveys, and psychometric influences in political science (Poole 2005). Beyond individual measures, it has become more common to aggregate issues or characterize

common variance becaue we are often less interested in how a question performs in isolation, but as contributors to larger political phenomena that any one question cannot give us latent dimensions (e.g. participation, knowledge, partisanship). Measurement models have been staples of mainstream political research, notably as factor analytic models, but having more recently branched into item response modeling. Verba and Nie's four modes of political participation (Verba and Nie 1987) were derived from a factor analytic model as well as Claggett and Shafer's analysis of dimensions of political issues in the public mind (Claggett and Shafer 1995). Item response models formed the basis for "What Americans Know about Politics and Why It Matters" (Delli-Carpini and Keeter 1996) as well as a more recent analysis of Polity scores by Jackman and Treier (2008). Notably, ideal point estimation is a perfect example of political psychometrics in practice: a landmark development of a theoretical approach to measuring Congressional policy positions. Since its initial application (Poole 1985), ideal point estimates have become a dominant measure of partisanship and partisan distance in Congress. Yet, it is based in item response theory and a direct psychometric contribution, acknowledging that any one vote is an inadequate representation of a Congress member's overall policy position on a left-right continuum. Together, congressional votes typify the polar ideological nature of congressional choice and provide a more stable measure of partisan or ideological leaning, and they do so in a way that cannot be discovered on any one issue vote. Similar models have been developed in a Bayesian framework to estimate policy locations of Supreme Court justices (Martin and Quinn 2002). Notably, the use of more complex measurement models may exacerbate the problem of response style by accentuating a common element of style in relation to a reduced common variance among an array of variables (Treier 2005).

Important differences exist in the standard application of item response models to the study of politics that separates the focus of political scientists and psychometricians. They lie in no small part in the violations of response modeling assumptions: unidimensionality, local independence, and functional form. To have confidence in psychological testing results, the first step in doing so is to have some reasonable academic/social consent that what is being measured is valid and that individual estimates do no harm to the subject. The political science tradition has been less unified in its theory of learning than educational psychologists, yet political facts are undoubtedly learned, issue positions are learned, and partisan attitudes are learned. The difference seems to stem from the unstructured way in which we are politically educated and instructed, the way we process information, and are motivated to apply what we have learned. Politics has no standard curriculum and produces few right answers. We are allowed to learn politically in a way that best suits us personally, rendering the correct approach a matter of circumstance−whether we choose to expose ourselves to political trivia, objective issue positions of major parties or candidates, or learn as much as we think we need from talk radio or political advertisements. Therefore, testing political learning is inextricably linked with value judgments about what knowledge is important and how that knowledge is obtained, assembled, and transmitted onto a survey. Judgments in educational achievement are reflective of both the quality of the curriculum and the individual student's ability in a structured, homogenous environment. Classrooms across the nation may have different specific characteristics, yet the overall modular structure of schooling has not fundamentally changed for centuries. Politics, on the other hand, is an arena of competitive messaging, devoid of standard curriculum, and is not structured in a way that easily lends itself to inference. When using response models, political and psychometric applications differ in focus and

parameterization by necessity to address problems that are endemic to the heterogeneous political curriculum, both in content and style.

Organization of Thesis

This dissertation examines the rudimentary how and why of response style for political surveys, examines bias, and explores some possibilities these styles have for expanded political research. Again, by habitual response, I look at the way respondents approach surveys independent of content. That is, how respondents choose endpoints or acquiesce to the valence of a survey question independently of the object of measure.

The first overall task is establishing that response style exists and providing some baseline characteristics. Chapters 1 to 3 look at response style through the lens of only Extreme Response (ERS) which is the most robust measure of response style, and arguably one of the most influential, experimentally validate the phenomenon, delineating a potential measurement solution, and inspecting extremity as a political value. Chapter 1 relies on experimental evidence to validate Extreme Response Style (ERS), chapter 2 examines ERS and its impacts with regards to political extremity and provides an avenue toward approaching corrective action and examining potential bias, and chapter 3 examines ideological issue areas with respect to ERS.

The second overall task is to extend the knowledge of response style to specific political content. Where engagement and knowledge of content are at issue, how might we interpret higher levels of response style by content? The argument here is primarily one that contends that "*yes*, response style can tell us how respondents evaluate issues and political objects." As a corollary, we might know how engaged survey respondents are on issues, and measure by response pattern (importantly not self-report), what the degree of knowledge of a respondent is

on issues and whether or not they are cognitively active on them. In essence, do we know or care enough to think critically about content, or are we relying on a "backstop" of a default, extreme or acquiescent answer (chapter 3)? Chapter 4 continues on the theme of response style to look at acquiescent response (ARS) in terms of confidence in political institutions, and finally, chapters 5 and 6 then look at response style through the lens of moderation, along with ideological content fidelity measures, as an outcome in formative civic engagement and in a specific case of contemporary religious groups.

Chapter 1. An Experimental Assessment of Extreme Response Style

Response style, the systematic tendency for respondents to choose categories on Likert scales independent of the intended object of measure, has been a measurement concern for over half a century. 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 chapter 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. A potential solution lies in isolating a class of extreme responders and reweighting analyses to reflect extreme response.

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 chapter, 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. 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 L scale structure itself.

Response style speaks to both facets of how we might assess the quality of the questions we ask. 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 such as 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, pp.5-6) 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, p. 279). 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, p. 162) By implication, if one were to look for response style, public opinion is as ideal a place to look as we might expect.

The analytical strategy in this chapter uses two approaches: 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).[2] Survey response styles have been theoretical and statistical issues dating to Cronbach (1946, 1954), and have most pressingly concerned marketing researchers who have attempted to account for irregularities that appear in different populations when 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 contexts (Cheung and Rensvold 2000). Yet, even within countries, individuals have importantly different approaches to surveys linked to 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 cognitive demand for scales with many options, which effectively limit responses to a few select options (Baumgartner 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, I emphasize one compelling cognitive argument for its existence and impact: 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

[2] We might also view these strategies as habitual or automatic response (Bargh 1997), specifically induced by a question's structure and not its content.

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 that those with less political information will have less ability or motivation to respond fully, will tend to use simplification strategies, and consequently will exhibit systematically less reliable responses.

The same process has been described in terms of one or more secondary question-answering strategies.[3] 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 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), where a correct answer is inferred from a narrow subset of all responses, and in such a way, response style is analogous to nested guessing.

Furthermore, 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 like extreme responses,

[3] 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.

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, if any, might be. 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) or a random "sampling" component.

By its nature 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 and political 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[4] 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

[4] In the context of Acquiescent Response Style (ARS).

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),but rather a more reflexive, automatic response disposition which is not attributable to substantive dimensions of interest.

Measuring ERS

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 to 20 percent in marketing research scales (DeJong et. al. 2008). Second, unlike acquiescent response, extreme response can be more credibly isolated analytically through psychometric methods developed and fine-tuned over the last several decades. That is, what we know about ERS can be known with greater confidence.

To measure ERS, I appeal to some more recent modeling work that 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 meaningful, political dimension from extreme response (Bolt and Johnson 2009).

To accomplish both measurements (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[5] which together tap a latent pattern of attitude from left to right on the political spectrum (Brady 1990, Brady and Sniderman 1985).[6] Because thermometer items have a uniquely extended format,

[5] 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.

[6] 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

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 that I use in these analyses[7]). 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: $P(X_{ik}|\theta_{1j}, \theta_{2j})$ is the probability of selecting a response category given a substantive latent variable, θ_{1j} , and an extreme response latent variable, θ_{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 parameters over each item is constrained to zero.[8] Non-responses, no opinion categories, and refusals are also constrained to exhibit no influence in their relationship to political attitude and

---

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).

[7] 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.

[8] 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).

ERS ($a_{i(NR)1} = 0, a_{i(NR)2} = 0$), which provides the interpretation of ERS conditioned on selecting a point in the scale.

$$P\left(X_{ik}\middle|\theta_{1j}, \theta_{2j}\right) = \frac{\exp\left(a_{ik1}\theta_{1j} + a_{ik2}\theta_{2j} + c_{ik}\right)}{\sum_{\mathrm{i}}^{\mathrm{h}} \exp\left(\mathrm{a}_{\mathrm{ih1}}\theta_{1j} + a_{ih2}\theta_{2j} + c_{ih}\right)} \tag{1.1}$$

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 (θ_{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 (θ_{2j}).[9] Respondents who have a greater tendency to choose endpoints, from right to left, will be increasingly drawn to the ends of the thermometer scale, separately from their political attitudes. In order to be most explicit, further confirmatory restrictions on ERS identify what we might expect of extreme 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 ($a_{i1} = a_{i9} = 1.0$), which constrain the remaining categories to sum to (-2.0) . The probability of selecting an endpoint will be symmetric and not dependent on the direction of

[9] These latent dimensions, like oblique factor rotation, are also allowed to correlate with one another.

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 ($a_{i1} = a_{-i1}, a_{i2} = a_{-i2}, \ldots, a_{ik} = a_{-ik}$).[10]

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

[10] 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.

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 take advantage of 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 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).[11] The character of each study in Krosnick and Berent's reference set differs slightly in structure, and several small *n* studies, for example, cannot be used here. These differences occur in the target questions themselves: sample size, differing base thermometer items, and the variable lag between the base and target questions.[12] What remains constant in these studies, however, is that 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 (n=2,461) with three focal issue questions.

To illustrate the comparability of the questions, an example of a branching and scaling is considered in the following presentation (Table 1.1) and in the accompanying diagram (Figure 1.1).[13] The scaled defense question is provided on the left and the branching equivalent is provided on the right.

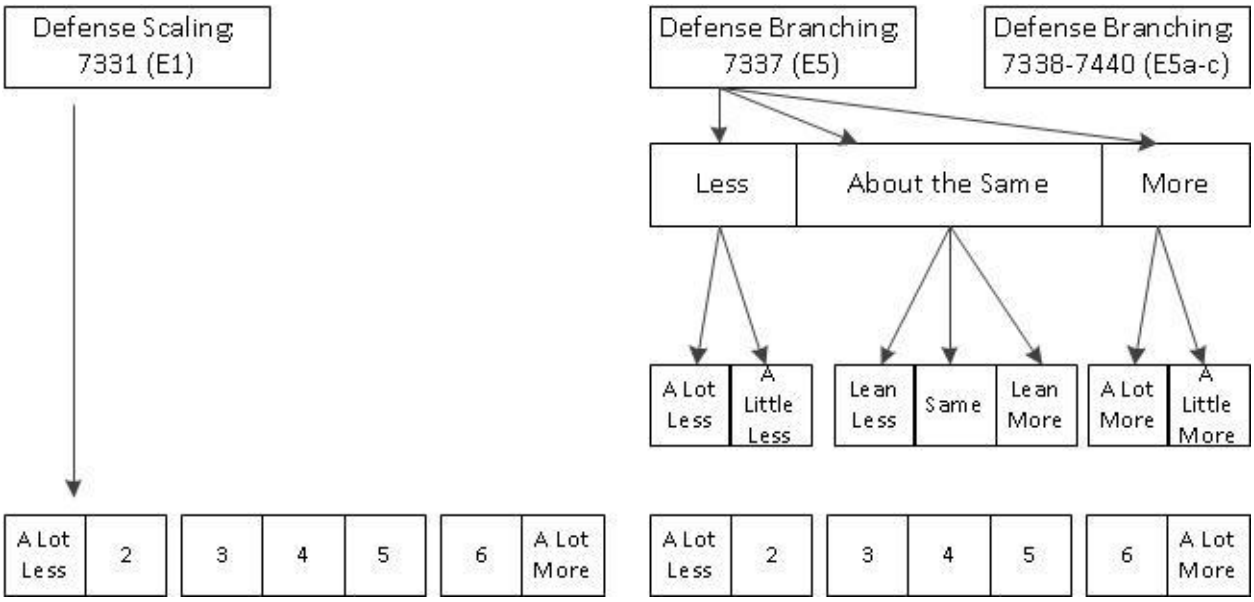
[11] Where branching and scaling items are equivalent in category number and category meaning. Notably, partisan identification differs in how conditions are labeled and more recent experiments in branching and scaling (ANES 2000 and ANES 2008) utilize different numbers of categories and different questions respectively.
[12] The administration profiles, base, and focal questions are presented in A1-A3 in the Appendix.
[13] Question text for the other scaling items is provided in the Appendix.

Table 1.1. Randomly Assigned Branching and Scaling Example

Condition #1: Scaling Defense (1989 ANES Pilot) Question 7331 (El)	Condition #2: Branching Defense (1989 ANES Pilot) Question Summary 7341 (E5x)
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)?	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? -- E5a. Would you say the U.S. should spend a lot less or a little less on defense? -- E5b. Would you lean toward spending less on defense or more on defense? -- E5c. Would you say the U.S. should spend a lot more or a little more on defense?

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 1.1).

Figure 1.1. Scaling Versus Branching Question Format Example



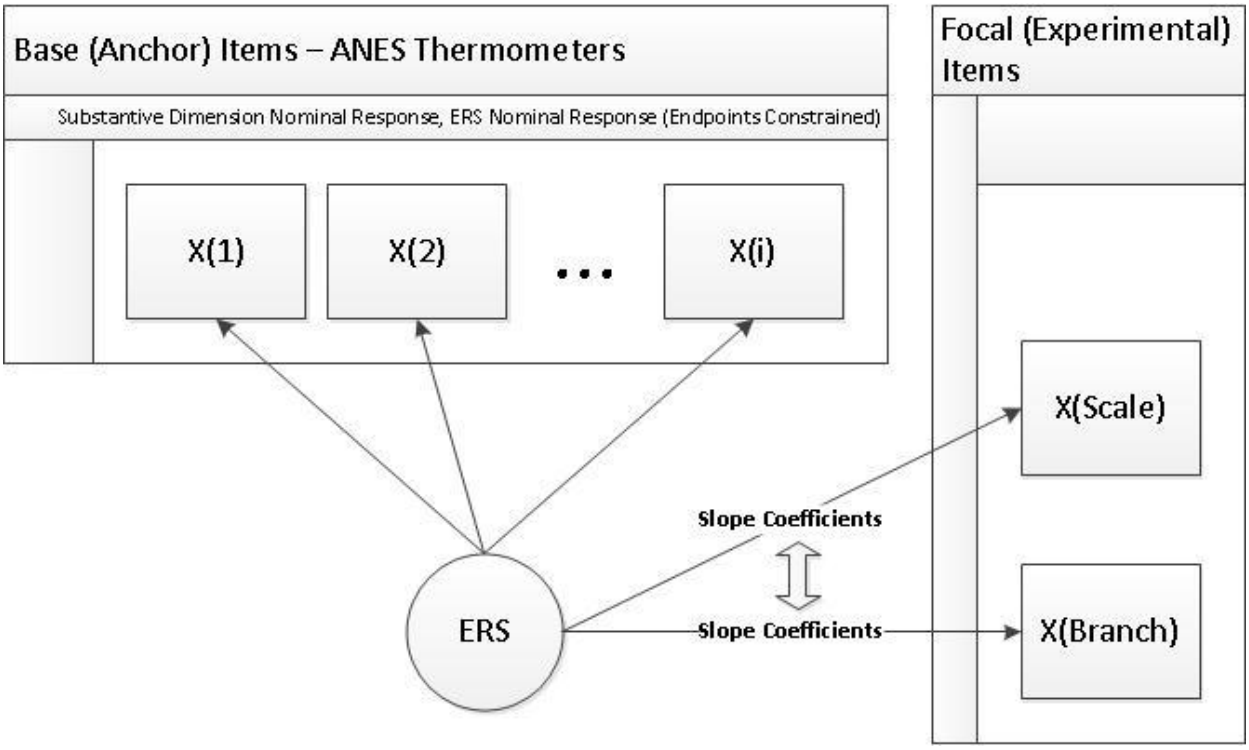
Random assignment between questions similar to the exemplar in Table 1.1 and Figure 1.1 insures that more or less equivalent groups receive both formats and that only a single facet is being manipulated: the format of the item. Additionally, by randomly assigning these branching and scaling items, the estimates will not have the 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 1.2 sketches the measurement model along the top, common variance between a set of base thermometer items (X_1 through X_I). 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 1.2 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 1.2 adds a path to both of the identical experimental and focal items.

Figure 1.2. Experimental Design Path Model: Focal and Base (Anchor) Items (X_I)



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 (1.1), estimated together.

Within each of the paths in Figure 1.1 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 equally distant from the endpoints 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 $(a_2 = a_6, a_3 = a_5)$, 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:

Hypothesis 1.

H_0: $a_{(Scale)1} \leq a_{(Branch)1}$
H_A: $a_{(Scale)1} > a_{(Branch)1}$
(endpoint difference in category selection by ERS scaled vs branching condition)

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.

Hypothesis 2a.

H_0: $a_{(Scale)1} \leq 0$
H_A: $a_{(Scale)1} > 0$
(endpoint increase in category selection by ERS in scaled condition)

Hypothesis 2b.
H_0: $a_{(Branch)1} \geq 0$
H_A: $a_{(Branch)1} \geq 0$
(no endpoint increase in category selection by ERS in branched 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 ($a_{(Branch)1}$).

Experimental Results

The first question in these items is of average endpoint selection effect. Do endpoint selection effects due to ERS appear to be significant when scaling is introduced experimentally? On nearly all issue questions, the answer is "yes." The results, comparing the branching condition to the scaling condition, are detailed in Table 1.2. The results are reported 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 −difference between items, the positive difference in scaling items, and the insignificance of branching− show some support. 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 1.2. Slope Coefficients for Category Selection due to ERS $a \in (a_1, a_2, a_3, a_4)$

			Slope Coefficient from Extreme Response (a_1) to Midpoint (a_4)			
Question and Difference Estimates (p-value – scaling endpoint mean selection greater)		Format	a_1	a_2	a_3	a_4
Defense	(0.003)	Branching	-0.10	-0.07	-0.18	**0.72**
		Scaling	**0.36**	0.02	**-0.29**	-0.16
Central America	(0.002)	Branching	0.09	-0.20	-0.18	**0.57**
		Scaling	**0.39**	-0.05	-0.15	**-0.37**
Gun Control	(0.001)	Branching	-0.16	-0.17	-0.18	**1.02**
		Scaling	0.15	-0.19	-0.07	0.23
			a_1	a_2	a_3	a_4
Limits on Foreign Imports	(0.000)	Branching	**0.18**	**-0.21**	0.00	0.06
		Scaling	**0.61**	**-0.23**	**-0.36**	-0.06
Sanctions on South Africa	(0.001)	Branching	**0.29**	-0.03	**-0.27**	**0.12**
		Scaling	**0.62**	-0.11	**-0.37**	**-0.28**
			a_1	a_2	a_3	a_4
Defense	(0.000)	Branching	0.12	0.13	-0.05	**-0.40**
		Scaling	**0.37**	0.06	**-0.26**	**-0.34**
Good job / Standard of Living	(0.000)	Branching	0.08	-0.01	0.12	**-0.39**
		Scaling	**0.56**	-0.11	**-0.32**	**-0.56**

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 note. Specifically, in the 1988-89 ANES Pilot, the three questions appear to replace endpoint selection with midpoint selection, a property that 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 1.2 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 1.3 and Figures 1.3 through 1.9 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 1.3. Probability of Endpoint Selection by Extreme Response Style

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

Because the ERS scale is approximately standard normal, each point on the scale may roughly approximate a percentile rank of extreme response tendency in a population. Presenting the range of the entire scale is instructive, namely to illustrate the functional form of endpoint selection in practical terms. With respect to the increasing propensity to respond extremely, the function is linear, yet in term of proportion, endpoint response becomes more influential in absolute terms as the scale increases toward the top of the scale. With respect to the proportion of endpoint selection overall, this might suggest a point at which this endpoint

tendency becomes problematic. As it stands, in Table 1.3, if you possess little of this tendency (say -1.5 on the scale as Respondent #1), you would still choose the endpoint for Defense 5.6 percent of the time and presumably it would be a "truly" extreme response. At 1.5 (Respondent #2), by contrast, you would select the endpoint 24.2 percent of the time. Ascribing meaning to this, we might interpret the respondent at -1.5 as having a "normal/conservative" range of selection and the respondent at 1.5 would have a less "normal" response that prefers endpoints. That leaves a difference of 18.6 percent of the time, on questions like Defense, Respondent #2 picks endpoints "independent of the object of measure." If we imagine the (unlikely) scenario in which a complete sample of respondents are a -1.5, or completely at the 1.5 level of ERS, at the very least, we have two samples that are wildly different in their distributions. At more, we have a tremendous source of potential bias that has nothing to do with an intended characteristic.

Figure 1.3. Experimental Results: Defense Spending,Branching v Scaling 1988 ANES and 1989 Pilot

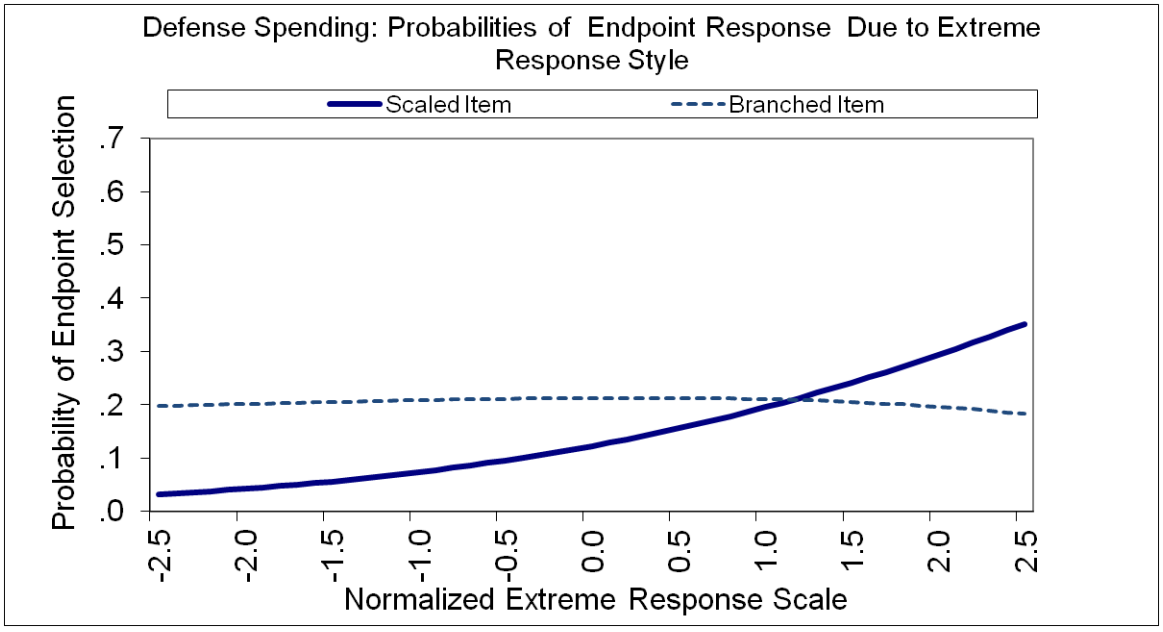


Figure 1.4. Experimental Results: Aid to Central America−Branching v Scaling 1988 ANES and 1989 Pilot

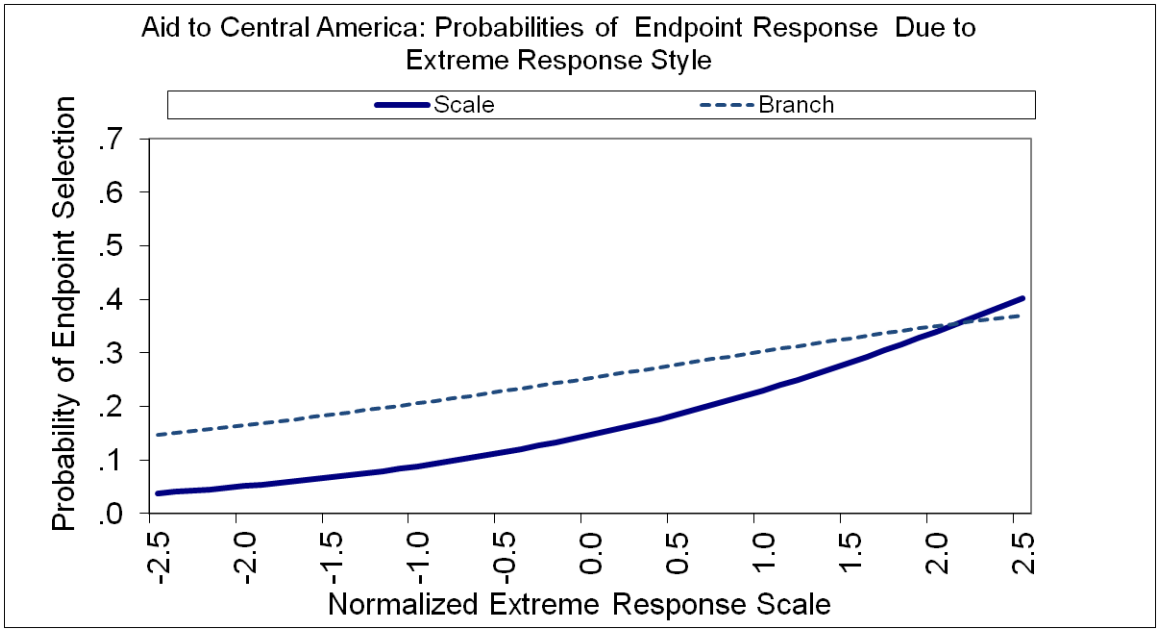


Figure 1.5. Experimental Results: Gun Control−Branching v Scaling 1988 ANES and 1989 Pilot

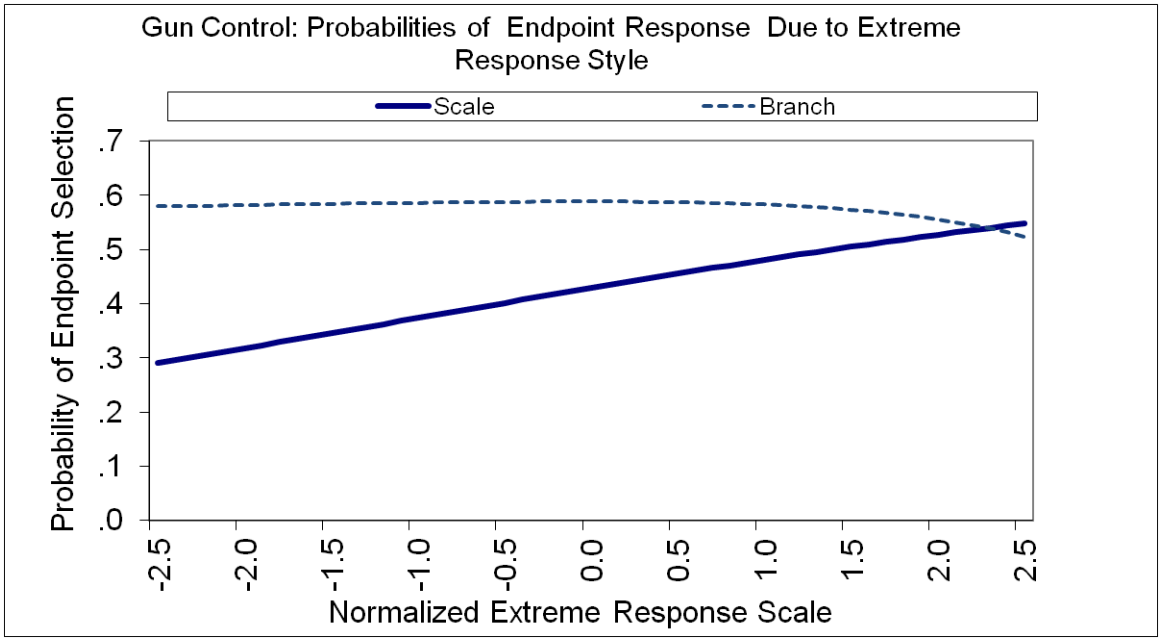


Figure 1.6. Experimental Results: Limits on Foreign Imports–Branching v Scaling 1990-1992 ANES Panel

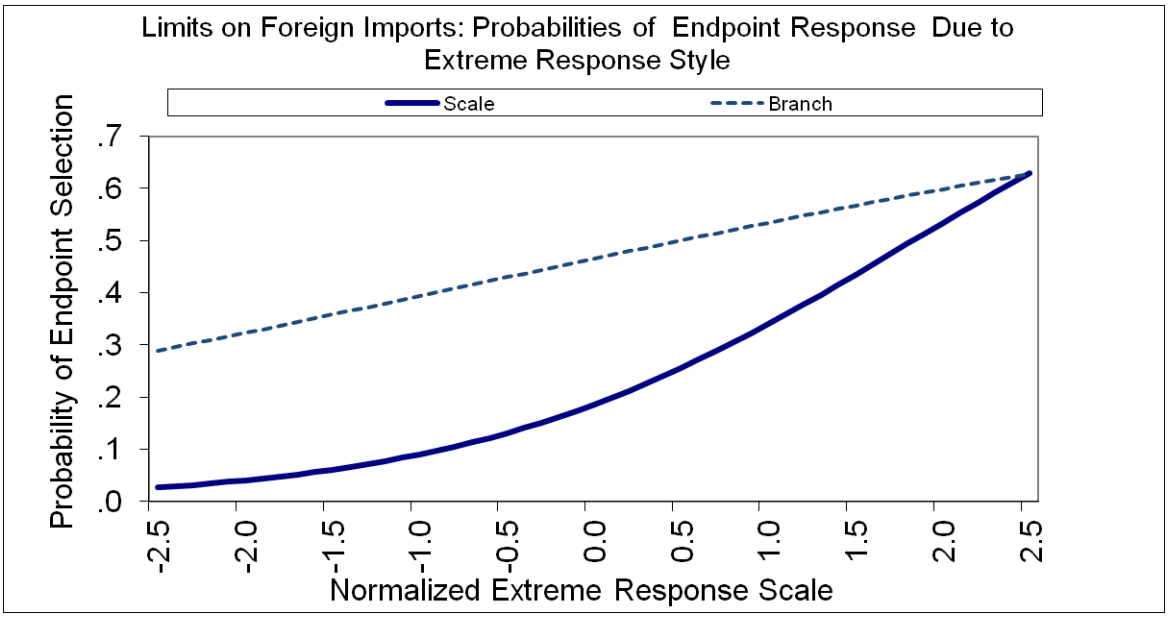


Figure 1.7. Experimental Results: Sanctions on South Africa, Branching v Scaling 1990-1992 ANES Panel

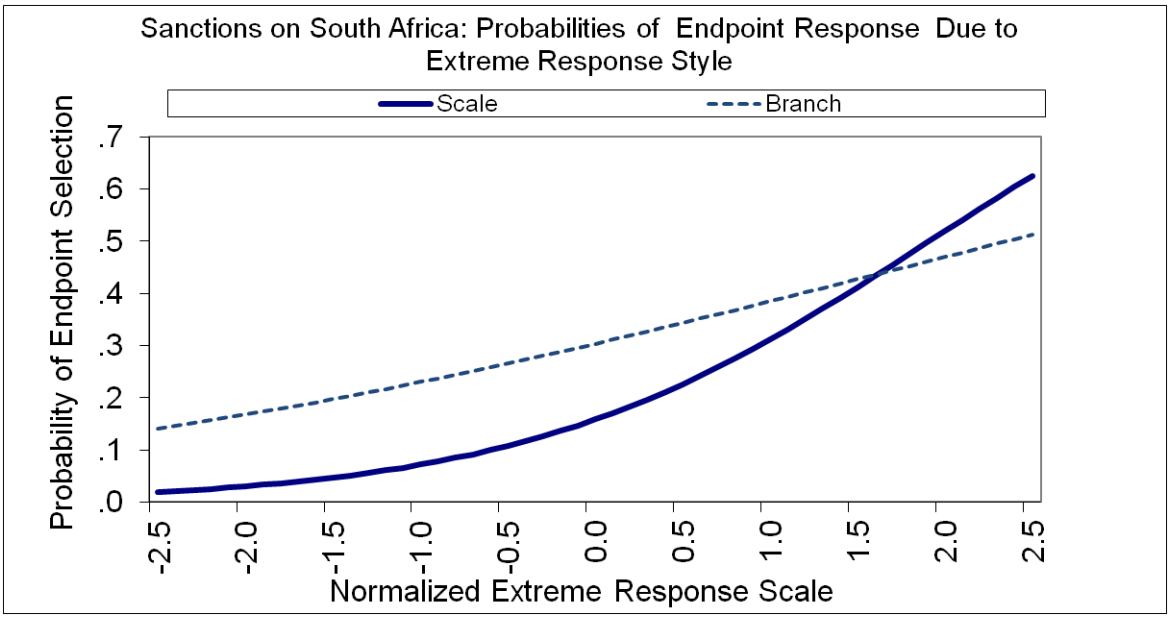


Figure 1.8. Experimental Results: Defense, Branching v Scaling
1982 ANES Methods Comparison Project

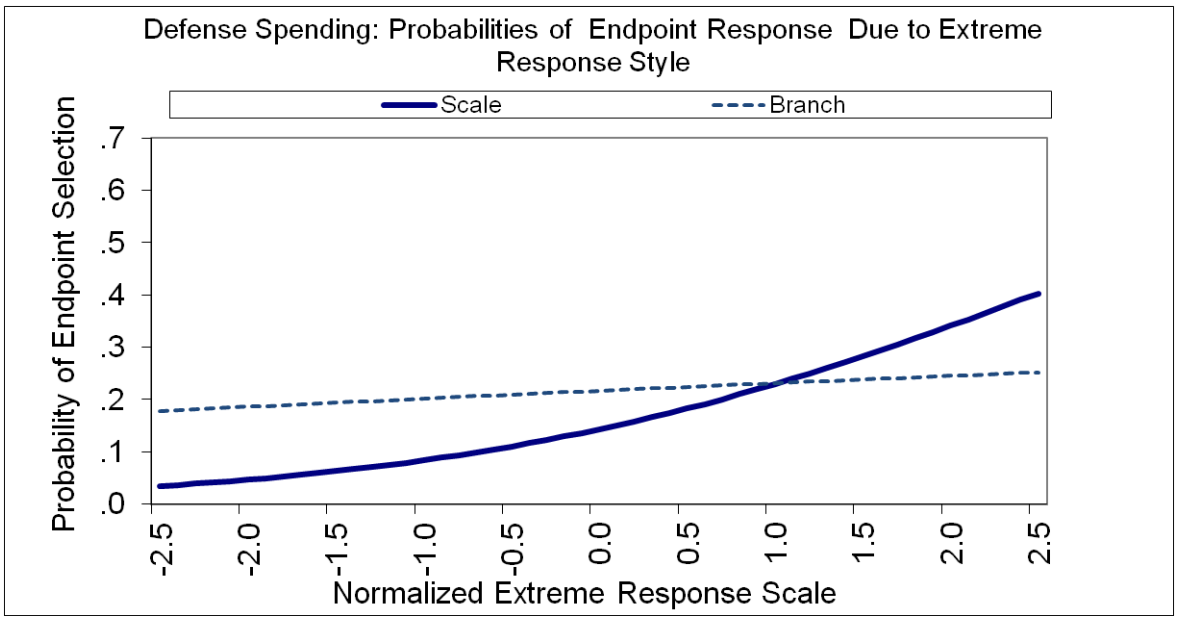
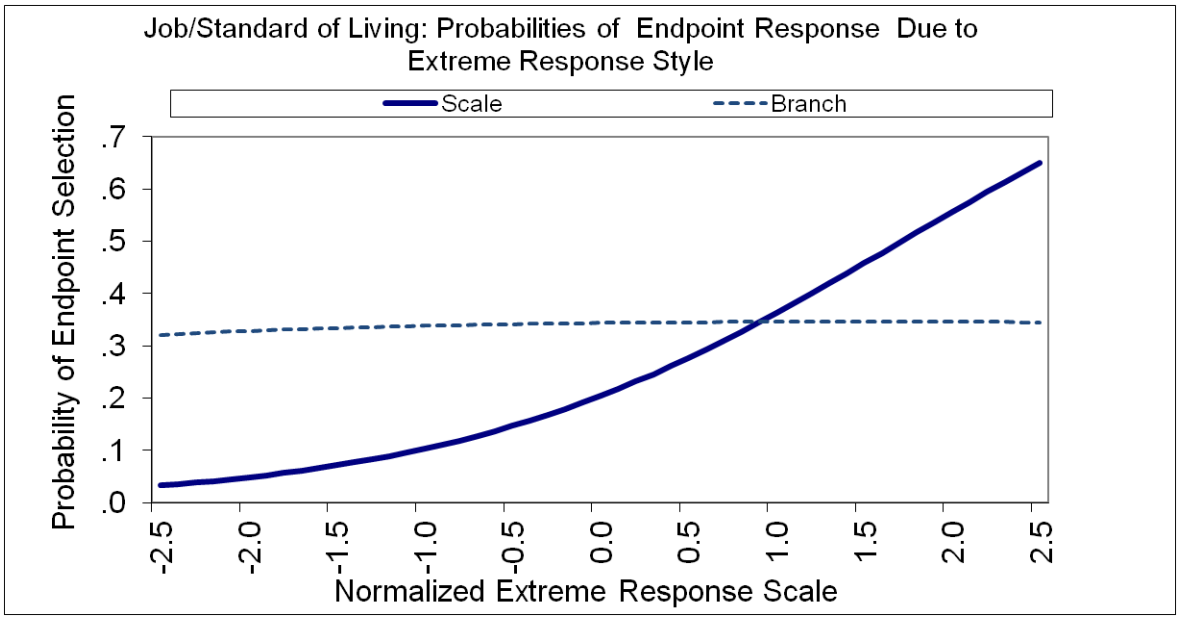


Figure 1.9. Experimental Results: Job/Standard of Living, Branching v Scaling
1982 ANES Methods Comparison Project



Overall, as ERS increases, as we might expect from the slope coefficients, the probability of selecting an endpoint on the scales increases. For more modest ranges of response style (-0.5 to 0.5), a change in ERS represents increased likelihood of residual endpoint selection, a range of roughly 50 to 100 percent greater, moving from a modestly "non-extreme" responder to a respondent who might more often exhibit this behavior. For this range of respondents, however, we might be less interested in the marginal increase of 6 to 13 percentage point increase, but the more pointed 9 to 18 percentage point increase apparent when we go from less modest ranges to stronger endpoint response (0.5 to 1.5). At 1.5 then, the top 93rd percentile of the distribution, these extreme response tendencies increase to include 25-50 percent of all responses.

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 ranges of the data. Content held constant, with the

exception of Gun Control,[14] 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 correlate to this selection. Otherwise, this increased variance may not be indicative of more serious substantive bias, it may merely indicate 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 speak to the cognitive optimizing and satisficing behaviors. Where more resources are brought to bear (political awareness), there should be an increase in the ability to optimize: process the question's meaning, retrieve relevant opinion, process the sampled opinions in memory, and deliver thye judgment to an appropriate response category.

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 (1.1) and performed simultaneously with estimation of the

[14] 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.

measurement model where the x_i in the following (1.2) represent "Fairly High," "Average," "Fairly Low," and "Very Low" political awareness, with the baseline, omitted category, of "High Awareness":

$$\theta_{2j} = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 \quad (1.2)$$

Each coefficient (β_i) estimate should be more positive, such that as awareness decreases, ERS should become stronger.

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 1.4. Adding Political Information as a Determinant of ERS

Question/Coefficient (SE)	Fairly High	Average	Fairly Low	Very Low
1988-89 ANES + Pilot	0.110	*0.337*	*0.459*	*0.893*
	(0.099)	(0.096)	(0.102)	(0.117)
1990-92 ANES + Panel	-0.171	0.141	*0.443*	*0.663*
	(0.109)	(0.105)	(0.109)	(0.126)
1982 MCP	-0.044	0.049	*0.261*	*0.744*
	(0.104)	(0.100)	(0.113)	(0.134)

*Very High information is omitted as a base category

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 responses for responders with varying degrees of 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 $\left(X_{ik} \middle| \theta_{1j}, \theta_{2j}\right)$ for each respondent according to equation (1.1) using item parameters from the scaled experimental items. The correlations in Table 1.5 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 a conservative estimate.

Table 1.5. Issue Simulations and Pearson Correlations at ERS Quintiles

	Mean Correlations of Variables with Differing Levels of ERS (n=5,000 – 30 sets of correlations)				
Question – ERS Quintile	Bottom	2nd	3rd	4th	Top
1989 Pilot Defense	0.067	0.072	0.092	0.099	0.092
1989 Pilot Central America	0.047	0.046	0.055	0.062	0.055
1989 Pilot Central Gun Control	0.004	0.007	0.004	-0.009	0.001
1992 Panel Foreign Imports	0.002	0.007	-0.002	0.011	-0.013
1992 Panel South Africa	0.047	0.058	0.064	0.068	0.086
1982 MCP Defense	0.068	0.076	0.078	0.080	0.091
1982 MCP Jobs	0.125	0.133	0.134	0.159	0.164
Increased r - 1982 MCP Jobs	0.364	0.356	0.369	0.404	0.392
Increased r - 1982 MCP Jobs	0.655	0.649	0.652	0.673	0.637

From these simulations, we can see that Pearson correlations increase mostly but 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 0.04-0.04 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 and 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 our 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 establish 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 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 to 100 percent over the heart of the distribution from -0.5 to 0.5, but increasingly more likely and consequential as ERS gets larger. 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.

Yet, 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? In scales where ERS is present, the bias can be substantial. Notably, and most simply, 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. If we suppose that response style increases with ambiguity of an issue , issue evaluations may show more correlation, and

more correlation relative to the magnitude of any true association, as questions become more ambiguous, difficult, or obscure.

With these results come new questions that have yet to be addressed and 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 in particular circumstances. 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 is 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 the 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 across issues. If one issue stands out as potentially instructive, it is gun control. Respondents to gun control questions, 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.

Finally, if the results here are even mildly convincing, there is sufficient ambiguity in response item, or the target population is more likely to be less informed, there are ample avenues toward solution. Solutions may be as simple as reweighting analyses to discount for respondents who us response style, using cut points more bluntly or establishing a likelihood of being in an extreme responder class. These response problems and their solutions should concern political scientists especially if effect sizes are small, and, recognizing that there are larger consequences for even smaller biases, whether it is in the big business of public polling or whether it is in more academic endeavors.

A 1.1. Administrations Used in Analysis

	1988 NES + 1989 Pilot	1990-1992 NES Panel	1982 Methods Comparison Project
Focal Issue Items	3	2	2
Content	Defense, Central American Aid, Gun Control	Foreign Imports, South Africa Sanctions	Defense, Jobs and Standard of Livings
Base Items	24 Lagged Thermometer (1988)	16 Concurrent Thermometer (1990)	7 Concurrent Thermometer
n	2,040 (Base) 614 (Focal)	1,980 (Base and Focal)	2,416 (Base and Focal)

A 1.2. Scaling Question Text Used in the Studies

7331 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)?

7332 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.)

7333 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

A 1.3. List of Base Thermometer Items for ERS

1988-1989 ANES + Pilot (Base *i*=24)	1990-1992 ANES Panel (Base *i*=16)	1982 ANES Methods Comparison (Base *i*=7)
Democratic Party	George HW Bush	Jimmy Carter
Republican Party	Mario Cuomo	Ronald Reagan
Big Business	Dan Quayle	Edward Kennedy
Labor Unions	Ronald Reagan	John Anderson
Liberals	Jesse Jackson	George HW Bush
Conservatives	Dem Party	Walter Mondale
Civil Rights Leaders	Rep Party	Dem Party
People On Welfare	Supporters Abortion	
Congress	Conservatives	
Environmentalists	Labor Unions	
Anti-abortionists	Women's Movement	
Federal Govt	People On Welfare	
Fundamentalists	People Pro Enviroment	
Feminists	Liberals	
Michael Dukakis	Poor People	
Lloyd Bentsen	Abortion Opponents	
George HW Bush		
Dan Quayle		
Ronald Reagan		
Kennedy		
Dole		
Carter		
Evangelical Groups		
Jesse Jackson		

A 1.4. 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];
```

Chapter 2. The Effect of Extremity on Behavior: The Bias of Response Style

Having examined whether or not extreme response style is meaningful and potentially influential in surveys from an experimental standpoint and having looked at potential impacts, a natural follow-up question is how it might be possible to control for ERS in surveys, and what impact ERS may have on simple outcomes, both inside the context of suspect survey questions and in a broader context. The proposed solution here is to estimate and control for ERS by considering Extreme Response as "non-compliant" survey behavior where possible, or, by the same token, consider Extreme Response as missing values.

As in chapter 1, I construct a measure of extreme response and connect it to ideological extremism according to respondents' reactions to political groups and persons. This creates a picture of measured symbolic ideological strength, and an unrelated extreme response disposition. Then, by looking at self-reports of political behavior (voting, donation, event participation, and influencing others), some validity of the ERS measure might be ascertained where strong behavior and strong opinion are linked. Finally, by adjusting for stylistic response in surveys, we might discover what the impact might be of strong partisanship on the decision to vote, and how likely stylistic responses may affect our assessment of this impact.

When Responses are Missing: Survey Fidelity

In the experimental world, in order to obtain causal effect estimates of a treatment, a treatment is randomly administered to a population and another group is set aside in a control. The net result is an unbiased estimate of a treatment's effect. However, subjects do not always adhere to treatments in a way they were intended, especially outside of laboratory settings, causing fidelity problems that require inspection in order to claim valid causal inference (Hulleman and Cordray 2009). In fact, it is often the case that a large number of subjects,

especially in field studies, do not comply with a treatment regime when offered. This presents a special problem for the experiment, since it is easily hypothesized that compliance behavior itself may introduce positive or negative bias.

Political pollsters know this situation well in another, related context. It is difficult to obtain a truly random survey sample from a population for purposes of inference. Those who do not choose to answer a telephone survey, actively engage in an online sample, or refuse to complete a survey may still show up to the polls on election day and by virtue of voting can be assumed to have opinions. Related to their refusal to engage, these hidden subjects may vote one way or another systematically. The answer to this conundrum is often to look at observable characteristics of a population and their behavior and re-weight them appropriately−what is often called the "art" of polling.

For extreme response in particular, I have demonstrated that habitual extreme respondents are indeed expressing something other than true opinion the way we intend. Thus, if this something else is being employed habitually, it may not correctly be viewed as a response at all even if the respondent has given answers or the respondent's opinion as recorded is only partial. Yet, traditionally, we believe generally that "a response is a response is a response" with regard to feelings and views. However, the presence of an answer is not necessarily proof of its validity. The, perhaps controversial, point of this thesis is that sometimes a response is true and sometimes it is not, or is partial in its truth, and to the extent we might measure and identify these untrue responses, they might be purged from analysis or partially suppressed. They have not really "taken the pill" from which we wish to measure effects, thought they seem to have swallowed it.

What do we do, then, with these respondents? One simple solution is to identify offending respondents and count these responses as "missing" for the purposes of analysis. This is not an analytically new problem in medical or social science research, which often have missing values, to which several strategies might suffice. First, we might simply exclude cases that are missing. If cases are missing "completely at random," then this strategy is adequate (Little and Rubin, 2002), giving less than accurate standard errors in the process. Two other potential alternatives are inverse-probability weighting (IPW) (Höfler et al. 2005) and multiple imputation (MI) (Little and Rubin, 2002). In IPW, again only complete cases are included in the analysis (excepting analysis of repeated measures, which we do not treat here). Importantly, however, weights are used to rebalance the set of complete cases so that it is representative of the whole sample.

This analysis takes a look at those who complete surveys in the American National Election Studies, but may not be giving scaled question complete attention in the way we design and intend them. I assess bias in effects across the electorate by estimating some basic models of behavior for the entire population surveyed, then look at how these estimates change based only on the "compliers" in the survey environment (missing completely at random), and, in turn using inverse probability weighting (missing at random).

Extreme Response, Extreme Ideologues, and Behavior

As a proof-of-concept, this study focuses on specific survey material in the ANES for which two exacerbating factors for extreme response bias may exist. The first is in choice of subject, inspecting participants who hold extreme views. The second is with data that are scaled, which are, by virtue of combination, more likely to be affected by systemic bias (Ansolabehere et al 2008). When we adjust for extreme response style, extreme ideologues

might be expected to bifurcate toward those with consistent true extremity, and those who choose extremes by virtue of habit or style.

Two targets are of interest here. First, I reinforce the case in this chapter that what is measured in ERS is different, incomplete response, and consequently, untrue opinion. Acknowledging that there is room for interpretation around response style is warranted, just as we might interpret a "don't know" or a "refusal" and look at their relative rates. This is also true in how we might interpret any latent variable that is constructed to measure personal attributes that are not directly observable through one response, yet evident through a pattern of responses across multiple questions. I discuss several variants and nuances of what an extreme response trait might represent, as measured in this chapter and previously. I discuss what extreme response *style* might mean and how it might complicate the interpretation in the following chapter's analysis. A hardened skeptic may believe that this endeavor is more akin to relating issue choice, partisanship, and voting to traffic infractions or some other personality characteristic. The virtue in this chapter is in the validation and inspection of the properties of ERS. Even if, at a minimum, the reader may refuse to accept any meaningful ascription to the concept of ERS or the measures I provide, it is, at the very least, the fidelity of "strength of opinion" that comes in question and comes into question differently for different opinions. Just as we might not ascribe any specific meaning to "don't knows" or "refusals," yet still note the importance and bias they might contain, we can note with concern how ERS varies between questions to show us strong opinions that are not objectively strong at all and remark how this erstwhile strength varies. If nothing else, ERS represents a source of potential differential bias with which to contend, and a serious threat to research that uses common opinion measures in the American National Election Survey (ANES).

As far as extreme opinions, I invoke some of what we know about ideological polarization. As background, political polarization is relevant for research as alignment along several dimensions of potential issue division, and categorizes respondents into groups around differing clusters of opinion. This has the effect of factionalizing the electorate. Thus, issue alignment, rather than extremity, is the part of the process that will most likely foster "deep" division in the electorate and is most concerning to some analysts. Our divisions are polarized inasmuch as we line up neatly along many issues simultaneously, even if we are not particularly extreme. Peoples' distances from one another in the political sphere are a by-product of correlated views, and not extreme views (Converse 1964), even though we often normatively fixate upon issue opinions in isolation that are extreme as constituting radicalization.

Second, we might not know exactly whether a response is untrue or not with simply a measure in hand that we nominally call response style and is set up to measure this tendency. For this, I consider the summary statement that is a baseline for opinion research. Inasmuch as we care about the intensity of opinion: "…strong *attitudes are predictive of* behavior, whereas weak *ones are* not" (Miller and Peterson 2004, p.848). With ERS, this truism is, in theory, violated such that strong, even extreme opinions are expressed that are not strong opinions, but vestiges of habitual behavior. If we care about this violation, then it may be useful to know when strong opinion is not strong opinion, and how that may bias results under two conditions: (a) if true opinion can be isolated from extremity and (b) if this has a meaningful relationship outside of the context of the response itself.

For this, I examine very rudimentary questions about how active respondents are on the American National Election Survey (ANES) using participation questions in the ANES and voter turnout. I then look at whether they are extreme responders or not and, consequently,

whether respondents are actually engaging in surveys as we require for inference based on our ANES sample.

The first hypothesis for behavior then comes from what we expect of those with strong views that are true: "Extreme views should prompt more active political behavior" and its corollary, "Extreme responses without extreme views are not indicative of more active political behavior." For those who are ideologically consistent and highly opinionated, we would expect a higher degree of political action. In contrast, for extreme responders, that political action should be no greater than for the population as a whole.

While there is much evidence to cite the existence and extent of ERS, there is still no consensus or discussion of what exactly makes a respondent extreme (or not) and how to delineate between respondents for whom responses might be aberrant and respondents for whom response adhere to our analytic constructs. A standards-based threshold might specifically link an accepted benchmark probability of who holds extreme views: what proportion of the electorate fits into the distinction of an ideologue (liberal or conservative), what kinds of issue and symbolic connection is necessary to qualify, or who holds any modicum of issue consistency. There is no accepted standard here, and in response style, there is even less of an accepted standard. As a result, I revert to a more arbitrary, normative distinction and set a 10th percentile threshold for ideologues and extreme responders, slightly more than have historically been cast as ideologues in a strict Conversian sense, but less than we might identify at large as "non-moderate" partisan identifiers through polls and surveys.

Measuring ERS

To identify those with extreme opinions, as well as extreme respondents in a model-based approach, the same strategy for measurement is used as in chapter 1. The measure of

primary interest is extreme response style (ERS), which rivals acquiescence (ARS) in importance as a stylistic response. As shown in the experimental items in chapter 1, extreme response is calibrated by election year using a multi-dimensional nominal response model as outlined in Bolt and Johnson (2009). The data are part of the flagship American National Election Studies (ANES) series through the cumulative data file compilation, arguably one of the core datasets in political research. The ANES data are replete with scaled Likert items that measure attitudes toward political objects and policy issues. The items here are again "feeling thermometer items" that have been examined extensively, and scaled to gage strength of partisan political feelings (Jackman and Sniderman 2002; Brady 1989; Brady 1990; Brady 1990a).

The design of the measure in this chapter and the previous chapter contains a couple very desirable analytical properties. First, it captures a mental process or processes that are unique in that respondents do know what their preferences are literally, yet have special difficulty marking an appropriate intensity, and sometimes refuse to do so. While there is some room to speculate on what this particular process might mean below the surface, on its face it has a clear meaning: the tendency to select endpoints on Likert items independent of the object of measure (Paulhus 1991). Second, and relatedly, it is a deliberately constructed measure that is importantly independent of political dimensions, and as plausibly exogenous a measure of survey compliance and cognition with respect to politics as we might ask. Counting the frequency of "don't know," "refusal," and endpoint responses cannot make as strong a claim to independence from content. Third, *why* we might need or want to know this comes from some doubt about the reliability of self-reports. Asking respondents directly what their priorities, knowledge, or deliberative capacities are may not be a reliable exercise if the political objects

they are evaluating are not familiar to them or do not resonate with them. A prerequisite skill for comparing priorities is the correct identification and conceptualization of the political concepts in question−issues, figures, and political groups− and asking directly will not be reliable. Observed behavior, however, can be an important, revealed indicator of low deliberation.

The meaning of ERS then, with this model, is presumptively: the portion of extreme response attributable to extremity that is independent of Political Evaluation (PE) and follows a constrained pattern of symmetry from a midpoint selection. From an individual standpoint, a pattern of response behavior is identified that communicates that "real" extremity on either political evaluation or social welfare, captured by a "real" tendency to choose endpoints that is common among items used to measure these dimensions, yet extremity that is non-substantive will be captured by a common endpoint selection tendency outside of this definition. Given the associations that these might exhibit,[15] the presumption is that the meaning of ERS here is a "best estimate" of classic, non-substantive ERS, a cognitive strategy employed by respondents on surveys that stems, descriptively, from a relative lack of knowledge or engagement with the question.

To examine differences in how ERS is used by question, endpoint selection propensity for each focal item is used as the indicator of issue "strength or priority" for each focal item, such that a high extreme response (ERS) independent of the objects of measure (Political Evaluation) indicates less knowledge and priority on an issue and a low ERS indicates more comprehensive thought and priority in the public. For each target item, the probability of selecting an endpoint is a function of general ERS such that, as ERS increases, selecting an

[15] Classic associations by political knowledge, income, and education being key hallmarks (Appendix)

endpoint becomes more probable. The general slope of this increase in endpoint selection is the strength or informativeness of ERS in predicting endpoint selection. Sharper rises in endpoint selection indicate a more pronounced use of ERS in a specific question (and for different segments of those who exhibit this behavior).

The formal specification follows as a nominal response model in three dimensions where *i* denotes the item, *j* the respondent, *k* the answer category, and *h* is a summation operator across categories for the same item:

$$P\left(X_{ik}\middle|\theta_{1j},\theta_{2j}\right)=\frac{\exp\left(a_{ik1}\theta_{1j}+a_{ik2}\theta_{2j}+c_{ik}\right)}{\sum\exp\left(a_{ik1}\theta_{1j}+a_{ik2}\theta_{2j}+c_{ik}\right)} \quad (2.1)$$

Each item is allowed to relate in the model, in order to distill the ERS estimates with minimal substantive bias. Importantly, the questions are chosen such that each question is introduced in every year, the feeling thermometers are balanced in their partisan direction (Liberals are offset by Conservatives and so on), lending face validity and continuity.

Similar to chapter 1, the model is identified by setting extreme responses to an equal and increasing propensity (1.0) and also identifies equality in each question by distance from the midpoint for other categories−the nearest two responses to a midpoint set equal to each other, as are near extremes, and so on. The calibration samples consist of ten thermometer items to gauge partisan affect and ERS is identified on all scaled Likert items.

Different from chapter 1, however, the thermometer items I use are the same concurrent items (i=10) in every ANES administration and the underlying scales are forced to be independent of one another (r[θ_{1j}, θ_{2j}]=0 : political, partisan feeling scale and extreme response scale are unrelated) so that respondents are more equally distributed between liberals and conservatives and will have little imprint of partisan bias by design. This insures that an extreme response style is completely independent of political partisanship.

Strong Opinion and Strong Response

Having constructed this model, as before, the results of inspecting the normative 10th percentile procedure are presented to see if it comports with what we believe are the properties of respondents that we expect, that is, roughly balanced on ideology by design, higher among minority populations, low income, low education, and those who are less politically informed. As a check of assumptions, a 20 percent threshold was also introduced and examined, which may include less extreme individuals, but include a larger sample for comparison.

Table 2.2. Frequency of Extreme Response and "Extreme Views"

10 Percent Identification

ANES Year	Extreme Response Only %	Extreme Views Only %	Neither %	Both %
1992 (n=2,485)	9.1	9.2	80.8	0.9
1996 (n=1,714)	9.0	9.0	80.9	1.0
2000 (n=1,807)	8.8	8.9	81.1	1.2
2004 (n=1,212)	8.6	8.6	81.4	1.5
2008 (n=2,322)	9.1	9.2	80.8	0.9

20 Percent Identification

ANES Year	Extreme Response Only %	Extreme Views Only %	Neither %	Both %
1992 (n=2,485)	16.1	16.1	63.9	3.9
1996 (n=1,714)	15.9	16.0	64.0	4.1
2000 (n=1,807)	16.1	16.1	63.9	3.9
2004 (n=1,212)	15.0	15.1	64.9	5.0
2008 (n=2,322)	16.3	16.4	63.6	3.7

Table 2.2 shows the division by design of 10 percent Extreme Respondents and 10 percent division on the strongest ideological preferences, or political evaluation and expanding to include 20 percent in the second portion of the table. The demographic composition of these extreme responders and those who hold more extreme opinions is shown in tables 2.3 and 2.4. As expected, Black and Hispanic groups are overrepresented as extreme responders, as well as Low Income, individuals with lower education, and those assessed as having low political awareness. At least on its face, consistent with the experimental assessment, extreme

responders conform to what we might expect and in the intensity we might expect of them. Additionally, as 20 percent of the sample is included, the imbalance tends to moderate somewhat, yet still maintains a fairly high level of extreme response relative to the baselines of 10 percent and 20 percent. Given the research on extreme response style, it appears as if (a) the demographic data suggest the measurement model is on the right track and (b) that an adjustment based on these demographics would be required in order to balance the sample through weighting so that a "missing completely at random" assumption would be an incorrect way on adjusting for any prospective bias.

Table 2.3. Demographics of Extreme Response (Race and Income)

10 Percent Identification

ANES Year	Black %	Hispanic %	Low Income %	Middle Income %	High Income %
1992 (n=2,485)	18.2	13.8	15.4	7.6	5.7
1996 (n=1,714)	25.3	15.2	14.9	7.9	5.9
2000 (n=1,807)	18.1	18.6	13.2	8.4	5.5
2004 (n=1,212)	17.1	13.6	11.8	9.8	7.7
2008 (n=2,322)	18.5	10.2	14.9	6.6	5.6

20 Percent Identification

ANES Year	Black %	Hispanic %	Low Income %	Middle Income %	High Income %
1992 (n=2,485)	31.4	28.9	28.1	15.4	14.3
1996 (n=1,714)	37.8	26.8	29.0	15.7	12.8
2000 (n=1,807)	37.1	31.0	25.8	16.4	12.5
2004 (n=1,212)	24.6	25.0	21.7	19.6	18.0
2008 (n=2,322)	32.2	22.1	26.2	15.7	14.2

Table 2.4. Demographics of Extreme Response (Education and Political Awareness)

10 Percent Identification

ANES Year	High School or Less %	College Degree or Higher %	High Political Awareness %	Average Political Awareness %	Low Political Awareness %	Very Low Political Awareness %
1992 (n=2,485)	13.6	5.8	8.2	9.1	13.1	17.0
1996 (n=1,714)	14.2	7.3	8.7	7.4	11.9	24.2
2000 (n=1,807)	11.8	8.3	9.2	8.7	11.9	12.9
2004 (n=1,212)	13.1	8.6	9.8	9.7	10.4	13.2
2008 (n=2,322)	11.5	6.9	6.7	11.5	11.2	14.7

20 Percent Identification

ANES Year	High School or Less %	College Degree or Higher %	High Political Awareness %	Average Political Awareness %	Low Political Awareness %	Very Low Political Awareness %
1992 (n=2,485)	25.2	13.5	16.1	18.3	25.0	37.6
1996 (n=1,714)	26.8	15.3	19.1	16.8	21.8	35.5
2000 (n=1,807)	23.8	17.9	17.7	17.8	22.8	29.2
2004 (n=1,212)	24.5	16.9	21.0	18.0	21.2	22.4
2008 (n=2,322)	22.4	15.0	17.4	20.4	20.5	28.8

Table 2.5. Demographics of "Extreme View" (Race and Income)

10 Percent Identification

ANES Year	Black %	Hispanic %	Low Income %	Middle Income %	High Income %
1992 (n=2,485)	2.8	5.3	6.0	12.5	12.7
1996 (n=1,714)	2.8	6.3	6.8	10.1	14.4
2000 (n=1,807)	4.3	4.4	7.9	10.5	13.7
2004 (n=1,212)	3.2	2.3	5.2	11.6	16.0
2008 (n=2,322)	6.9	4.9	5.7	11.0	17.9

20 Percent Identification

ANES Year	Black %	Hispanic %	Low Income %	Middle Income %	High Income %
1992 (n=2,485)	9.7	11.2	13.7	21.5	26.3
1996 (n=1,714)	10.6	16.1	15.8	20.1	25.9
2000 (n=1,807)	10.0	12.4	16.5	18.9	27.7
2004 (n=1,212)	5.9	15.9	11.6	22.6	30.9
2008 (n=2,322)	13.9	10.2	12.6	22.4	31.9

Table 2.6. Demographics of "Extreme View" (Education and Political Awareness)

10 Pecent Identification

ANES Year	High School or Less %	College Degree or Higher %	High Political Awareness %	Average Political Awareness %	Low Political Awareness %	Very Low Political Awareness %
1992 (n=2,485)	5.0	9.4	17.8	7.7	3.6	0.0
1996 (n=1,714)	6.0	10.6	18.4	9.4	1.6	0.0
2000 (n=1,807)	5.6	11.9	18.2	10.6	2.6	0.0
2004 (n=1,212)	4.9	7.3	17.8	5.9	2.8	2.6
2008 (n=2,322)	7.1	10.0	18.2	7.9	3.9	0.6

20 Percent Identification

ANES Year	High School or Less %	College Degree or Higher %	High Political Awareness %	Average Political Awareness %	Low Political Awareness %	Very Low Political Awareness %
1992 (n=2,485)	11.9	20.6	30.9	17.8	9.9	1.8
1996 (n=1,714)	14.0	21.2	32.5	19.9	7.5	0.8
2000 (n=1,807)	14.3	24.4	32.0	22.8	7.7	1.2
2004 (n=1,212)	12.9	14.8	32.1	15.8	6.6	2.6
2008 (n=2,322)	14.2	25.4	33.7	18.0	7.8	0.6

Tables 2.5 and 2.6 round out a rudimentary look at those who hold more extreme, or strong, ideological views. In contrast to those who are identified with ERS, those who hold more extreme ideological views are less often minorities, less often designated low income, less often in lower education categories, and less often identified as having low political awareness. In sum, the demographic markers that are associated with extreme response style are the opposite markers that predict extreme views or ideologues. This also suggests the utility of these factors in making population adjustments when looking at this subject matter. Those who have strong ideology, with these opposite observable markers, may exhibit strong sample bias in their effects simply due to the character of those who exhibit strong ideologies.

Strong Opinion, Strong Response, and Behavior

With these observations of who exhibits extreme response style, however, is the question of whether this trait is somehow influential "in action;" that is, if those who feel

strongly but not consistently with a left/right divide will exhibit stronger associations in political behavior. If the extreme response scale constructed is either contaminated with a latent variable that is not simply response or if there is true opinion mixed in with extreme response, then we might expect this to be the case. Subjects who feel strongly, as opposed to those who are more ambivalent, might be expected to exhibit more active behavior if the strength is of consequence and not simply a matter of incomplete response.

Table 2.7 shows that this is largely not the case. Those with extreme views (Figure 2.1), or the strongest ideological opinions, are categorically more inclined to act, presumably upon these strong views. This difference is both strongly significant and highly meaningful. In particular, the likelihood of voting in these ANES sample varies from 62.0 to 67.2 percent, the vote likelihood of those with strong opinions is squarely above 90 percent. Contrast this observation with extreme response style (Figure 2.2), and we can see the difference most clearly. Those who have extreme response style are, on balance, somewhat less likely to vote than the population as a whole, if at all.

Figure 2.1. Voting: Strong Views (Top 10 percent) vs All Respondents by Year

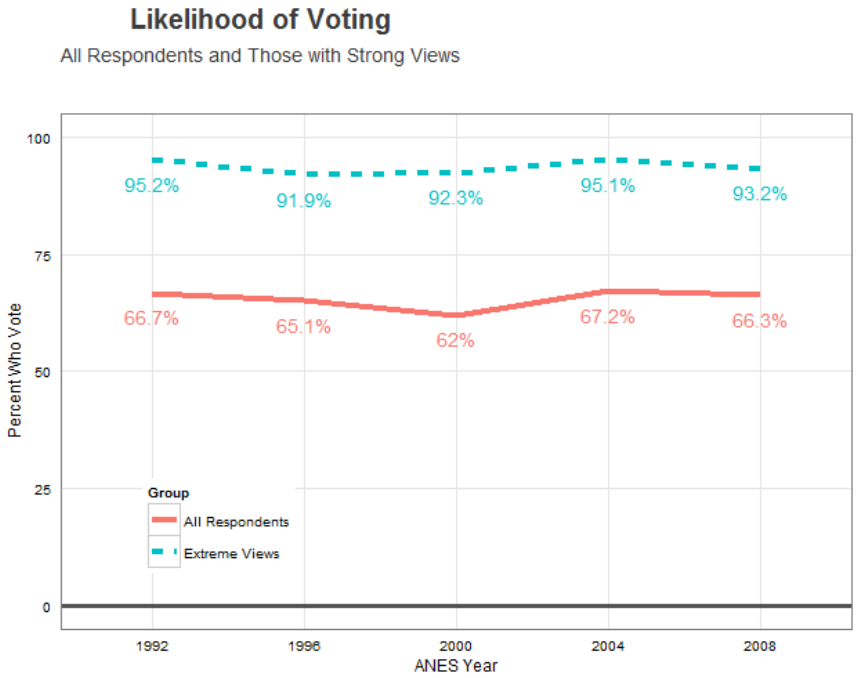


Figure 2.2. Voting: Extreme Response (Top 10 percent) vs All Respondents by Year

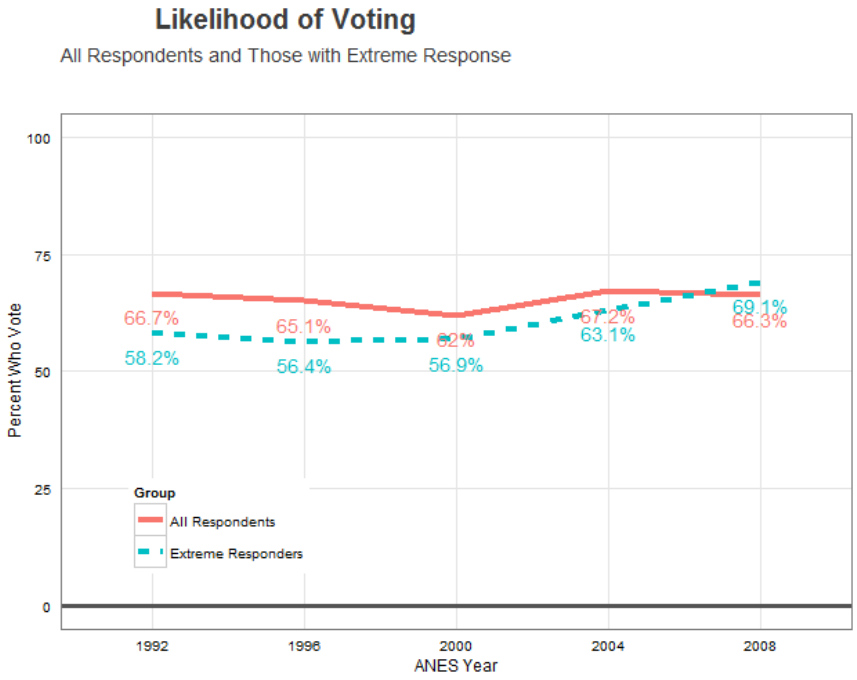


Table 2.7. Extreme Responders, Extreme Views, and Behavior

10 Percent Identification

		Vote %	Tried to Influence Others %	Attended Political Meetings %	Donated Money %
1992	Extreme Responders	58.2**	38.9	4.6*	4.2*
	Extreme Views	95.2***	67.1***	16.5***	14.9***
	Total	66.7	37.6	8.1	7.3
1996	Extreme Responders	56.4**	34.0	6.3	14.6**
	Extreme Views	91.9***	55.0***	13.5***	21.1***
	Total	65.1	28.9	5.9	8.9
2000	Extreme Responders	56.9	39.0	7.1	14.3**
	Extreme Views	92.3***	57.5***	12.7***	17.7***
	Total	62.0	35.1	5.5	9.3
2004	Extreme Responders	63.1	59.6**	12.5**	13.5
	Extreme Views	95.1***	79.5***	18.9***	40.2***
	Total	67.2	48.5	7.6	13.5
2008	Extreme Responders	69.1	53.5***	12.9**	14.3
	Extreme Views	93.2***	68.2***	20.6***	31.3***
	Total	66.3	43.0	9.3	11.1

20 Percent Identification

		Vote %	Tried to Influence Others %	Attended Political Meetings %	Donated Money %
1992	Extreme Responders	56.9***	37.3	7.5	5.8
	Extreme Views	90.7***	59.1***	14.4***	13.4***
	Total	66.7	37.6	8.1	7.3
1996	Extreme Responders	57.7***	36.3***	7.9	12.6**
	Extreme Views	89.2***	49.7***	12.6***	17.5***
	Total	65.1	28.9	5.9	8.9
2000	Extreme Responders	50.8***	38.2	6.9	11.5
	Extreme Views	87.3***	53.2***	11.2***	18.4***
	Total	62.0	35.1	5.5	9.3
2004	Extreme Responders	65.8	60.0***	14.0***	22.0***
	Extreme Views	95.9***	73.8***	16.4***	32.8***
	Total	67.2	48.5	7.6	13.5
2008	Extreme Responders	67.1	47.8**	13.0***	13.7*
	Extreme Views	89.3***	65.3***	17.6***	24.9***
	Total	66.3	43.0	9.3	11.1

Chi-Squared test for homogeneity applied
* - significant at 10% level
**- significant at 5% level
***-significant at 1% level

The presence of some inconsistency notwithstanding, by and large, these results point to an effective separation of "true" strength of views as opposed to strength which is not indicative of a true underlying view. At most, the preceding tables and figures reinforce that there is a

subset of those who answer strongly to survey questions who do not transfer this strength of conviction into political action.

Extreme Response Bias: Strong Partisanship as a Predictor of Voting

Having inspected the validity of the ANES models by year, it appears that the attempt to isolate the trait of response style has largely succeeded in producing what we expect from extreme response independent of political evaluation: a strong association with expected demographic characteristics, and weak or inconsistent associations with variables that indicate political action. These data also suggest that assuming these data are missing completely at random is likely not a prudent assumption.

Therefore, the remaining, and focal, question is whether there is a substantial bias in these ANES samples with respect to an expected result, namely does strong partisanship affect a respondent's likelihood of voting? Several notes accompany this "new question": (a) it is not an exact match of the previous indication of extreme views according to scaled thermometer items, but is a "summary self-report" of strong partisanship. And while ideologues may react consistently with respect to partisan political figures and traditional partisan groups, many do, in fact, deviate when asked to summarize their views. The second note is (b) partisan political self-report, importantly, is a result of a branching item. A respondent is asked to identify as a partisan, and then the strength of identification is a follow up question. As such, there should be no response style associated with this strength.

With a two-phase design, I estimate a simple logistic regression model of voting as a function of Strong Partisanship, toward the expected relationship that strong partisans are more likely to vote. To get an idea of the potential bias, I estimate a model for each ANES year in which no adjustment for ERS is made ("No Control"). Subsequently, a model is estimated that

excludes extreme responders from the sample at a 10 and 20 percent rate, respectively, a listwise deletion that is appropriate under the conditions of missing "completely at random." However, because we do know that there is a distinct demographic profile that is part and parcel of extreme response, a third strategy is employed that uses inverse probability weights to reassemble the original sample with respect to race, income, education, and political awareness. The reweighted sample is then used to estimate the regression again to inspect the differences. I estimate the sampling probability weights using logistic regression as described by Robins et al (1994) in the R programming language using the survey glm library. In sum, the weights increase representation of cases that are underrepresented by excluding respondents with ERS, and decrease those where excluding cases creates an overestimate. For each year the results are given in Table 2.8.

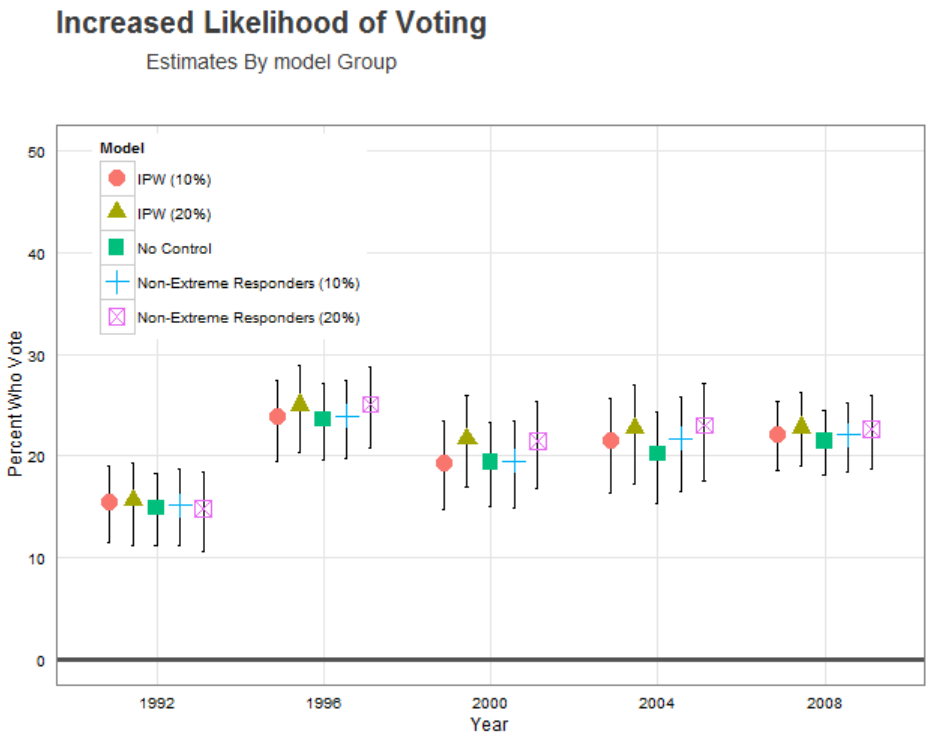
Table 2.8. Strong Partisanship Effect on Voting

ANES Year	Model	Intercept		Slope		Partisanship Percent		Lower Bound Percent (95%)	Upper Bound Percent (95%)	Diff Percent
		Coeff	SE	Coeff	SE	Not Strong	Strong			
1992	No Control	0.51	0.05	0.72	0.10	62.5	77.5	73.8	80.8	14.9
	Non-Extreme Responders (10%)	0.57	0.05	0.76	0.11	63.8	79.0	75.1	82.5	15.2
	IPW (10%)	0.53	0.05	0.76	0.11	62.9	78.4	74.4	82.0	15.5
	Non-Extreme Responders (20%)	0.65	0.05	0.77	0.13	65.7	80.5	76.3	84.1	14.8
	IPW (20%)	0.58	0.05	0.79	0.13	64.2	79.8	75.4	83.5	15.6
1996	No Control	0.31	0.06	1.16	0.12	57.6	81.2	77.2	84.7	23.6
	Non-Extreme Responders (10%)	0.37	0.06	1.22	0.14	59.1	83.0	78.8	86.5	23.9
	IPW (10%)	0.34	0.06	1.19	0.14	58.4	82.3	77.9	85.9	23.9
	Non-Extreme Responders (20%)	0.41	0.06	1.34	0.16	60.0	85.1	80.8	88.7	25.1
	IPW (20%)	0.36	0.06	1.29	0.16	58.9	84.0	79.2	87.8	25.0
2000	No Control	0.23	0.06	0.88	0.11	55.8	75.2	70.9	79.1	19.4
	Non-Extreme Responders (10%)	0.27	0.06	0.89	0.12	56.7	76.2	71.6	80.2	19.5
	IPW (10%)	0.25	0.06	0.88	0.12	56.3	75.6	71.0	79.8	19.4
	Non-Extreme Responders (20%)	0.35	0.06	1.04	0.14	58.6	80.0	75.4	84.0	21.5
	IPW (20%)	0.31	0.06	1.04	0.14	57.6	79.3	74.5	83.5	21.7
2004	No Control	0.44	0.07	1.01	0.15	60.8	81.1	76.2	85.1	20.2
	Non-Extreme Responders (10%)	0.46	0.08	1.13	0.16	61.3	83.0	77.9	87.1	21.7
	IPW (10%)	0.45	0.07	1.11	0.16	61.1	82.7	77.5	86.8	21.6
	Non-Extreme Responders (20%)	0.47	0.08	1.23	0.19	61.5	84.5	79.0	88.7	23.0
	IPW (20%)	0.47	0.08	1.21	0.19	61.5	84.2	78.7	88.5	22.7
2008	No Control	0.39	0.05	1.07	0.11	59.5	81.1	77.7	84.0	21.5
	Non-Extreme Responders (10%)	0.39	0.06	1.11	0.12	59.6	81.7	78.1	84.8	22.1
	IPW (10%)	0.38	0.06	1.11	0.12	59.3	81.6	77.9	84.7	22.2
	Non-Extreme Responders (20%)	0.40	0.06	1.15	0.13	59.8	82.4	78.6	85.7	22.6
	IPW (20%)	0.39	0.06	1.16	0.13	59.5	82.4	78.5	85.7	22.8

One way to look at the results is by inspecting the increase in the likelihood of voting from an unadjusted condition to the adjusted condition. Figure 2.3 is such a look, showing each coefficient relative to each other and showing the coefficients' standard errors. From this basic look, the story is one of similarity. The difference between years is generally larger than any model adjustment may make, and the standard error of the coefficients largely overlap. It becomes apparent that the amount of bias we might be looking at is quite small, but the magnitude in this case tells us only one part of the story. Because both voting and strong partisanship will not contain extreme response, any consistent deviation may be small, but no less meaningful depending on context. Accepting what results we see as a construct-independent tendency, and that it may spill over into content, another look may be appropriate

which checks the consistency of result and how the results change with alternate sensitivity to exclusion and with weighting for excluded responses.

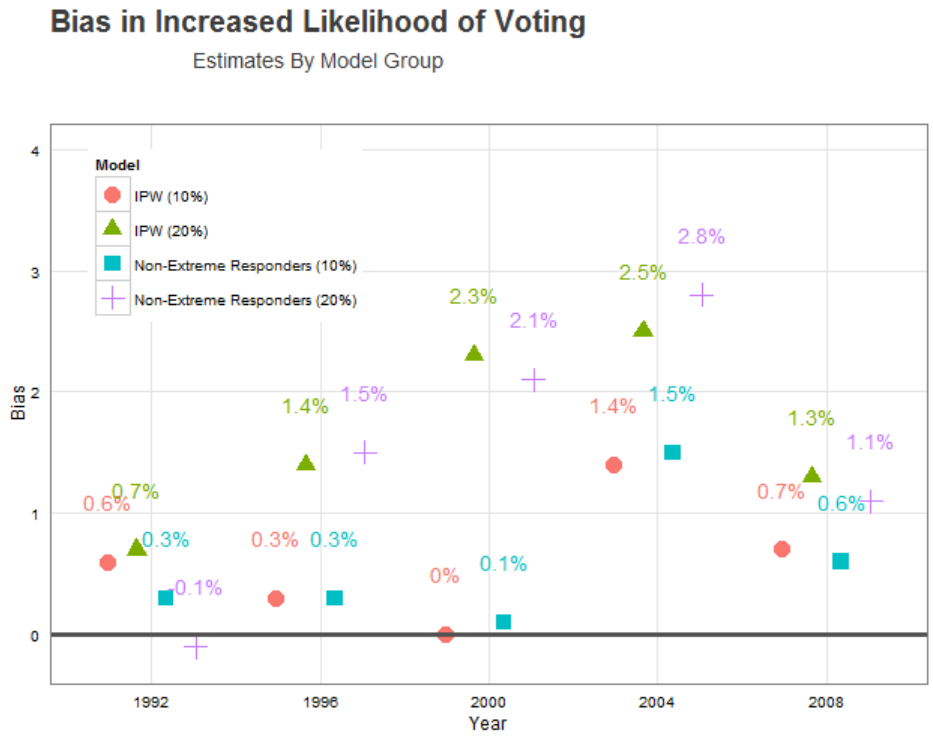
Figure 2.3. Increased Likelihood of Voting−Bias Estimates by Model



Even though the bias we observe is small compared to variance of model year, it may be more informative first to inspect whether the small effect here is consistent across years, second, whether the demographics used in constructing inverse probability weights mitigate the effect of sample adjustment on the coefficients we observe, and third, whether or not a broadening of the definition of extreme response from 10 to 20 percent of the sample exacerbates or mitigates what biases we might see in these models. Figure 2.4 speaks to these three questions, calibrated to the extra effect above and beyond the models with "No Control" for extreme response. In essence, this compares a "do nothing" approach to approaches which assume missing completely at random (exclude extreme respondents) and missing at random

(exclude respondents and reweight sample based on observables that have proven significant, both in prior research and in this dissertation.

Figure 2.4. Omitting Extreme Response Creates Downward bias on Estimates Outside of Intended Target: Increased Likelihood of Voting−Bias Estimates by Model



The first observation from Figure 2.4 is that nearly all the effects of strong partisanship on voting are more positive when controls are made for extreme responders. From the previous table, this might be foreshadowed in some part. When a large segment of subjects is eliminated from analysis, and ostensibly they have no effect (or negative effect) on voting, the remaining data (the square points) will necessarily have a higher propensity for voting overall. What is not intuitive, however, is the effect of partisanship increasing: the slope coefficient. That might lead us to expect lower intercept values for voting, yet that does not happen. In fact, those without strong partisanship actually have a higher likelihood of voting when these observations are removed. It bears repeating too that this relationship, neither in voting nor in partisanship, has any explicit relationship with extreme response at all. One explanation is that a certain kind

of individual is omitted for whom this relationship is not as strong: there is contamination by a dimension that is common among years and related in the same way as response style. Another possible explanation here for the modest effect is that, by suppressing those with extreme responses, *error is removed from other survey responses that are not themselves extreme responses*. Those who prefer endpoint observations, that is, might employ other tendencies to respond that are not opinion expressed as we expect and are more akin to measurement error than bias. Additionally, through the IPW, the coefficients are not substantially altered (save for 1992). By weighting for covariates in these samples, covariates that we know to be related to response style (the circular points), not a lot is gained or lost in estimating coefficients. The conclusion might be, if we accept the hypothesis that the measure reflects something substantive other than noise, that what we are seeing as bias is not a simple reflection of "who is excluded" (extreme response is embedded in something that is simply demographic or a function of awareness).

This general tendency of negative bias (not controlling for ERS subdues a true effect in the population) is repeated for the 20 percent exclusion sample, only magnified. The 10 percent exclusion samples seem to be the models that have modest bias, mostly below 1 percentage point in their effect. Excluding at a 20 percent rate increases the bias.

Discussion/Conclusion

Previously, I had used some simple simulations to demonstrate likely bias due to extreme response and how it might affect overall response frequency in issue items. Here, I extend that examination of bias by examining the effects of that bias by overtly controlling for ERS, and inspecting the difference between raw regression results and regression results using the conditioned, re-weighted data set.

In terms of validity of the extreme response model, the demographic covariates confirm what we expect to see with respect to race, income, education, and political awareness. Political extremity also behaves like we expect, tilted in the opposite direction in every respect; toward white respondents, higher income, higher education, and political awareness. The completely contrasting character of these respondent profiles also suggest that there may be good reason to believe that extreme respondents will skew results simply due to the very different nature of their populations. Those who are genuinely extreme and those who are simply picking extreme response are very different in character.

This validity check by demographics is reinforced by looking at actual political behavior between those who respond extremely and those who hold extreme (or strong) substantive political views. Those with stronger views are much more likely to engage in political behaviors: voting, trying to influence others, attending meetings, and donating money. As for extreme "respondents," the behavior profile is much different, largely having little effect: voting behavior may see a decrease in some years, and increases in other behaviors are seen in some years and not others. Sometimes these are noteworthy deviations, yet it is difficult to reason that some political campaign activities should be suppressed for a group while others will be increased. This may suggest that somehow, between years, certain behavior profiles are being picked up and may deserve separate analyses. Perhaps a choice between two candidates where strong conflicting actual ideas may be unpalatable, yet meeting attendance and monetary contribution are seen as more viable outlets for conflicting partisan feelings.

Finally, by controlling for extreme response, excluding those responses who do not conform to the survey construct as we might expect, and by conducting a substantive analysis

that has both strong expectations (strong partisanship will associate with higher rates of voting) and is free, itself, of a direct relationship of extreme response, we see a small, but consistent negative bias. That is, when extreme respondents are viewed as incomplete respondents, the relationship between strong partisanship and voting becomes stronger. It does so in a way that also suggests that extreme response operates in a substantive way that is independent of its demonstrated effects by demographic group. For example, if we expect low income individuals to have substantially more extreme response, and are more often identified as displaying these tendencies, yet are less often strong in their ideological views, the results should be responsive to a weighting adjustment. Yet, when taking into account variables such as race and income, reweighting the sample using these variables simultaneously, and inspecting results, there is very little evidence that coefficients meaningfully change. The suggestion, at least, implies that "something" is happening in the sample, that "something" has to do with response style itself (or some vestige thereof), and that something is also operating largely independently of these fixed population attributes.

The solution here is an important step. At the very least, we might say that sample correction is absolutely warranted in cases where outcomes or covariates of interest are directly affected by offending items, including, but not limited to scales that consist of offending items. However, even if there is no direct relationship with an offending item, response style adjustment may be advantageous, and this advantage may be heightened in analyses that transfer to other ambiguous constructs, require more complex cognitive tasks, have systematic demographic bias in one or more areas identified here, or rely on small coefficients that might lend significance in large samples or in important contexts. Conservatively, alternate analyses, setting aside offending cases may go some way toward correcting this bias, and in theory,

inverse probability weighting or some other reweighting strategy is necessary in order to overcome the difficulty with subjects being missing at random.

A 2.1. Latent Gold Model Syntax

```
model
title Model1;
options
  algorithm
    tolerance=1e-008 emtolerance=0.01 emiterations=250 nriterations=50;
  startvalues
    seed=0 sets=10 tolerance=1e-005 iterations=50;
  bayes
    categorical=1 variances=1 latent=1 poisson=1;
  montecarlo
    seed=0 replicates=500 tolerance=1e-008;
  quadrature  nodes=10;
  missing  excludeall;
  output
    parameters=effect standarderrors probmeans=posterior profile bivariateresiduals;
variables
  dependent VCF0218o nominal, VCF0224o nominal, VCF0209o nominal, VCF0210o
nominal,
    VCF0211o nominal, VCF0212o nominal,
    VCF0424o nominal, VCF0425o nominal, VCF0426o nominal, VCF0427o nominal;
  latent
    CFactor1  continuous,
    CFactor2  continuous,
    Cluster nominal 1;
equations
  (1) CFactor1 ;
  (1) CFactor2 ;
  CFactor1 <-> CFactor2 ;
  Cluster <- 1;
  VCF0218o <- 1 + Cluster + (a1)CFactor1 + (b1)CFactor2;
  VCF0224o <- 1 + Cluster + (a2)CFactor1 + (b2)CFactor2;
  VCF0209o <- 1 + Cluster + (a3)CFactor1 + (b3)CFactor2;
  VCF0210o <- 1 + Cluster + (a4)CFactor1 + (b4)CFactor2;
  VCF0211o <- 1 + Cluster + (a5)CFactor1 + (b5)CFactor2;
  VCF0212o <- 1 + Cluster + (a6)CFactor1 + (b6)CFactor2;
  VCF0424o <- 1 + Cluster + (a7)CFactor1 + (b7)CFactor2;
  VCF0425o <- 1 + Cluster + (a8)CFactor1 + (b8)CFactor2;
  VCF0426o <- 1 + Cluster + (a9)CFactor1 + (b9)CFactor2;
  VCF0427o <- 1 + Cluster + (a10)CFactor1 + (b10)CFactor2;



// constrain ERS
// set non-response equal to 0
```

```
// first category is zero (0) [x,1]
  a1[1,1]=0;
  a2[1,1]=0;
  a3[1,1]=0;
  a4[1,1]=0;
  a5[1,1]=0;
  a6[1,1]=0;
  a7[1,1]=0;
  a8[1,1]=0;
  a9[1,1]=0;
  a10[1,1]=0;

// missing are 0
  b1[1,1]=0;
  b2[1,1]=0;
  b3[1,1]=0;
  b4[1,1]=0;
  b5[1,1]=0;
  b6[1,1]=0;
  b7[1,1]=0;
  b8[1,1]=0;
  b9[1,1]=0;
  b10[1,1]=0;


// fix extremity on ERS variables
  a1[1,2]=1;
  a2[1,2]=1;
  a3[1,2]=1;
  a4[1,2]=1;
  a5[1,2]=1;
  a6[1,2]=1;
  a7[1,2]=1;
  a8[1,2]=1;
  a9[1,2]=1;
  a10[1,2]=1;

  a1[1,9]=1;
  a2[1,9]=1;
  a3[1,9]=1;
  a4[1,9]=1;
  a5[1,9]=1;
  a6[1,9]=1;
  a7[1,9]=1;
  a8[1,9]=1;
  a9[1,9]=1;
```

```
a10[1,9]=1;


a1[1,3]=a1[1,8];
a2[1,3]=a1[1,8];
a3[1,3]=a1[1,8];
a4[1,3]=a1[1,8];
a5[1,3]=a1[1,8];
a6[1,3]=a1[1,8];
a7[1,3]=a1[1,8];
a8[1,3]=a1[1,8];
a9[1,3]=a1[1,8];
a10[1,3]=a1[1,8];

a1[1,4]=a1[1,7];
a2[1,4]=a1[1,7];
a3[1,4]=a1[1,7];
a4[1,4]=a1[1,7];
a5[1,4]=a1[1,7];
a6[1,4]=a1[1,7];
a7[1,4]=a1[1,7];
a8[1,4]=a1[1,7];
a9[1,4]=a1[1,7];
a10[1,4]=a1[1,7];

a1[1,5]=a1[1,6];
a2[1,5]=a1[1,6];
a3[1,5]=a1[1,6];
a4[1,5]=a1[1,6];
a5[1,5]=a1[1,6];
a6[1,5]=a1[1,6];
a7[1,5]=a1[1,6];
a8[1,5]=a1[1,6];
a9[1,5]=a1[1,6];
a10[1,5]=a1[1,6];

// For testing new item, set endpoints to equality
//a11[1,1]=a11[1,4];
//a11[1,2]=a11[1,3];
end model
```

Chapter 3. ERS and Political Issues: Extreme Response Style (ERS) Differences in the American National Election Study (1992-2008)

This chapter begins to consider extreme response in the context of different political issues under the umbrella of a right to left ideology, the ideology that comes from symbolic feelings toward partisan political groups and people. To do so requires a deeper consideration of what it means to respond extremely. If respondents are using some cognitive strategy to answer items and this cognitive strategy is transferred to different situations in different ways, it may appear in some items and not others, and at some times and not others. This is the crux of the situational hypothesis that undergirds what might be categorized as response set or style, the nomenclature for which varies by discipline.

Building on prior work, if extreme response style can be reliably estimated, has effects in scaling and not branching items, and controlling for its effects yield bias that is not meaningfully captured entirely by demographics, the avenue is open to considering ERS as a contextual device to answer questions. The issues used to define extreme response mirror the issues examined in chapters 1 and 2, but with a slightly more elaborate procedure, utilizing the strength of an anchor scale of feeling thermometer items, the addition of a scale of issues, and eliminating any attribution of style to substance by constraint, where extreme response should not be applied in items that do not scale. Thus, we might improve the base measure used previously, and look to extend its potential inference.

The previous chapters have been more methodological and confirmatory in nature: an introduction to response style, an experimental assessment of ERS in chapter 1, and an impact assessment in chapter 2. In chapter 1, the case for the measurement of ERS was tested in as strict a causal sense as possible and the case for response style was reinforced as a meaningful indictor of knowledge and engagement with specific subject matter. That is, specific

knowledge did appear to drive the "range" of response choice (the tendency to pick categories close to an extreme or a midpoint) independent of the object of measure, and this response was differentially detectable between issues. This section makes the theoretical jump in accordance with research on optimization behavior[16], inspecting whether we might measure cognitive engagement with issues and political groups by measuring extreme response tendencies. Importantly, I do this, not simply by asking respondents what they feel might be important, but by observing whether they exhibit behavior that is indicative of cognitive engagement and knowledge in the content of the issue area or around the political group or figure.

As the following sections extend the observations / results in the previous chapters by looking at the possible meaning of response style measures in politics and political cognition/engagement, I inspect the differences in response style between widely-used measures of issue and opinion. This chapter looks at ERS through an extended multidimensional IRT model to further isolate ERS from any potential substantive construct to focus on specific item types and items over time. Some analysts point out that response style, in this case Extreme Response Style, should not vary between items and does not meaningfully change in some psychological question batteries (Weijters 2006). Contrary to this common assertion, I have demonstrated a link between political information and ERS, which, at the very least, suggests that there may be meaningful variance between items and within items over time that may tell us something about how (and what) the public thinks about politics. What the public tells us by "giving up" on some questions over time and less others may provide some independent insight into commonly asked political questions shed light onto which electoral skills are more consciously present, especially for the least involved.

[16] Optimization is the terminology used for the thoroughness of a thought process from processing a question's subject matter to the point of response selection.

The methodological chapters preceding have provided some referendum on the question: issues do exhibit meaningful differences in extreme response style. From a methodological and measurement standpoint, two schools of thought exist about the nature of any such tendency to respond extremely. One contends that response style is *invariant*. That is, a response style applied on question *i* is applied uniformly on question *i+1* and is structural in nature: response style is largely independent of question content. Weijters (2006) in his treatment of response style in consumer research makes the invariance argument. The other argument is that the subject matter and extreme response are interactive (De Jong et al 2008), that is, ERS exists as a trait that is "drawn out" by the question itself. Combined with the *optimization* hypothesis of response (Krosnick 1999), questions that are more accessible or more important to the respondent will dampen automaticity of response, and ERS will be dampened as a result. Conversely, questions that are ambiguous or obscure will evoke response defaults much more often, and with more strength.

The Meaning of Extreme Response Style

The latent trait of extreme response style (ERS) presents a unique difficulty in interpretation. An explicit refusal or a self-report that acknowledges that a subject does not have the requisite knowledge to complete a question as desired ("I don't know") is a simpler proposition in theory. When a respondent receives a question, looks into available resources to answer, and finds little there, we might take this at face value. Response style, however, contains no such explicit roadmap by which it might be interpreted meaningfully. It remains a kind of black box into which we might project a number of related traits. This requires a little more digging and hypothesizing if we are to differentiate between what response style is likely to mean. To do so, I repeat what is an increasingly consensual definition from which we might

build. The direct behavior is the tendency to choose endpoints independent of the object of measure.

Indirectly, however, we might extend this meaning to be more inclusive of the reasons why this direct behavior exists, and find it instructive to explore what these reasons imply. In that light, I reach to cognitive science for answers. First, this category of response may represent what we might call a truncated decision process in which extreme, strong responses are used as a default response without regard to content. As such, it might be a unique way of showing how issues or political objects elicit kneejerk response more than others.

This line of reasoning has a good deal of merit for how cognition is thought to work in political science. There has been a long tradition that focuses on lack of consideration that is exercised in the mass public when contemplating political ideas and affiliation. As far as political issues and ideology, this tradition has its iconic beginnings in Converse (1964), showing how the electorate behaves when confronted with surveys of political opinion. Subsequently, kinder renditions show that mental sampling (Zaller 1992) or measurement error (Achen 1975, Ansolabehere 2008) may overstate the public's apparent lack of knowledge. Kinder still is that political decisions require very little thought, and that the rational option for voters is to rely on simple signals to stand in for more complex, perhaps unnecessary thought (Popkin 1992).

In discussing what is measured in ERS as meaningful and indicative of a truncated thought process, I might start out by acknowledging some room for interpretation, just as we might interpret a "don't know" or a "refusal" and look at their relative rates or just as we might interpret any latent variable that is constructed to measure personal attributes that are not

directly observable through one response, but evident through a pattern of responses across multiple questions. Some respondents simply do not know, and that is an admission from a mental search and finding relatively little resource. I have discussed several variants and nuances of what an extreme response trait might represent, as measured in this chapter and in previous ones. I have also discussed how it might complicate the interpretation in the following chapter's analysis.

A hardened skeptic may feel that data exploration in this way may be more akin to relating issue choice, partisanship, and voting to traffic infractions or some other personality characteristic, which mandates an extra step, validating the properties of ERS at each step. Even if, at a minimum, the reader may refuse to accept any meaningful ascription to the concept of ERS beyond the literal definition by the measures I provide, it is, at the very least, the fidelity of "strength of opinion" that comes into question, and comes into question differently for different opinions. Just as we might not ascribe any specific meaning to "don't knows" or "refusals," yet still note the importance and bias they might contain, we can note with concern how ERS varies between questions shows us strong opinions that are not objectively strong at all and remark how this erstwhile strength varies. If nothing else, ERS represents a source of potential differential bias with which to contend and a serious threat to research that uses common opinion measures in the American National Election Survey (ANES), even while it is not an ignorable fact that ERS, like non-response, contains meaning and regularity that can be usefully interpreted.

But with this minimum, we might consider, however, the summary statement that is a baseline for opinion research and a clear result from chapter 2. Again, inasmuch as we care about the intensity of opinion: *"…strong* attitudes are predictive of *behavior*, *whereas weak*

ones are *not."* (Miller and Peterson 2004, p. 848). With ERS, this axiom is violated such that strong opinions are expressed that are not strong opinions, but vestiges of habitual behavior that are not intense. If we care about this violation, it is, at a minimum, useful to know when strong opinion is not strong opinion and for which issues.

In previous research, extreme response does have certain stable properties. As noted previously, there are differences between cultures, countries in the effect of ERS, and even within countries in different ethnic subgroups – notably among African Americans and Latino minorities in the United States. This follows a line of research that views the tendency to respond extremely as somehow culturally imbedded, and not necessarily a dynamic cognitive mode for dealing with ambiguous content. I have shown how optimization can be a helpful strategy for viewing and interpreting ERS as the byproduct of a truncated deliberation process. Furthermore, findings in this dissertation show that ERS does appear to have meaning for a respondent's thought process about the subject matter in question, notably a respondent's level of political information is associated with extreme response.

In this light, one of the basic findings in this dissertation, and this chapter specifically, reinforces that ERS can be isolated and validated (as here) and does appear to have meaning for a respondent's thought process about the subject matter in question. And, in turn, it is this thought process that colors the subsequent conclusions of substance. Among them are contrasts in how questions and political phenomena vary with respect to this ERS measure and what we may learn from it. The notable results come primarily in the areas of *policy issue and political cognition*. If how the mass public processes different domains of policy, political groups and people contains more ERS, the question will be minimally (a) at greater risk for bias if it is

included in linear regressions or factor analyses and likely (b) indicative of opinions of respondents with lower knowledge, engagement, or deliberation where ERS is present.

Because response style traits are most pronounced for populations that are the least aware, an individual question's meaning takes on additional importance. What do potential "between question" differences mean about the substance of politics that comes out most starkly for this lower information, automatically responding category of respondents? In a derogatory way, whether these respondents are more "true" in their assessments and beliefs or whether a response is simply a fleeting, or automatic impression. What might this population find to be engaging enough, or clear enough to answer in a way that we would like? To go even further, what does this population find engaging enough to first have an opinion at all, but then abandon nuanced judgment when considering how intensely they might feel about a political group or issue?

If *situational response style* is present, the implication is that endorsement of endpoints apart from any meaningful content reveals a differential awareness, both between objects and within objects over time. At first it might be reflective of knowledge in the electorate: high endpoint selection due to ERS will reveal low awareness and engagement among respondents on an item or topic, and low endpoint selection due to ERS will reveal higher levels of competence. More specifically however, the process of endpoint selection is one of the primacy of contextual cues that distort the underlying "utility function" of a response (Böckenholt 2012; Tsestos, Usher, and Chater 2010). As such, a distinction might be made between the "rapid and effortless" judgment of ERS and the "slow, deliberative" judgment of more tempered category options. That is, ERS is the second in an series of response processes, the first representing true lack of knowledge (a midpoint, "don't know" or refusal), while the

second belies a possible contextual certainty that is seen as a pattern without a deliberative process added. As such, it might also be a referendum on a debate about the substance of partisan behavior, such that strong partisanship that is subject to ERS can be, at worst, an abdication of complete thought and compliance, and at best, a more conscious decision to simplify the political environment through the use of its most basic structural characteristic. Therefore, the meaning of this process is perhaps not a hallmark of pure lack of knowledge, but that *endpoint selection due to ERS represents opinion with a lack of calculated deliberation*.

Issues, Political Evaluation, and Response Style

I will start with a supposition of issue thought and priority, such that issues that are more important will deserve more calculated deliberation, and others will not. For example, any issue, on its face, might be reported a high priority issue, yet if it receives very little deliberation, we might call the issue important only nominally. By virtue of receiving less deliberation in the public for those who might respond stylistically, we might view this as evidence that the issue in question is not as deliberated as we might have thought over the time periods in question.

People, as with the issue of Abortion, may feel both strongly about their position, and it may be in the forefront of the public mind as a litmus test for cultural divisions now, yet has not always been the case. As late as the 1980s, abortion as an issue was not particularly important, yet it had grown in importance to be perhaps the popular litmus test issue for behavioral norms or a signal of adherence to traditional cultural values. Another way to look at the problem is available here as a referendum on the question or importance. ERS, as mental accessibility, engagement, or deliberation can tell us what is relevant enough to merit consideration on a nuanced level, potentially read as a measure of importance.

We have asked about issue importance intermittently, yet have not, for whatever reason, pursued importance, priority, or deliberation of issues as a mainstay of issue analysis. One notable exception is the work of John Zaller in *the Nature and Origins of Mass Opinion* (1992) and *The Myth of Massive Media Impact: Revived* (1996). As issues become more widely known, awareness of issues themselves tends to move opinion. While Zaller highlights the middle of the knowledge spectrum in his work, this look at awareness asks a different question of a markedly less informed population, a population for which we are perennially attempting to ask questions and receive very weakly interpretable information. If a respondent is so weakly aware of issues, how might we expect respondents to prioritize importance? In these questions, is it possible to detect changes in the relevance of issues at a very minimal level, by inspecting a specific type of response behavior? What may be seen between issue areas represents another facet of what makes a policy area "easy" (Carmines and Stimson 1980, Carmines and Stimson 1986). The crux of work on issue difficulty, relevance and deliberation, in the public is a question of which issues are difficult to navigate and which are "easy." Between issues, then, and between political objects, an automatic response will be more pronounced for issues that are less relevant, well-considered, or easy. Contrary to what we might think of as importance, the notion that some symbolic issues are easier to digest stands somewhat in opposition to what might be called "importance" or "priority." If a set of issues over time strengthen or weaken in their response style effects, the conclusion might plausibly show that they are less important or more important, respectively, at any given time. Yet by Carmines and Stimson's logic, instead, they are simply more difficult and less difficult to digest in the public due to the symbolism, direct character, and not simply a focus in national dialogue.

Three issue examples provide an initial basis for evaluating priority or importance in this analysis, extended to a purposive sampling of other political objects as well. First, a recurring theme for several decades running has been the American experience with persistent federal budget deficits. Politically speaking, it is a bone of contention, with Republicans routinely calling for lower spending because of it, and Democrats routinely blaming Republicans for shepherding large deficits through their presidential terms. Regardless of where one falls on these issues, the reality in public opinion is, at least, that the deficit and the growing debt are facts of life that command public opinion on spending. Another fact of life is that American perceptions about the role of government are additionally conflicted, such that in combination with the deficit and debt, it is unclear whether the public favors the cuts necessary to bridge the gap, or alternately prefers to find revenue to sustain spending and services at the current level.

One puzzling question in the ANES reinforces this classic American bind: "do you prefer taxing or spending?" and the neat answer to this question defies even a minimal association with ideology. So it would seem that the federal budget deficit should be increasingly important to people, yet when called to task to solve it, the public is conflicted. That is, the public is not just polarized on a solution to deficit spending, but finds a real lack of impetus within itself to confront the problem. Does this signal real importance, then, or is the deficit relatively meaningless? The vehicle for looking at this potential question is the issue of "Spending and Services".

Second, the devolution and re-emergence of the issue of defense in American political priority has undergone some periodic change. From Cold War politics of the pre-80s, we have witnessed a two decade lull in explicitly foreign military preoccupations. From Vietnam to the

fall of the Soviet Union to the bombing of the World Trade Center and the Pentagon in September of 2001, our preoccupation with defense at some times, and not in others, is well documented in public opinion (Claggett and Shafer 2010). Since 2001, defense has been a more pressing issue, especially in the 2004 election where it is arguably the key to the Bush re-election.

Third, the American evolution on the question of cultural values has evolved. For abortion, since Roe v Wade in 1972, abortion-rights or pro-life stances have galvanized and anchored what we might call a latent cultural vein in American political thought and behavior. This vein was not activated overnight, but became an issue over the course of several decades (Carsey and Layman 2002, Adams 2008), evolving into a new playing field for a party-led "culture war" (Hunter 1992, Fiorina et al 2005). However, given the time frame of the abortion issue, another vein of cultural values is examined as well, namely questions about traditional families and moral relativism. Abortion may not exhibit variance over this time period in a way that other issues might as an emerging cultural issue that separates traditionalist and progressive ideological leanings.

With issues as an exemplar, it is not far-fetched to extend this logic to more expansive realms of political cognition as well. This may reveal important clues to how areas of politics have behaved over the last two decades, how cognitive structures behave with respect to political stimuli. How does the mind of the electorate, ostensibly those who are least engaged, work and what do the cognitive and deliberative capacity between major political issues and objects say about us? How do Democrats, Republicans, liberals, fundamentalists, our presidential candidates have in common? How do they differ? How can our decision processes on questions help understand what it means to have political ideas, be exposed to political

people? The question might have important implications for policy areas, elections, and revealing the character of public dialogue and perception.

Measuring ERS

The design of the measure in this chapter and previous chapters contains a couple very desirable analytical properties. First, it captures a mental process or processes that are unique in that respondents do know what their preferences are literally, yet have special difficulty or refusal marking an appropriate intensity. While there is some room to speculate on what this particular process might mean below the surface, on its face it has a clear meaning: the tendency to select endpoints on Likert items independent of the object of measure (Paulhus 1991). Second, it is a deliberately constructed measure that is importantly independent of political dimensions, and as plausibly exogenous a measure of survey compliance and cognition with respect to politics as we might ask. Third, *why* we might need or want to know this comes from some doubt about the reliability of self-reports. Asking respondents directly what their priorities, knowledge, or deliberative capacities are may not be a reliable exercise if the political objects they are evaluating are not familiar or do not resonate. A prerequisite skill for comparing priorities is the correct identification and conceptualization of the political concepts in question, whether issues, figures, and political groups, and asking directly will not be reliable. Observed behavior, however, can be an important, revealed indicator of low deliberation.

As in previous chapters, the measure of issue importance is extreme response style (ERS), which rivals acquiescence (ARS) in importance as a stylistic response. As shown in the experimental items chapters 1 and 2, extreme response is calibrated by election year using a multi-dimensional nominal response model as outlined in Bolt and Johnson (2009). The data

are part of the flagship American National Election Studies (ANES) series through the cumulative data file compilation, arguably one of the core datasets in political research. The data is replete with scaled Likert items that measure attitudes toward political objects and policy issues, the most well-worn of which (and most consistent) are the feeling thermometer objects in this analysis. Again, the strategy separates items into *base* and *focal* items. *Base* items provide the measurement model with desired characteristics and the *focal* item is allowed to vary freely with respect to endpoint response. The base items in these cases are feeling thermometer items for Political Evaluation (PE) (Jackman and Sniderman 2002, Brady 1989, Brady 1990, Brady 1990a) and "Social Welfare" (SW) (Trier and Hillygus 2009, Claggett and Shafer 2010) items from the American National Election Surveys (ANES) cumulative data file, with the focal items by year in this chapter being:

Issues
vcf0839 – Spending and Services
vcf0843 – Defense
vcf0838 – Cultural: Abortion
vcf0853 – Cultural: Traditional Families
vcf0854 – Cultural: Moral Standard Tolerance

Feeling Thermometers
vcf0211 – Liberals
vcf0212 – Conservatives
vcf0218 – Democratic Party
vcf0224 – Republican Party
vcf0224 – Democratic Presidential Candidates
vcf0225 – Republican Presidential Candidates
vcf0232 – Gays and Lesbians
vcf0234 – Christian Fundamentalists

The measurement model tackles ERS by adjusting the model from chapter 1 to address two main concerns: (1) is it possible that ANES has gotten better at asking issue questions face to face over time, thereby producing a decrease in (or erratically behaving) response style? (2)

is it also possible that some issues are less integrated into the partisan debate and carry with them specific item variance due to ERS?

If item clarity and administration improve (1), it would likely come through as a decrease in ERS over time and while it may not be entirely eliminated, it should be possible to see some general improving trends by looking at issues over this extended time-period. If there is a nuisance dimension at work (2), it might be corrected well by introducing a useful tertiary dimension that captures something we might suspect contaminates the ERS measure. To ameliorate this possible difficulty, a third dimension is included that calibrates issues of social welfare. Yet a third strategy might prove productive. Some of the items that constitute Social Welfare are scaled and some are 3-point responses (Social Security increase/decrease, Education increase/decrease, etc.). The extended scale items, therefore, should be allowed to vary with an ERS component, while the 3-point response will be constrained to zero, no effect. The dimensions of Social Welfare and Political Evaluation, naturally, are allowed to associate with one another. The result is a better estimate of ERS using multiple dimensions and introducing questions, like our controls in chapter 1, that have no theoretical bearing on response style, while setting their loading to zero (Bolt and Newton 2011).

The meaning of ERS then, with this model, is presumptively: the portion of extreme response that is attributable to extremity that is separate from Political Evaluation (PE) and Social Welfare (SW) issues, is common to all scaled Likert items in this study, is not common to the unscaled questions, and follows a constrained pattern of symmetry from a midpoint selection. From an individual standpoint, a pattern of response behavior is identified that communicates real extremity on two scales, identified by a real, yet separate, tendency to choose endpoints that is common among affected items, and excluded where it will not

theoretically be present. Given the associations that these might exhibit[17], the presumption is that the meaning of ERS here is a best estimate of classic, non-substantive ERS, a cognitive strategy employed by respondents on surveys that stems, descriptively, from (a) a relative lack of knowledge or engagement with the question, (b) a lack of engagement that indicates a level of knowledge that marginally exceeds "no knowledge," and/or (c) communicates an importance or priority in individuals who are less engaged.

To examine differences in how ERS is used by question, endpoint selection propensity for each focal item is then used as the indicator of issue "strength and/or priority" for each focal item, such that a high extreme response (ERS) independent of the objects of measure (Political Evaluation and Social Welfare) indicates less knowledge and priority on an issue and a low ERS indicates more comprehensive thought and priority in the public. For each target item, the probability of selecting an endpoint is a function of general ERS such that, as ERS increases, selecting an endpoint becomes more probable. The general slope of this increase in endpoint selection is the "strength" or "informativeness" of ERS in predicting endpoint selection. Sharper rises in endpoint selection indicate a more pronounced use of ERS in a specific question (and for different segments of those who exhibit this behavior).

The formal specification follows as a nominal response model in three dimensions where *i* denotes the item, *j* the respondent, *k* the answer category, and *h* is a summation operator across categories for the same item:

$$P\left(X_{ik}\middle|\theta_{1j},\theta_{2j},\theta_{3j}\right)=\frac{\exp\left(a_{ik1}\theta_{1j}+a_{ik1}\theta_{2j}+a_{ik1}\theta_{3j}+c_{ik}\right)}{\sum\exp\left(a_{ik1}\theta_{1j}+a_{ik1}\theta_{2j}+a_{ik1}\theta_{3j}+c_{ik}\right)} \quad (3.1)$$

The progression by year in Table 4.1 represents how each is allowed to relate in the model, in order to distill the ERS estimates with minimal substantive bias. Importantly, the questions are

[17] Classic associations by political knowledge, income, and education being key hallmarks

chosen such that each question is introduced in every year, the feeling thermometers are balanced in their partisan direction (Liberals are offset by Conservatives and so on), and items are introduced in the area of social welfare in which response style will not be present - lending face validity and continuity. Inspecting the item parameters for these base items will also prospectively lend face validity to each construct. Again, the analysis is not as focused on the content of the substantive dimensions as with the ERS component that it identifies.

Another note accompanies these specific measurement models in terms of specification. The sample also excludes African Americans and the Latino population. It is likely, even probable, that culture plays some independent role with respect to response style (Bachman and O'Malley 1984, Marin et al 1992, Cheung and Reinsvold 2000) and these populations specifically exhibit special issue characteristics and relationships in American political life, and in their propensity to respond extremely in prevailing theory. Even though the exact cultural mechanism is undefined for these populations, since the target is otherwise political cognition as a process, these special, empirically relevant cases are removed to isolate this mechanism, even though a full political story might and should contain these populations.

Table 3.1. Items Selected for Multidimensional IRT Model

Variable	Categories	θ_{1j} Extreme Response ERS	θ_{2j} Political Evaluation PE	θ_{3j} Social Welfare SW
vcf0218 Democratic Party FT	9	X	X	
vcf0224 Republican Party FT	9	X	X	
vcf0209 Big Business FT	9	X	X	
vcf0210 Labor Union FT	9	X	X	
vcf0211 Liberal FT	9	X	X	
vcf0212 Conservative FT	9	X	X	
vcf0424 DemPres Candidate FT	9	X	X	
vcf0425 Rep Pres Candidate FT	9	X	X	
vcf0426 Dem VP Candidate FT	9	X	X	
vcf0427 Rep VP Candidate FT	9	X	X	
vcf0809 Jobs/Std of Living 7 pt	7	X		X
vcf0839 Spending/Serices 7 pt	7	X		X
vcf0886 Spending: Poor People	3			X*
vcf0887 Spending: Child Care	3			X*
vcf0890 Spending: Public Schools	3			X*
vcf0894 Spending: Welfare Prgs	3			X*

*Not subject to Response Style (validity check: constrained to zero effect)

Similar to previous chapters, the model is identified by setting extreme responses to an equal and increasing propensity (1.0) and also identifies equality in each question by distance from the midpoint for other categories, the nearest two responses to a midpoint set equal to each other, as are near extremes, and so on. The calibration samples consist of ten thermometer items to gauge partisan affect, six items to gage social welfare ideology (two scales and four non-scaled funding questions), while ERS is identified on all scaled Likert items.

Each focal item, then, is allowed to associate freely with each dimension in each year.[18] The strategy of allowing a focal item to be freely identified while all other items are held constant constitutes a special case of examining *Differential Item Functioning* (DIF). The

[18] The focal item may or may not be a particularly strong indicator on substantive dimensions – and if already associated with a dimension, it is only allowed to load on the dimension of interest. Abortion (VCF0838) for instance, is not a very good bellwether of social welfare ideology and be stronger on partisanship. Spending and Services (VCF0839) should associate strongly with both Social Welfare ideology and partisan affect.

strategy is standard in identifying group differences in item parameters for which there may be a substantive theoretical difference for item behavior, whether it is between time periods (as in this analysis) or by other substantive factors (gender, race, or other discrete factor).

Measurement Results

The purpose of this multidimensional measurement effort is to distill the ERS properties of the focal issue items. I look first to policy issues, then examine the "stock" categories of Political Evaluation (Liberals, Conservatives, Democrats, and Republicans), then progress to the culturally sensitive thermometer items (Gays and Lesbians and Christian Fundamentalists), and then move to the Presidential Candidates. Again, higher coefficients will speak to (a) a relative lack of knowledge or engagement with the question, (b) a lack of engagement that indicates a level of knowledge that marginally exceeds "no knowledge," and/or (c) communicates an importance or priority in individuals who are less engaged. Patterns of results will also speak to which meaning might best be attributed to response style in these models.

Validity of ERS

Before inspecting the models themselves, some due diligence is required to insure that the models are measuring what we believe they are supposed to. Measurement models are calibrated separately for each election year from 1992 to 2008, five election years' worth of ANES administrations, with three focal issue items and six feeling thermometer items. With each of these administrations, the coefficients for the focal items are reported as the primary coefficients of interest. However, with the complexity and number of parameters and models, an overall characterization is often warranted. The main results that run through all the models, as with models in Chapter 2: (1) reflect the primary dimensions of partisan feeling and social

welfare run in the expected directions for all items, that is, selecting categories that are "more liberal or conservative" as the categories become warmer or cooler[19] and likewise the ERS coefficient for endpoint selection is the maximum category and decreases toward the midpoint in every case, (2) covariate association between years in models are mostly stable and run largely in the expected directions, and (3) selected branching items in each calibration year do not exhibit ERS. Given these validity checks, it is highly likely that what is being measured is prototypical ERS in all years, largely unrelated with substance, indicative of lower political information and engagement, and not present under branching conditions where ERS should be absent.

Table 3.2 shows the introduction of a linear regression component to the measurement model where ERS (θ_{1j}) is a linear function of political information, education, and income indicators (X_k), primary covariates of response style and indicators of mental simplification, a lack of engagement, or deliberation.

(3.2) $$\theta_{1j} = \beta_k X_k$$

The regression component is estimated with the measurement model simultaneously. As expected, these covariates strongly relate to ERS. Compared to the omitted group of political information "High or Fairly High Political Information," Low Political Information respondents display a coefficient of roughly 0.50, or ½ of a standard deviation in ERS while "Fairly Low Information" respondents have slightly lower coefficients. This is true, even independent of education and income variables, which show that lower education and lower SES respondents have significantly less extreme response. This result is noteworthy, again, given research in

[19] Face validity is also satisfied by a tendency on some extreme PE categories, for selection probabilities to be less strong than their near extreme counterparts. The implication is that "substantive" extremity behaves like we might expect in an ideal point analogue, rising in probability of selection until a respondent is essentially too liberal or conservative, and then decreasing.

political science that more extreme respondents on substance are slightly more educated and knowledgeable than average, thus our measure of extremity is more plausibly aligned with extremity that is substance-free. This table reveals a surprisingly divergent, yet plausible pattern of results: that instead of being more educated, higher SES, and higher on a scale of Political Information, those who exhibit extreme response *style* are actually the opposite, less affluent and less educated.

Table 3.2. Multiple Regression Results - ERS by Year

		1992	1996	2000	2004	2008
Political Information	Coeff	*0.5321*	*0.5164*	*0.3746*	*0.5748*	-0.0350
(Low)	(SE)	*(0.1241)*	*(0.1473)*	*(0.1405)*	*(0.1930)*	(0.1112)
Political Information	Coeff	*0.4555*	*0.3145*	*0.2262*	*0.2999*	*-0.2176*
(Fairly Low)	(SE)	*(0.0856)*	*(0.1115)*	*(0.1002)*	*(0.1308)*	*(0.0767)*
Political Information	Coeff	0.0848	-0.1213	-0.0840	0.0297	-0.0918
(Average)	(SE)	(0.0667)	(0.0846)	(0.0825)	(0.1026)	(0.0621)
Some College	Coeff	*-0.3724*	*-0.5671*	*-0.3483*	*-0.2319*	*-0.1642*
	(SE)	*(0.0710)*	*(0.0857)*	*(0.1155)*	*(0.1029)*	*(0.0831)*
College Degree	Coeff	*-0.5814*	*-0.9113*	*-0.5536*	*-0.6501*	*-0.2286*
	(SE)	*(0.0791)*	*(0.0934)*	*(0.0807)*	*(0.1163)*	*(0.0721)*
Middle Income	Coeff	*-0.4747*	*-0.3271*	*-0.2419*	-0.0446	*-0.3446*
	(SE)	*(0.069)*	*(0.0822)*	*(0.0809)*	(0.1031)	*(0.0581)*
High Income	Coeff	*-0.5241*	*-0.4773*	*-0.4563*	*-0.1944*	*-0.3718*
	(SE)	*(0.0694)*	*(0.0929)*	*(0.0827)*	*(0.1066)*	*(0.0732)*

Coefficients in italics are nominally significant at the 5% level.

As another check on validity, we might expect that measures that are not expected to exhibit ERS should not. Table 3.3 shows the results of the branching items of Gay Discrimination and Affirmative Action. As expected, endpoint selection in these items is largely unrelated to ERS, consistent with minimal bias in these measures. Even when this relationship appears, and is significant, only one appears (Affirmative Action in 2004), and has a very low impact (0.125 SD).

Table 3.3. ERS Endpoint Coefficient for Branching Items by Year

		1992	1996	2000	2004	2008
Gay Discrimination	Coeff	0.0032	-0.0405	0.0004	*-0.1293*	0.0806
	(SE)	(0.0474)	(0.0500)	(0.0427)	*(0.0713)*	(0.0545)
Affirmative Action	Coeff	*-0.0985*	0.0235	0.0004	*0.1253*	0.0806
	(SE)	*(0.0446)*	(0.0494)	(0.0427)	*(0.0660)*	(0.0545)

Overall, the models appear successful in accomplishing the intended goal, representing a substance-free measure of extreme response. The parameters of the model run in the expected direction, some important traditional associations of response style are confirmed, while theoretically unresponsive questions demonstrate little association. Covariates examined are standard in ERS literature, including race, income, and education, augmented with the interviewer-observed political awareness.

Endpoint Selection

Before examining the full results, it becomes useful to examine how endpoint selection behaves between focal items to be able to contextualize our subsequent results. A first look at extreme response entails looking at a general trend of extremity in the target items over the 1992-2004 election year administrations of the ANES. This centers how we might expect endpoint selection to vary over this time period and obtain a general sense of how these items behave without regard to whether opinions are exaggerated by response style, or are real opinions. Notably, this is from both a substantive and non-substantive standpoint. Extremes in this regard should ebb and flow with how respondents in each sample feel about these political objects and policy issues.

Table 4.4 Frequency of Extreme Response in Focal Items

Endpoint Selection Percent Variable/Year	1992	1996	2000	2004	2008
vcf0218 Democratic Party FT	9.7	13.1	13.1	12.0	19.0
vcf0224 Republican Party FT	7.6	7.7	9.7	14.4	15.0
vcf0211 Liberal FT	5.0	5.0	6.8	5.3	7.5
vcf0212 Conservative FT	3.7	3.6	5.6	5.8	7.0
vcf0424 Dem Pres Candidate FT	9.7	19.5	11.6	12.3	23.6
vcf0425 Rep Pres Candidate FT	7.1	10.9	4.7	9.5	7.8
vcf0232 Gay/Lesbian FT	21.9	19.9	16.1	16.3	18.6
vcf0234 Christian Fund FT	10.2	9.1	7.7	10.5	11.1
vcf0839 Spending/Services 7 pt	12.4	10.0	15.3	15.9	23.8
vcf0843 Defense 7 pt	11.0	8.2	6.6	12.8	20.8
vcf0838 Abortion 4 pt	56.4	53.5	54.2	56.2	55.8
vcf0853 Traditional Family 5 pt	50.5	46.6	50.0	42.3	42.1
vcf0854 Moral Standards 5 pt	28.0	23.4	27.6	24.0	30.3

To characterize this table briefly, endpoint selection has risen over this time for most items. For Democrats, Republicans, Liberals, and Conservatives, the frequency has roughly doubled in size. Defense and Spending and Services as issues saw less frequent endpoint selection in 1996 and 2000 followed by a jump that exceeded 1992 levels. By contrast, most cultural values thermometer and issue endpoint selections remain flat to decreasing. Finally, the presidential candidate responses fluctuate idiosyncratically with spikes at Clinton 1996, Dole 1996, and Obama 2008. Between questions, the level of endpoint selection is notably higher for questions that are not feeling thermometers, Abortion, Traditional Family, and Moral Standards and notably lower for Liberals and Conservatives.

Exemplar: Democratic Party 1992

Reporting is complicated by the large number of parameters estimated, so the plots here report only targeted parameters, notably the coefficients and charting of extreme responses due to Extreme Response Style. The example in Table 3.5 and Figure 3.1 serves to demonstrate how these behave at the most basic level in their entirety, and how differences are examined in

this chapter. The Democratic Party category selection behavior with respect to ERS is reported in coefficients in Table 6.2. In this table, each category c1 to c10 has a "constant," and a "slope" parameter for each relevant dimension. For the Democratic Party feeling thermometer item, only Extreme Response (ERS) and Political Evaluation (PE) are allowed to load – as seen in Table 3.5. It is evident that in both sets of category parameters that the sum is set to zero for purposes of identification. Categories with a negative constant are less frequently endorsed, and those with a positive constant are more frequently chosen.

Table 3.5. Democratic Party Coefficients in 1992 Base Model

Democratic Party Answer Category		Constant	se	ERS	Se	PE	se
Non-response	c1	-0.2465	0.1129	0	.	0	.
Lower Extreme (0-5)	c2	-1.6100	0.2160	1	.	1.3117	0.1782
	c3	-1.1649	0.1942	0.1048	0.0354	1.5918	0.1484
	c4	-0.2593	0.1293	-0.4183	0.0268	1.1597	0.1160
Near Extreme (Cold)	c5	0.3480	0.1028	-0.6713	0.0264	0.8554	0.0940
Midpoint (46-54)	c6	1.3241	0.0670	-0.0304	0.0622	-0.1139	0.0816
Near Extreme (Warm)	c7	1.0292	0.0871	-0.6713	0.0264	-0.4975	0.0829
	c8	0.9629	0.0845	-0.4183	0.0268	-0.9732	0.0906
	c9	0.5492	0.0945	0.1048	0.0354	-1.3242	0.1075
Upper Extreme (95-100)	c10	-0.9328	0.1666	1	.	-2.0098	0.1565

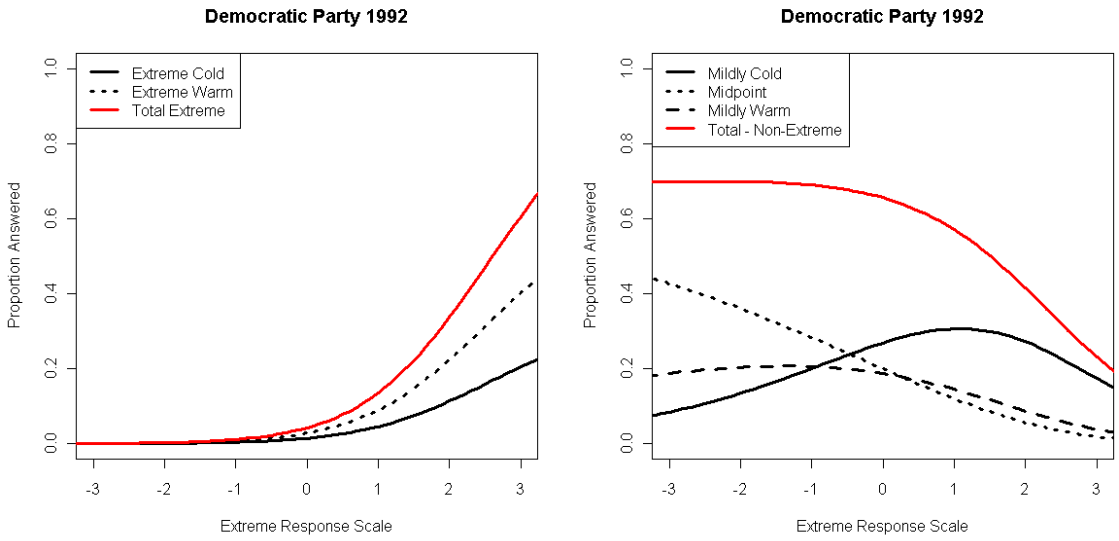
For Political Evaluation in Table 3.5, the tendencies to select categories are easily interpretable trough the coefficients. From lower to higher categories (conservative to liberal in this case), the likelihood of selecting a more liberal category rises from category 2 ("cold endpoint") to category 10 ("warm endpoint") and the coefficient regulates largely what we expect to see – liberal category selections result from warmer liberal feelings.

For ERS in Table 3.5, category selection tendencies are also fairly intuitive. By constraint, the endpoints are set to equality, though they are not necessarily bound to be the maxima. In this case (and in every case for every year), they are the maximum category – strictly increasing in ERS. Additionally, every near extreme is lower, and each coefficient

becomes less indicative of ERS to the midpoint. Midpoint selection, however, is somewhat of an anomaly in how it behaves. In this case, midpoint selection carries a slight negative coefficient, seen in other studies of ERS (Moors 2003, Bolt 2009), possibly an artifact of both a tendency to respond in non-extreme terms and a tendency to simplify a question by choosing an easily accessible middle category.

These simpler interpretations notwithstanding, because the coefficients are couched in a nominal response model, it is not readily obvious what the exact shape of the category curves should be. The relative slopes and intercepts interact, similarly to standard regression models, to produce the resulting category characteristic curves (CCCs) that are most readily interpretable (Figure 3.1).

Figure 3.1 Plots of Democratic Party from Table 6.2, 1992[20]



Thus, we might find more utility of inspecting the graphs themselves, which I repeat in subsequent sections. What is more obvious from the figure (Figure 3.1), on the normalized ERS

[20] All charts are presented "conditioned on valid response" such that "Don't Know," "Refusal," and "Not Asked" are removed from response probability, even though they are estimated. This insures comparability of estimates that may fluctuate in their missingness by differing survey designs (such as 2008 in Table 3.4).

scale, is that endpoint selection increases substantially as the scale extends from average values of ERS (around 0) to the more heavy extremes. On the left, as respondents move from 0 (the midpoint) to +2 (standard deviations in the direction of extremity), endpoint selection rises from 5 to roughly 40 percent. This is a fairly strong contribution from "non-substantive" ERS. It is also notable that extremely cold selections (the solid black line in the left hand figure) are far less frequent, but have the same logistic slope parameter, a property not readily identifiable through a coefficient table, but evident in the more negative constant in the cold extreme (c2=-1.61) compared to the top extreme (c10=-0.55) evaluated at the midpoint (0) of ERS.

On the right hand graph, by contrast, are the midpoints and "near midpoint" categories on the scale. All of the near midpoints have negative coefficients and are generally decreasing (the heavy dashed line and solid line in the right hand figure), yet the slopes in the relevant range of display are affected by category tradeoffs such that the milder coefficient of the lightly dashed midpoint line decreases more sharply in visual terms than its coefficient might suggest (-0.030 from Table 3.2; by contrast, -0.671 for the "near midpoints"). The combination of near midpoints and midpoints tell a clear story (the top red line): for moderates, midpoint responses dominate the lower half of the ERS distribution and quickly evaporate as ERS rises. As Extreme Response rises, midpoint selection decreases more over the range of the graph than either "near midpoint" selection without the strong negative coefficient. This exemplar points out in modest terms what might be learned from coefficients alone, requiring a look at the curves as they interact.

At this point, then, it becomes necessary to define some terminology germane to interpreting these slope and constant (intercept) coefficients for ERS that become necessary to keep in mind when interpreting the graphs. For ERS, positive coefficients indicate a rise in

endpoint selection probability as the latent variable (ERS) rises. For future reference, I will adopt the terminology of item response and call this slope rise *information or sensitivity* of endpoint selection to ERS. This means, as Extreme Response Style rises, the probabilities of endpoint endorsement increases, and midpoint/near midpoint selection decreases – sharper increases indicate more "information" or "sensitivity" of endpoint selection to ERS. The constant parameter indicates to what extent the frequency or *level* of endpoint selection increases overall at ERS when evaluated at zero. When both Political Evaluation (PE) and Extreme Response style are held at their midpoint (0), the constant reveals the frequency of endpoint selections at this point. The meaning of -1.61 for the "extreme cold" and -0.55 for "extreme warm" come into focus in relative category selection terms on a normal scale: from the formal equation, exp(-1.61) being 80.0 percent less frequent than average and exp(-0.55) being 57.7 percent less frequent than an average selection at this point. The result, with a ten category average of 10 percent is 2.00 percent extreme cold and 4.23 percent extreme warm when evaluated at zero. It follows, then, that when we move to a quantity of 1 sd to the side of extremity that these parameters shift accordingly by 1.0 (the constrained effect), such that an extreme cold response becomes exp(-0.61) resulting in a 5.43 percent frequency and the warm response becomes exp(0.45), a 15.68 percent frequency. In this way, the equation can be extended to include the political scale as well, where a 1 standard deviation increase in liberalism increases warmth to Democrats by a transformed factor of its extreme warm coefficient (-2.01) and its extreme cold coefficient (1.31), and so on. Like logistic transformations for marginal effects, the system of equations presents regular interpretations that may sometimes require graphical transformations. By inspecting these parameters together, as we must in the logistic regression analogue, some general regularities can be

brought into sharper focus even if they might be easily recognized and classified by slope and constant parameters.

One further, and related, decision about display is made here, one that risks falling by the wayside for many observers, but is important in this analysis through this example. To wit, some may wonder "where has political evaluation gone in this graphical mix?" The answer is not that it is ignored in these charts, but that it is held constant at zero. Because zero is the "middle of the road" in the case of political evaluation, the extreme values reported are generally as low as we might expect them to be. In fact, those with little to no response style at all will be rightly expected to exhibit endpoint selection probabilities of zero. What else might we expect of a moderate who additionally shies away from endpoints? As a result, and given the relative lack of movement below the midpoint (0) the figures reported will concentrate on the top half of the distribution in these specific examples, where extreme response has the most influence. Importantly, in common interpretation, this does not mean that values of +4 and above standard deviation hold little meaning, but that they have specifically little meaning *for political moderates,* understanding that any liberal or conservative movement will make endpoint selection more frequent. The meaning of the otherwise graphical "out of sample" values (+2 on ERS scale or above) will be more central when extended to more substantively liberal or conservative respondents where curves will follow a roughly parallel horizontal path.

Secondly, and more importantly, this small example of Democrats in 1992 shows how these coefficients might be interpreted largely on their own (slope and intercept), though it necessitates a graphical display to check on the meaning of the slope coefficients with respect to any specific level of ERS and PE, and to inspect the result of what could be more complex category tradeoffs as mediated by the constant parameters. That is, the "information content"

(*information/sensitivity*) of the question of Democrats in ERS is one facet of the curve's shape, yet the "overall non-substantive response" (*level* - the constant, horizontal shift), is yet another. Each holds meaning that is contained in, yet not always readily identifiable by its coefficients.

At this point, then, it is necessary to present statistical caveats as well. Differential Item Functioning analyses like these can be prone to fluctuation based on, not only sampling, but the particular circumstances of elections. Standard errors of estimation alone cannot account for the variability in estimating ERS. In these models, inconsistent measurement of ERS or the primary dimensions of interest might induce smaller differences in particular years that are not particularly meaningful. It is still possible, even probable, that ERS mixes two or more response processes/meanings as well, a first process that resembles true lack of knowledge, and another that resembles deliberation. It is also possible that contaminating dimensions may have influence in one or more years. Overall, however, given the survey of models run, the number of target items and years, a picture of ERS starts to emerge in its political influence. The overall strategy has been to allow all possible information to be accounted for toward estimating ERS properly, and it is this power that guides what we see in the results. Because there are a number of model variables held constant (intercept most importantly), the measurement model here is assumed to be the most appropriate (even though other specifications can be used), and the model constraints involved are just one set in a number of what may be plausible alternatives, we cannot be certain that the following provides a definitive view of cognitive reality.

What comes from these snapshots of extreme response, then, are more aptly described as a set of impressions that may be highly suggestive of what is happening in response patterns and provide hypotheses that are more likely to be correct than others. What do not come are

specific refutations of any hypotheses beyond the very basic: the situational variance in response style appears to have meaning toward defining how we respond to political objects. Looking at these results, there are some conclusions that appear, and others that might have been expected that do not appear.

ERS Issue Item Differences: Over Time and Between Items

Given the measure of selection appears valid, how do the parameters between issues and within issues over years differ? Does issue endpoint selection (and its presumptive awareness and engagement) change with respect to ERS over the period from 1992 to 2008? The quick answer is mixed. *Within single issues between time periods*, there is possible evidence of change, yet the more likely story is that this is an incomplete picture of change over time and not indicative of anything that is readily interpretable. Change in issues might be idiosyncratic to measurement process or to the context of the time period and do not show a consistent pattern of deviation. A reasonable expectation might have been that issues that appear central to campaigns or party divisions at any given time (Defense after 2001, Spending and Services after the recession in 1992, then after the recession in 2008).

This said, it is very notable what kinds of changes appear *between issues*. Abortion, Traditional Families, and Moral Standards all lag in their sensitivity of endpoint selection as a result of ERS. Defense and Spending/Services however, show a stronger relationship with ERS. Results are presented in the following sections. Because the question text is an important facet of how items are evaluated, I report the text for each question and look at each in turn.

Spending and Services

> *VCF0839. Some people think the government should provide fewer services, even in areas such as health and education, in order to reduce spending. (2004:*

Suppose these people are at one end of a scale, at point 1.) Other people feel that it is important for the government to provide many more services even if it means an increase in spending. (2004: 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.)

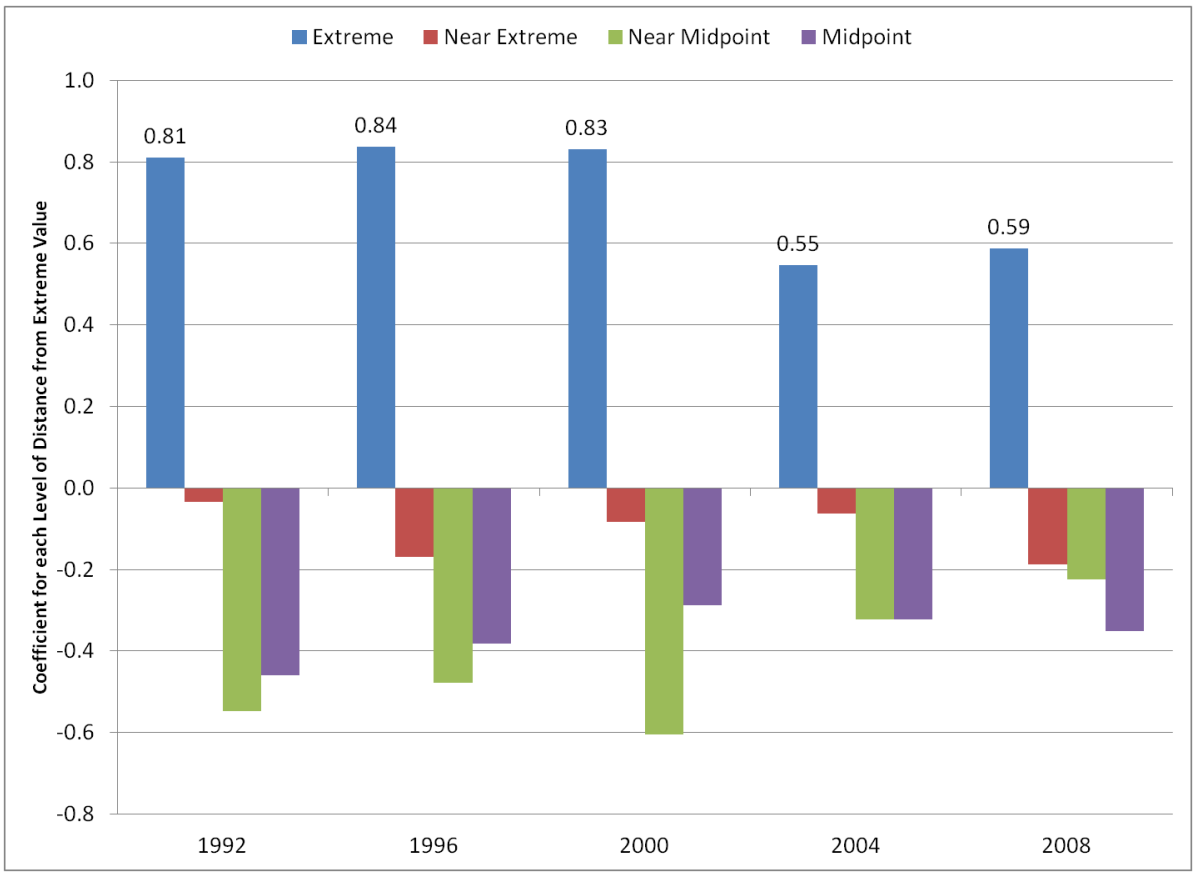
Where would you place yourself on this scale, or haven't you thought much about this? (7-POINT SCALE SHOWN TO R)

VALID_CODES:
---
1. Government should provide many fewer services: reduce spending a lot
2.
3.
4.
5.
6.
7. Government should provide many more services: increase spending a lot

For Spending and Services, the expectation might also be a general decrease in endpoint selection as concern over deficits get larger over time. As this issue becomes more salient, the electorate may become more strongly, and truly, opinionated. Alternately, as economic concerns reach a peak (1992 and 2008), we might see decreases in response style as economic concerns dominate the election cycle. However, in Figure 3.2.a, what we see is some, but not a meaningful or interpretable, change in the sensitivity of each question to ERS (blue bar). The other response types are reported as well, but it does not appear there is a big difference in how these selections deviate either (red, green, and purple bars). A very aggressive interpretation might confirm the notion that 2004 and 2008 are relatively low. This coefficient graph seems to indicate that there is possibly a detectable effect of ERS *sensitivity* or *information* over time in spending and services, however the drop appears to be stark in 2004 and 2008 for little reason. The timing is somewhat suspect in terms of making any valid point. No primary thread of

practical or academic discussion has called out government spending or services in 2004 or 2008 any more than in previous years.

Figure 3.2.a. Spending and Services "Sensitivity"/"Information", 1992-2008

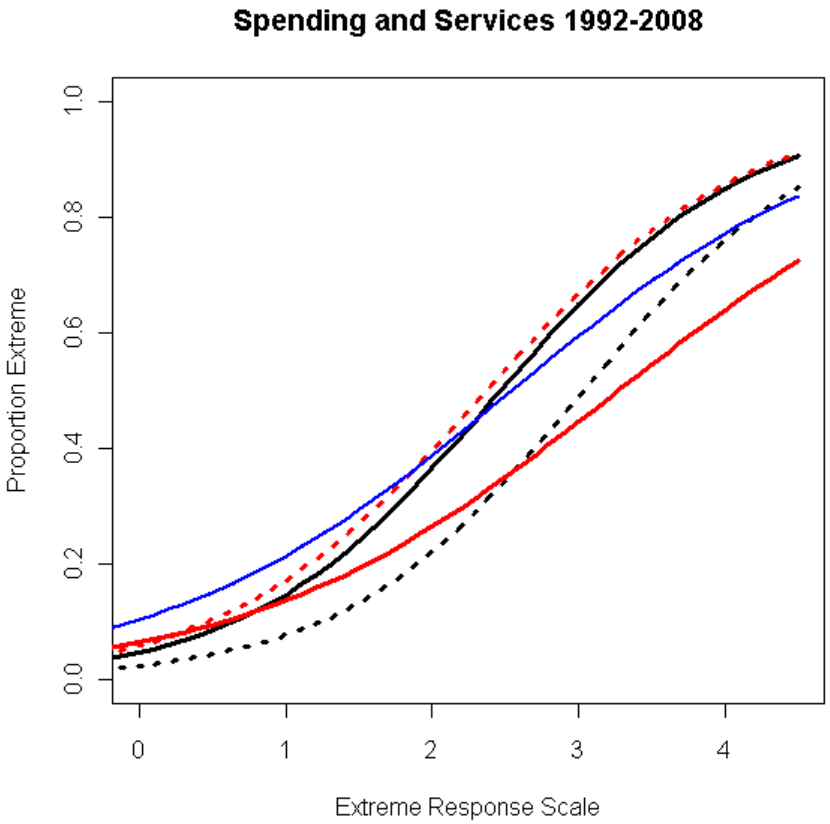


7-point scale: Extreme values at 1 and 7, Near Extreme: 2 and 6, Near Midpoint: 3 and 5

What is evident in Figure 3.2.b, however, is a general similarity in the *sensitivity* of non-substantive extreme response, but possibly in *level*. Figure 3.2.b indicates much the same phenomenon of modest change in slope, yet displays a slight difference in general trend of *level* from 1992 to 2008 in odd year – 1996 and 2004. In conjunction with Figure 3.2.a, the implication is that barring very specific contextual differences which could be slight differences in how questions relate from year to year, respondents are not particularly more prone overall to automatic responses in 2008 than they were in 1992. For 1992, 2000, and 2008, those who are

extreme to, say, 10 percent highest in ERS, are predicted to select extreme responses 30-40 percent of the time (black, dotted red, and blue respectively), while 1996 and 2004 samples show a predicted extreme response selection of 10-20 percent. If we were to derive expectations in this regard, we might have expected the exact opposite, more careful consideration in 1992 and 2008, and less so (higher endpoint selection due to ERS) in other years.

Figure 3.2.b. Endpoint Selection due to Extreme Response Style: Spending and Services, 1992-2008



Again, the graphical analysis, the combination of coefficients for Spending and Services contains only modest change with respect to ERS. The difference is found in the overall *levels* of endpoint selection, but not in the sensitivity overall. Around the midpoint of extreme

responders, the difference is very little in terms of increased or decreased probability of selecting an endpoint (0-10 percent). Yet, for moderate to high extreme responders, the difference can be more prominent in selecting an endpoint for reasons that are limited to a common response pattern of extremity. This modest tapering, however, may not be indicative of much except variance in how the measurement model behaves. Nominally "significant" in 2004 and 2008, these patterns do not seem to signify much change. Given the standard, modeled uncertainty and given the uncertainty of differential item functioning between years (unmodeled uncertainty), this interpretation of change might be overzealous.

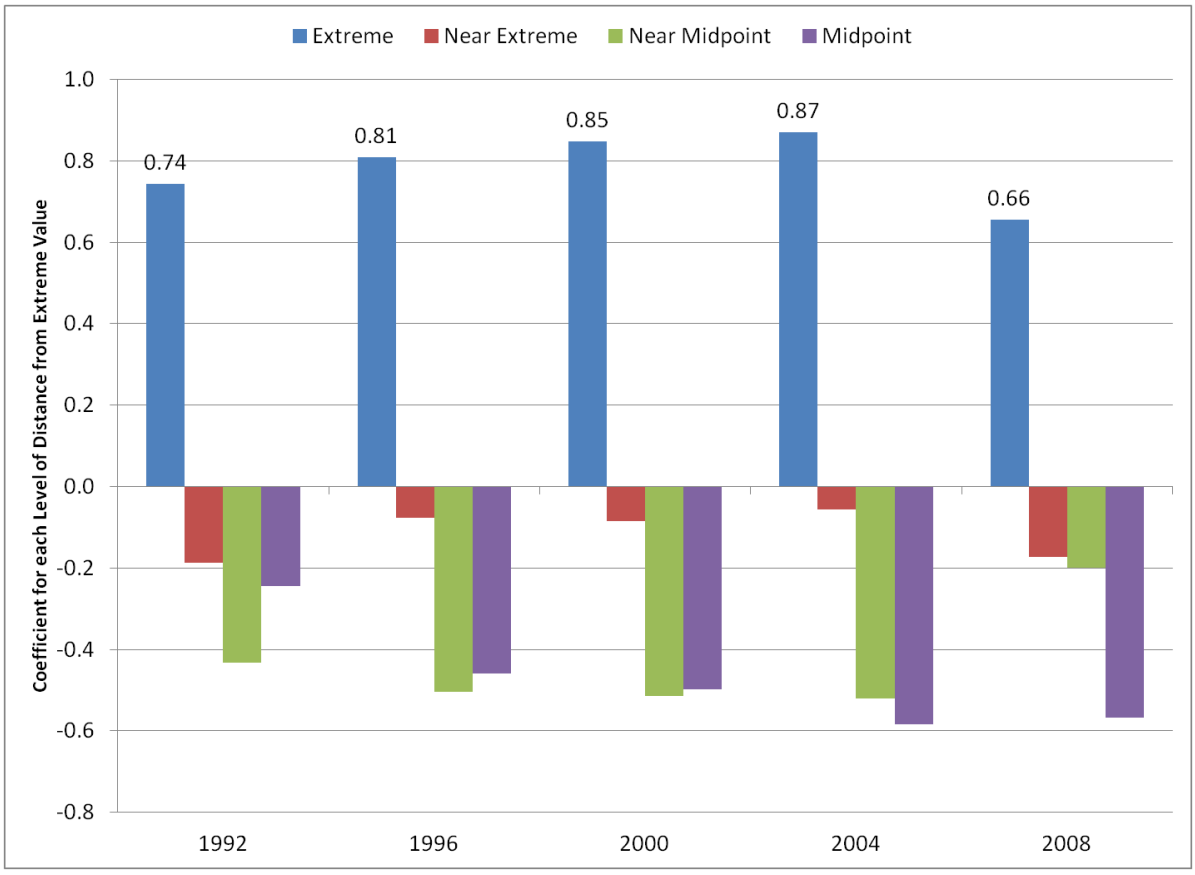
Defense

> *VCF0843. Some people believe that we should spend much less money for defense. (1996,2004: Suppose these people are at one end of a scale, at point 1.) Others feel that defense spending should be greatly increased. (1996,2004: Suppose these people are at the other end, at point 7.) (2004: 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? (7-POINT SCALE SHOWN TO R)*
>
> *1. Greatly decrease defense spending*
> *2.*
> *3.*
> *4.*
> *5.*
> *6.*
> *7. Greatly increase defense spending.*

On the standard issue of defense, I again find no precise pattern of change in ERS with respect to *sensitivity* that is apparent for Spending and Services. The coefficients we may be most interested in (blue bar) rise somewhat to 2004 and then take a dip downward. Categories at the extreme have the same amount of non-substantive information across this time period with respect to ERS except in 2008, which may be idiosyncratic due to the nature of other

endpoint selection tendencies as well (Figure 3.3.a), where the model appears to pick up either other characteristics or a different type of patterned response (red, green and purple bars). This is contrary to the expectation that non-substantive response might decrease during periods of international crisis, threat, or war. As the focus of the aftermath of the cold war shifts toward the "war on terror," very little changes in the realm of response sensitivity here, and it does not do so in years where we might expect defense to be an issue, more or less salient.

Figure 3.3.a. Defense "Sensitivity"/"Information", 1992-2008

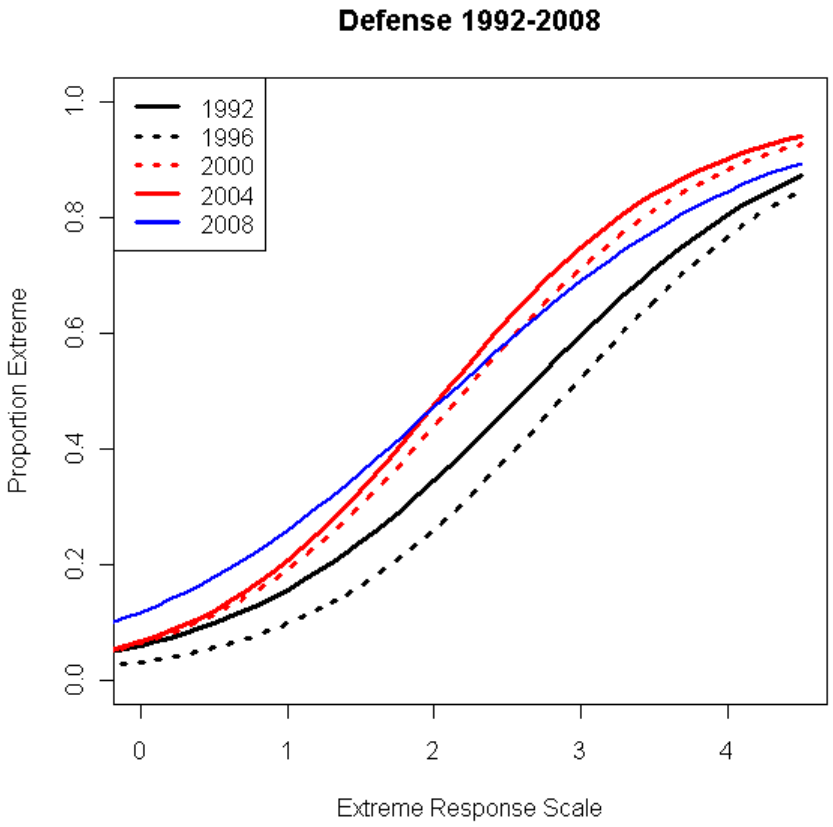


7-point scale: Extreme values at 1 and 7, Near Extreme: 2 and 6, Near Midpoint: 3 and 5

Graphically in Figure 3.3.b, however, shows a slightly different story to Spending and Services, such that there is an increase in non-substantive extremity over the same time period horizontally, a uniform shift in level of endpoint selection as ERS decreases. Between 1992

and 1996 (black and black dotted lines), extreme responses are elicited much less frequently than afterward for extreme responders (roughly 1.5-2.5 on the extreme response scale). Yet this runs into problems of interpretation again. Any attempt to establish a pattern here lacks any type of theory about the issue of defense that puts the year 2000 as the pivot point.

Figure 3.3.b. Endpoint Selection due to Extreme Response Style: Defense, 1992-2008



If we are forced to conclude much, Defense follows the pattern of a relatively shift toward non-substantive evaluation. For Defense, non-substantive extremity after 1996 is more highly endorsed *across the full range* of ERS, even though the slopes are largely similar. Yet temporal change appears to be less of a clear story has any kind of rationale.

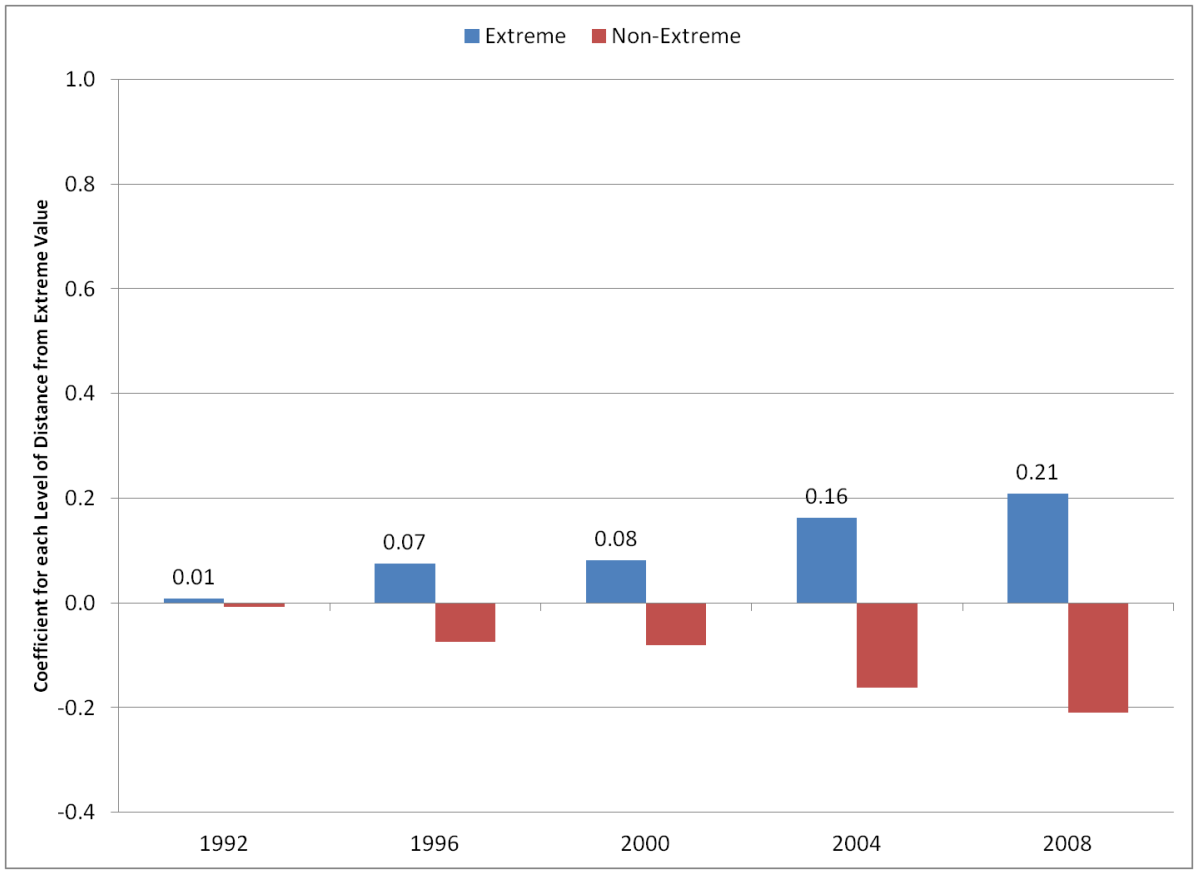
Abortion

VCF0838. There has been some discussion about abortion during recent years. (RESPONDENT BOOKLET) Which one of the opinions on this page best agrees with your view? You can just tell me the number of the opinion you choose.

> *VALID_CODES:*
> ---
> *1. By law, abortion should never be permitted.*
> *2. The law should permit abortion only in case of rape, incest, or when the woman's life is in danger.*
> *3. The law should permit abortion for reasons other than rape, incest, or danger to the woman's life, but only after the need for the abortion has been clearly established.*
> *4. By law, a woman should always be able to obtain an abortion as a matter of personal choice.*

For Abortion, the story is much different and is meaningful. For the abortion question, there is not only no meaningful change in ERS, but there is no detectable ERS to speak of. Figure 3.4.a shows this in the starkest terms, especially given the difference between Spending/Services and Defense, that show much steeper rises. By contrast, Abortion in any year is not significantly different from a zero slope in every year except 2008. It might be notable that this appears to be a trend from 1992 to 2008, yet because of the similarly sized, or larger) dips and peaks in Spending/Services and Defense that do not comport with any theory, caution is likely warranted.

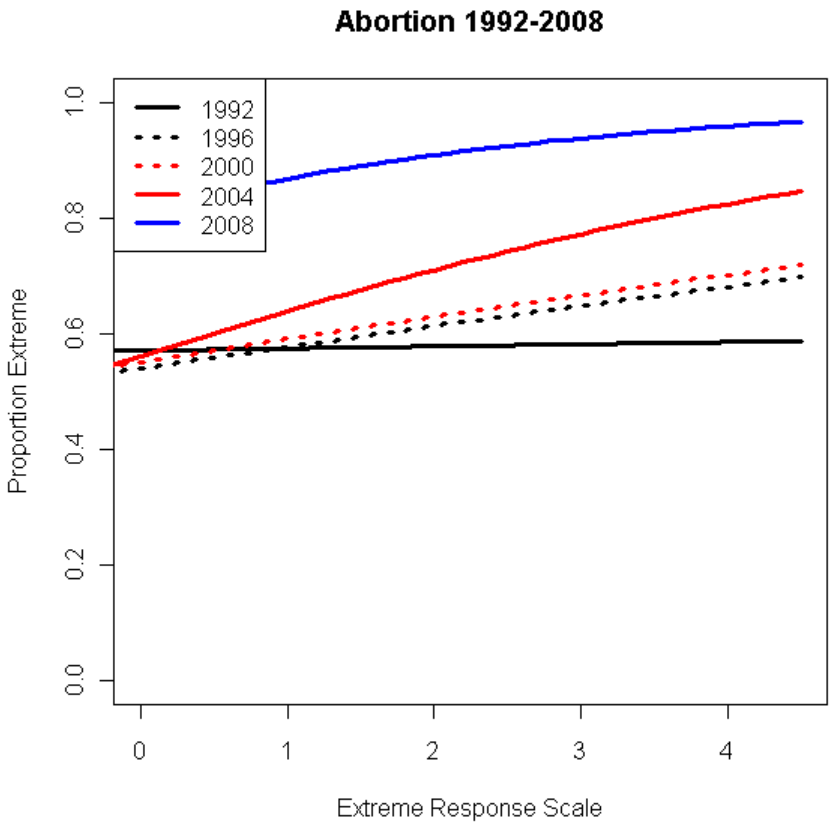
Figure 3.4.a. Abortion "Sensitivity"/"Information", 1992-2008



4-point scale: Extreme: 1 and 4, non-Extreme 2 and 3

The conclusion is dampened somewhat in Figure 3.4.b, but remains statistically true, that given uncertainty, the lines are no different than flat, no extreme response difference across the range of the Abortion question. The main difference here is in 2008, in which ERS is flat across the relevant, displayed range even though the slope is significantly positive and meaningful only at extremely mild levels of ERS. The Abortion question, overall, has an inconsequential level of ERS, and 2008 appears to be meaningfully deviant. When likelihood of an extreme response is so high in this range, a very narrow statement can be made for 2008 (the blue line), such that those who have a propensity to ERS are extremely likely to pick an endpoint, but this likelihood is not particularly steep in the range of plausible ERS.

Figure 3.4.b. Endpoint Selection due to Extreme Response Style: Abortion, 1992-2008



The relatively non-existent response sensitivity for Abortion presents one comfortable interpretation and one unanswered question. It is reassuring, on one hand, that an item that is "easy," amply labeled, and simply structured (4-point scale) would result in little, if any, response style impact. The null result, if attributable to a clearer scale, at least validates that the series of models appears to be working to isolate what I had hoped overall. On the other hand, it is unclear whether the null result is largely an effect of the question structure itself or a byproduct of its content. Previous research indicates that "structure," or number of answer choices, may have little impact in and of itself (Kieruj and Moors 2010), so provisionally, this indicates that Abortion is, among the issues thus far, the least ambiguous, most relevant, and/or

the most deliberated issue among those surveyed so far. But is Abortion simply clearly articulated in the public mind or is the question itself unambiguous? To look at this, I take a look at two other similarly cultural questions to track their sensitivity to ERS: Traditional Family and Moral Standards.

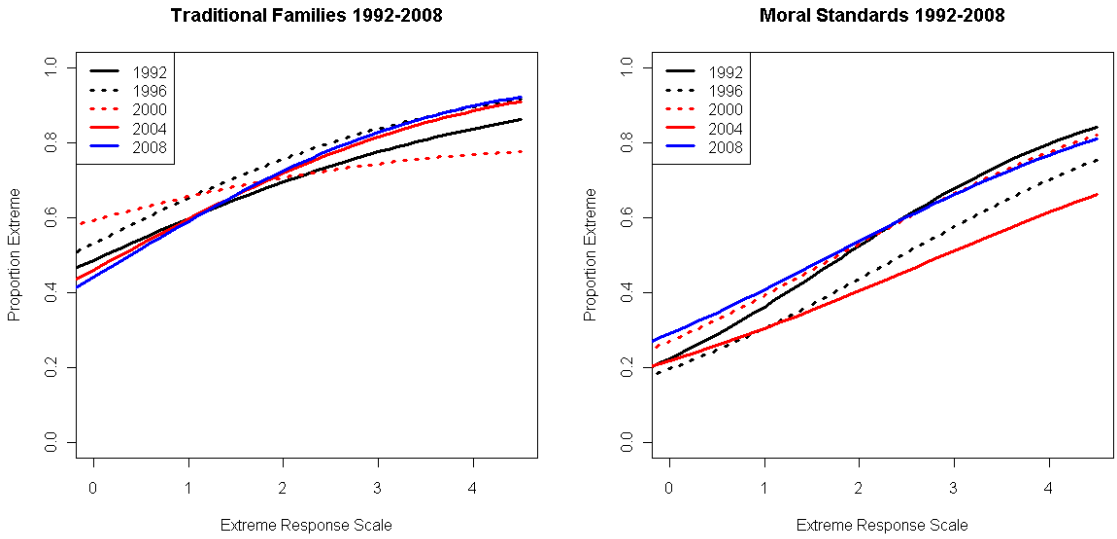
Traditional Family and Moral Standards

> *VCF0853. 'This country would have many fewer problems if there were more emphasis on traditional family ties.' (2000,2004: do you agree strongly, agree somewhat, neither agree nor disagree, disagree somewhat, or disagree strongly with this statement?)*
>
> *VALID_CODES:*
> ---
> *1. Agree strongly*
> *2. Agree somewhat*
> *3. Neither agree nor disagree*
> *4. Disagree somewhat*
> *5. Disagree strongly*
> *8. DK*
>
> *VCF0854. 'We should be more tolerant of people who choose to live according to their own moral standards, even if they are very different from our own.' (2000: do you agree strongly, agree somewhat, neither agree nor disagree, disagree somewhat, or disagree strongly with this statement?)*
>
> *VALID_CODES:*
> ---
> *1. Agree strongly*
> *2. Agree somewhat*
> *3. Neither agree nor disagree*
> *4. Disagree somewhat*
> *5. Disagree strongly*
> *8. DK*

Traditional Family Values and Moral Standards augment the inclusion of Abortion as bellwether cultural items. If Abortion, by virtue of its content (highly salient), or by virtue of

its structure (text heavy with no corresponding middle option) is not responding, do other questions in the mold of traditional versus modern values respond to ERS?

Figure 3.4. Endpoint Selection due to Extreme Response Style: Traditional Family and Moral Standards, 1992-2008

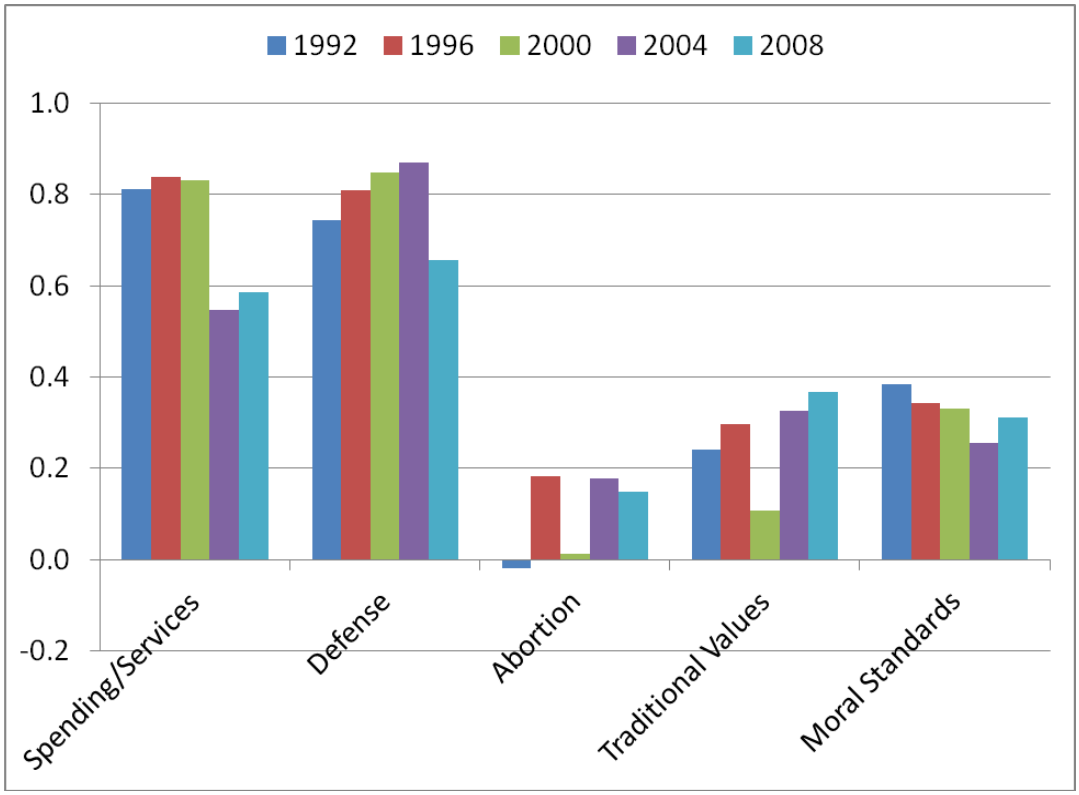


The conclusion from these items is that they do respond, and are statistically significant every year, but respond very weakly, and by the same token, show no apparent change between election years in the ANES that might warrant interpretation. While it might be tempting to read special meaning into the lower levels of Moral Standards endpoint selection due to ERS in 1996 and 2004, there is no particular reason to do so. These curves, given the standard error of measure in slope and intercept and modeling uncertainty, paint a picture of similarity overall. Yet, the striking difference comes in comparison to Figures 3.1.b (Spending/Services) and 3.2.b (Defense). By comparison, these curves show a very weak relationship of endpoint selection to ERS.

The results reveal that ERS in issues is largely consistent from year to year in these items, given reported standard errors and measurement uncertainty. For Abortion, there is effectively no response style in the 4-point question across all years. For Spending and

Services and defense (both 7-point scales), there is a modest range of endpoint selection due to response style, but little interpretable variance over time, while Defense appears to increase in endpoint selection due to response style. The results might suggest to an analyst perhaps that there is a possibility of *invariance* in response style, and that content between years is, indeed, irrelevant.

Figure 3.5 Endpoint Selection Coefficients for Policy Issues, 1992-2008



Several observations make the invariance argument less forceful as evidenced in Figure 3.5. Combined with the results in chapter 2, this conclusion of invariance is likely unwarranted. Additionally, in this analysis we see, if not meaningful variance over time, that variance between issue areas does exist. In addition to Abortion, Traditional Family Values and Moral Standards are consistently less meaningful indicators of endpoint selection on ERS than Spending/Services and Defense. The lack of strength belies a quality in these cultural questions

that does not draw out endpoint selection due to ERS as strongly as those in Defense or Spending/Services. We might, given the consistency and magnitude of these relationships here, draw the conclusion that cultural values are important, "easy" (Carmines and Stimson 1986), or deliberated enough for respondents to have informed and nuanced opinion. This runs slightly contrary to what colloquial expectations might be, namely that cultural values is often characterized in partisan debate as a litmus test set of issues which we might ascribe to the public a certain automatic and perhaps ill-considered reaction. On the contrary, this set of issues in this analysis presents the opposite. Compared to the flagship items of economic values and defense, this set of issues appears relatively well-considered.

Abortion represents a different conundrum for which there is no fully adequate answer here. The question structure itself could be clear enough to preclude ERS[21], or, perhaps, the issue is not ambiguous. If we rely on previous research, namely Kieruj and Moors (2010), response style should be present in any extended format, including 4-point scales. Krosnick (1991) contends, on the other hand, that labeling may effectively reduce response style. Considering the weight of evidence in this dissertation as a whole, more likely and pertinent is the prospect that abortion is a readily interpretable question – much like Gun Control (chapter 1). The apparent non-result is telling, namely that the hot button issue of the last several decades, and by extension a set of cultural issues that evoke passion in the electorate, is not a kneejerk phenomenon, but a well-considered question of substance for respondents.

ERS Thermometer Variance: Feelings about Selected Political Objects

Nevertheless, cultural dimensions of conflict provide a bridge toward the inspection of political groups/persons. Since a question's structure (4-point, 5-point, 7-point, 9-point, or

[21] Abortion (vcf0838) has no midpoint which may force evaluation in a respondent's mind.

specific wording or labeling) may have some bearing on ERS, it is natural to look for items that change, not by question structure, but entirely by content.

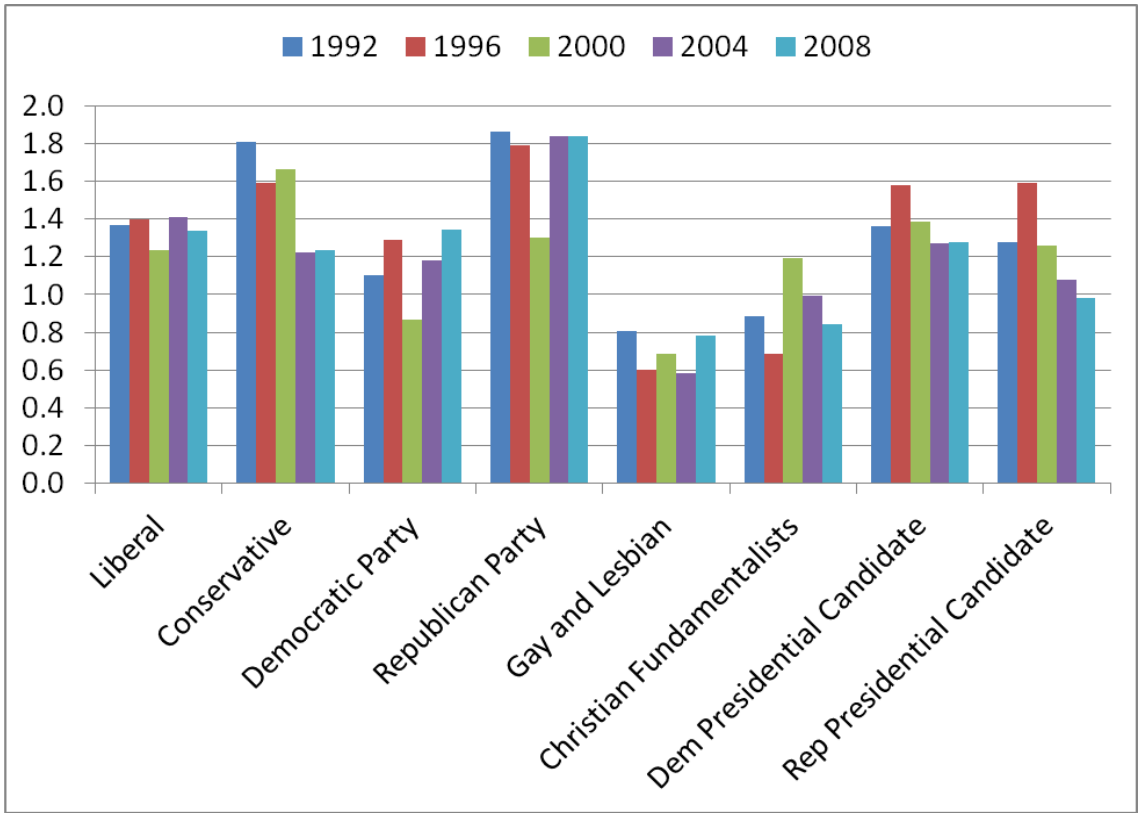
For this task, we might rightly utilize the thermometer item itself. Feeling thermometers, as discussed in chapter 2, have the benefit of being one of the most ambiguous constructs in the American National Election Survey, consisting of a point assessment between 0 and 100 (or nine verbal prompts). The other benefit of the feeling thermometer battery is that the structure of the questions is wholly identical except for the object in question. If we ask of these questions - which vary due to extreme response differentially due to content alone? – we may most meaningfully assess this by looking at thermometers in the ANES. In this way, we might place them side-by-side and inspect their differences as solely due to content on their slopes, the sensitivity with which respondents choose endpoints due to ERS.

The Feeling Thermometer series of questions is prefaced with a card of response options (Figure 3.5) and asked:

"...using the thermometer how would you rate the following:"

Figure 3.5. Thermometer Rating Options

Figure 3.6. Extreme Response Coefficients, Selected Thermometer Items, 1992-2008



All Thermometer Items 0-100: Anchors from Figure 3.5 and Extremes at 0 and 100

The summary is presented in Figure 3.6. Again, the picture of endpoint selection due to ERS *within items over time* is mixed. Some trends appear relevant with respect to Conservatives and the Republican Presidential Candidate. Secondarily, Republican Presidential Candidates appear to undergo a modest decline. In terms of movement and unique behavior, the conservative Republican political objects appear to be highlighted while the liberal Democrats remain largely constant. The Republican Party maintains the steepest endpoint selection propensity, and with them, Conservatives follow in 1992-2000. However, these observations warrant the further inspection by the endpoint category curves which follow, the symbolic evaluations of Liberals, Conservatives, Democrats and Republicans, followed by the traditional social identification, or less political cultural objects, and ending with the

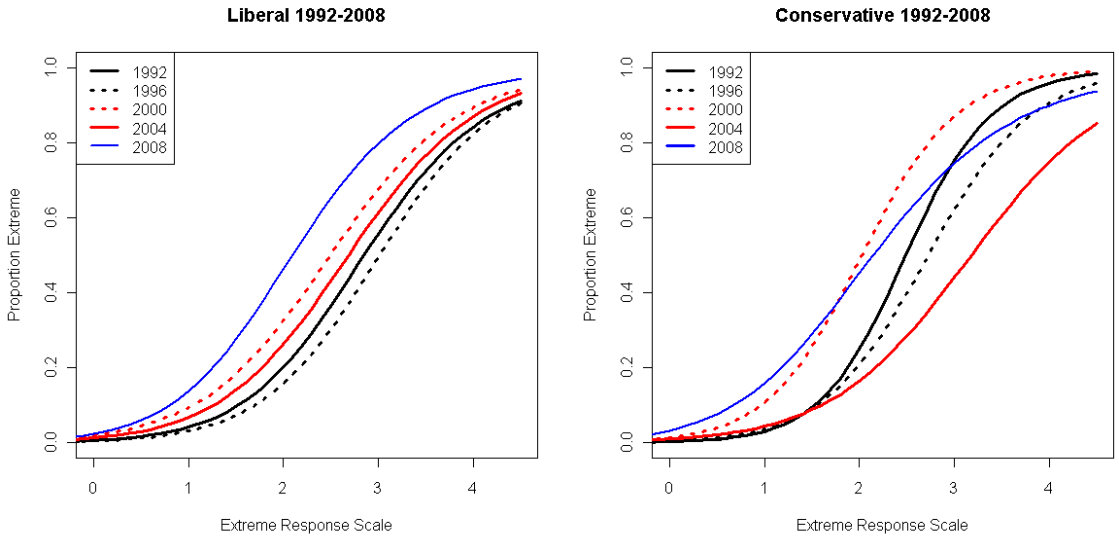
presidential candidates. These more minor interpretations are merely exploratory, and what we see over time is not a truly strong move in any of these measures over time. However, again we might move to the hotter cultural items of Gays and Lesbians and Christian Fundamentalists to notice the largest difference. These questions keep a consistently lower slope overall. These more minor interpretations are merely exploratory, and what we see over time is not a truly strong move in any of these measures over time. However, as with the issues analysis that preceded, the main source of difference is *between the political objects themselves*. One difference that might be noticed is between the Democrat and Republican feeling thermometers which show very high levels of ERS used toward Republicans as a group. And again we might move to the hotter cultural items of Gays and Lesbians and Christian Fundamentalists to notice the largest difference. These questions keep a consistently lower slope overall. As with issues, the main source of difference is between the political objects themselves that may focus on more culturally salient issues and groups.

Liberals, Conservatives, Democrats, and Republicans

The first set of results in Figure 3.7 provides some nuance to the slope coefficients. Two notes, in particular, do stand out. First, considering the absolute level of endpoint selection due to ERS, even though these key target questions provide mixed evidence of change in response style over time in slope, actual percentage of endpoint selection presents strong temporal contrast. Over the 1992 to 2008 elections, Liberals actually increase uniformly in endpoint selection to levels that and Conservatives have inconsistent curves, yet peak in 2000 and 2008. These may be overinterpreted, however at face value it appears there is some evidence that Conservatives and Liberals are being evaluated less substantively in most recent years, and not only less substantively, but very much more so for respondents in more relevant ranges (20-40

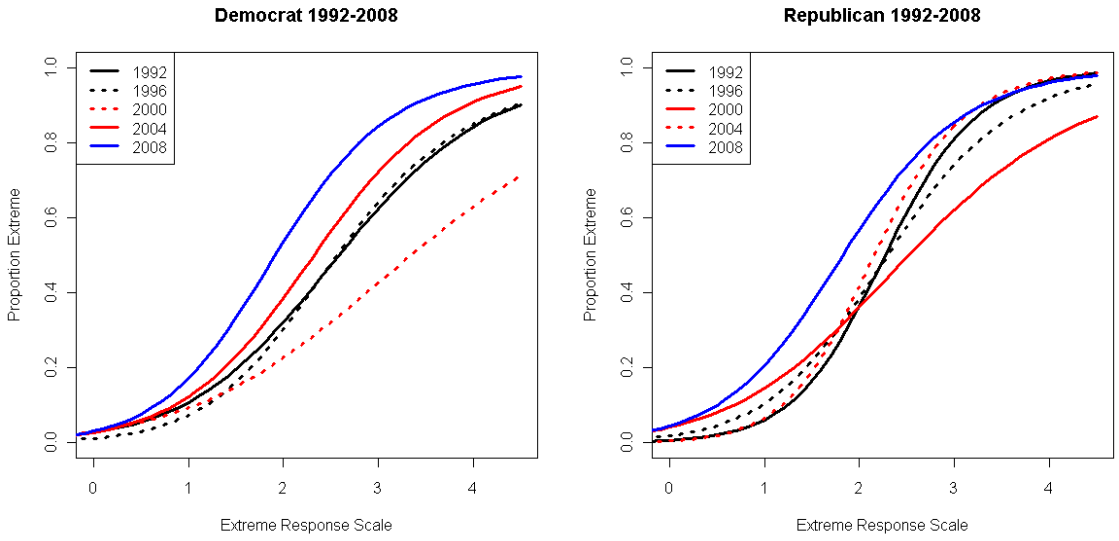
percent likelihood of selection in the range of 1.0 to 2.0 on the ERS scale). Perhaps even more striking is that the overall tendency to evaluate Liberals and Conservatives extremely is remarkably low overall (not exceeding 10 percent), yet the likelihood of an extreme response for no substantive reason among higher appears to be many times as strong. Due to the very low frequency with which endpoints are selected here and uncertainty in the estimation of ERS from year to year, it is an impressionistic venture to say that these charts are indicating an increased non-substantive Liberal endpoint response and a contextual and vacillating Conservative one. In particular, the 2004 measure for Conservatives appears to behave differently, as does the Liberal thermometer of 2008. Again, without a particular reason or specific trajectory, these results over time are difficult to ascribe to anything other than particular measurement anomalies generally. Yet there is an impressionistic story to be told that is plausible, and that story leans toward these labels, and the groups behind them, are being less deliberatively, and more habitually, evaluated.

Figure 3.7 Endpoint Selection due to Extreme Response Style: Liberals and Conservatives, 1992-2008



If there is a trend that is evident in the parties (Figure 3.8), it is an uptick in the level of Democratic Party non-substantive endpoint selection and again, a vacillating, seemingly contextual Conservative set of curves. Along with Conservatives, the 2004 Republican curve is below the others and has the mildest slope, while for the Democratic Party, 2000 is lower and 2008 looks like a consistent horizontal shift. Again, there is nothing particularly striking about the partisanship of these years that would allow a conclusion other than a particular aberration of measurement in terms of slope. As far as level, however, parties follow a more defined ascent in endpoint selection. More recent years tend toward the 20-40 percent endpoint selection range between ERS ranges in 1.0 to 2.0 for Democrats and a jump in 2008 for Republicans, though it is not apparent if the Republican estimate is an aberration in scale. Combined with an increase in endpoint selection overall, it is possibly more meaningful that respondents are much more likely to choose an extreme response that has little to do with an underlying feeling about partisanship or policy.

Figure 3.8 Endpoint Selection due to Extreme Response Style: The Democratic and Republican Parties, 1992-2008



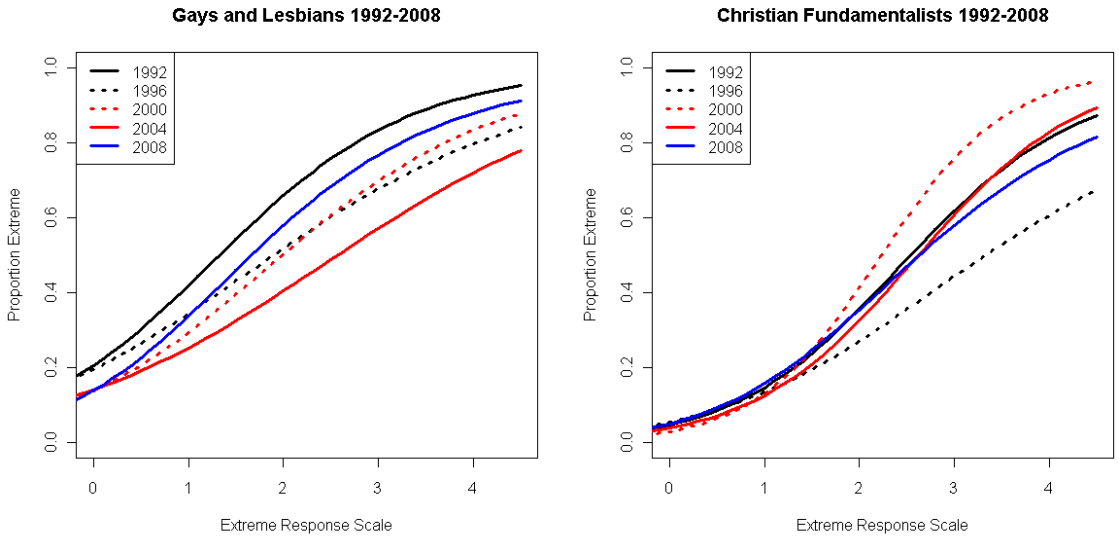
Put together, the overall story for ideological groups and parties is one that shows they are becoming very mildly less meaningful and relevant to respondents over this time period, or

are simply not being thought through. Among Liberals and Democrats, this tendency is slightly more discernible. On one hand, these results are not terribly consistent across years, enough so that the force of any conclusion wanes. On the other, a higher probability of endpoint selection combined with an increased probability that this endpoint selection is non-substantive may indicate that these partisan/ideological labels provide less real meaning in later years to respondents who are less aware.

Cultural Feeling Thermometer Items, Gays/Lesbians and Christian Fundamentalists

Turning to the more social identification and cultural feeling thermometers, the story is largely not of difference across time, but of the contrast between these thermometers and the "stock" feeling thermometers of liberalism and conservatism. These feeling thermometer items are of more interest given that cultural issues such as Abortion seem to be less strong in their ERS effect.

Figure 3.9. Endpoint Selection due to Extreme Response Style: Gays and Lesbians and Christian Fundamentalists, 1992-2008



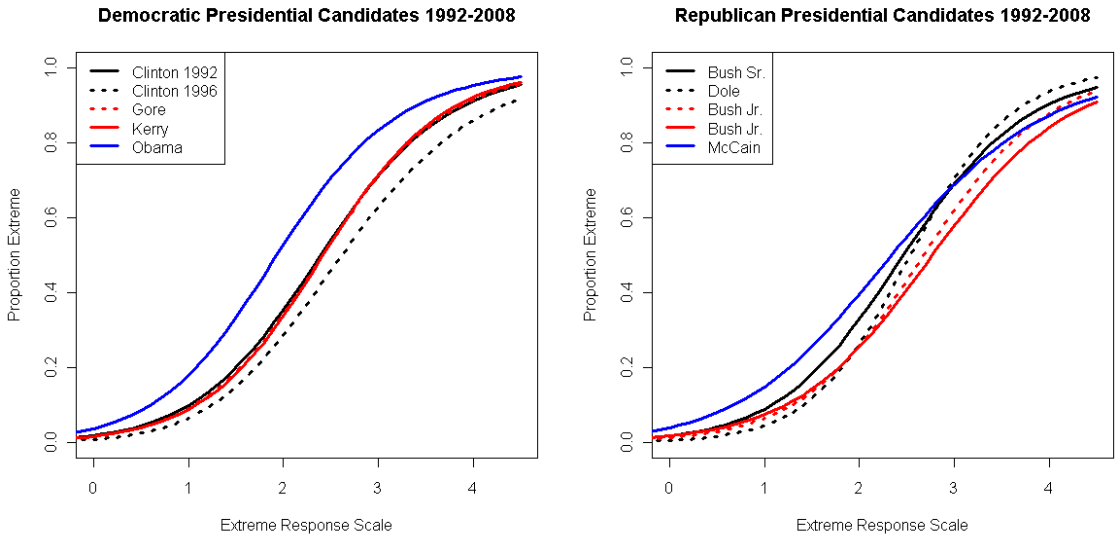
In these thermometers, both Gays and Lesbians and Christian Fundamentalists display no outstanding pattern of rise or fall over time, even taking into account the levels of non-

substantive endpoint selection. From Figure 3.9, if pressed, Gays and Lesbians show a mild decrease in the level non-substantive endpoint response over time. And, the Christian Fundamentalist thermometer shows particular responsiveness in 2000 and unresponsiveness in 1996, patterns for which there is little theoretical explanation. The main story, however, is the contrast from other political feeling thermometers, the comparatively flat trajectory in both, particularly among Gays and Lesbians. The results are clear in comparison to liberal and conservative core groups. Following the issue results, it may be the case that cultural objects elicit less automatic response again. Alternatively, however, because these represent groups of people rather than strictly partisan groups, respondents are less likely to pass summary judgment, that less overtly political objects are more thoroughly considered than political ones.

Democratic and Republican Presidential Candidates

The following takes a look at Democratic and Republican Presidential Candidates over time. In general, the provisional expectation here might be that ERS will be more marked for candidates who are less well-known (Clinton 1992, Obama, and Kerry in open elections, Gore as a VP in 2000), and less pronounced for Clinton 1996 as an incumbent. Likewise for Republicans, Bush Sr. and Bush Jr. in 2004 should be more recognizable.

Figure 3.10. Endpoint Selection due to Extreme Response Style: The Democratic and Republican Presidential Candidates, 1992-2008



These expectations, however, do not hold. In Figure 3.10, only Obama is an outlier – otherwise these two charts are the very picture of consistency over time. They are notably consistent given the fluctuation in endpoint selection that is characterized by the 1996 election candidates with an ebb in 2000. If Obama is an exceptional candidate, these data would seem to indicate it. However, given normal fluctuations and the exceptions in measurement in 2008, it may not mean that Obama or McCain are anything but in normal range. The consistency by level of endpoint selection in the upper ranges (1.0 to 2.0) is also remarkable, with the exception of 2008. For these upper ranges of response style, increase in endpoint selection that in non-substantive does not typically exceed 20 percent. What may initially seem very relevant, turns out to be quite mild. Even among those most prone to response style, true opinion for these respondents appears largely true. This is perhaps a testament to the exposure of the presidential campaigns, a real effort being made on the part of voters. It may also be a nod to the constraint of Political Evaluation at the average, where partisans and not moderates tend toward summary judgment, both of their own party and the other.

Discussion/Conclusion

Through a latent variable of extreme response style (ERS) and how issue questions and partisanship are subject to it, we can begin to see what it might mean to have automatic attitudes and how political ideas might be expressed summarily, whether an automaticity is indicative of the more derogatory, prejudicial type, a hallmark of low awareness, low compliance, or simply a product of incomplete deliberation. We also see where it finds outlets, and possibly more importantly where it does not.

Overall, three notable findings come from this analysis. The first is both methodological and substantive. These models confirm that response style is not, in fact, invariant to subject matter. I am cautious in this chapter not to overextend smaller conclusions that may stem from modeled and unmodeled uncertainty, given that year to year fluctuations in measurement may be truly contextual, or be induced by particularly strong or weak associations between particular objects. Even so, as far as validating the scale, important differences between items do exist in questions where there is a strong indication that they should, is absent in items where they should not be, and the properties of the ERS behavior is largely consistent with the covariates that are traditionally used to describe them.

Even with the prospect of unmodeled uncertainty, several overall notes bear further scrutiny. Most notably in ANES feeling thermometer items that are directly linked to political stimuli, there is more sensitivity to ERS which may not be surprising given the ambiguity of question format. Liberals, Conservatives, Democrats, and Republicans all have marked effects of ERS on endpoint selection that outstrip the effects of even other items of its identical type. The effective levels of endpoint selection due to non-substantive response are also very high comparatively. There is also some mild evidence that this tendency has increased during this

time period. From this, it is difficult to make an argument that partisan and ideological identification has become less important, or prioritized less over this time. Saying so would be antithetical to what we see as an increase in partisan polarization. Further saying that these partisan and ideological groups are somehow consistently less important than other groups does not comport with what we know about the consistent historical effect of partisanship and ideology on vote choice. What seems more likely is that these labels and groups have actually gotten harder for the least informed, and most moderate, in the mass public to process fully or deliberately. The full implication, if we choose to take it there prematurely, is that polarization, for these respondents, has little to do with conscious, deliberative choice and more to do with gut reaction.

It is also important to note where ERS is less influential, or absent which lends more credence to the "easy issue" hypothesis. Particularly on questions of political culture, traditional values issues and people (Abortion, Traditional Families, Moral Standards, Gays and Lesbians, and Christian Fundamentalists), respondents exhibit markedly less endpoint selection due to ERS. When it comes to cultural dimensions of the public mind, therefore, cultural values may be "easy" issues (Carmines and Stimson 1980, Carmines and Stimson 1986). The nuanced conclusion that comes from ERS is not the mere fact that cultural values are easy to comprehend in politics, but that they are well-considered in public judgment – and by virtue of this, may be characterized as high relevance or high priority issue areas. Being easy in this case does not mean simply that they fit well with other issues, or that we know which side we might find ourselves, but that cultural-dimensioned issues are, in fact, being deliberated less automatically on average. When it comes to cultural dimensions of political and social conflict, the public is less prone to extremity for the sake of extremity, but in contrast to other political

objects and issues, does not summarily judge as frequently, and consequently perhaps, less easily moved. This observation is important. In public discourse, cultural ideology has been hotly debated, and cultural issues have polarized over the last two decades. The contribution that ERS behavior makes to this debate is the possibility that, for the rancor in public discourse that might surround these issues, "the other side" may not be engaging in these issues lightly, as political discourse is likely to suggest.

What this analysis presents is a more complex picture of response style than has been confronted by experts in public opinion, and by overlaying the findings in this chapter, several conclusions come into focus. Some of these conclusions are inward looking. That is, what can be said about response style can illuminate what it is we believe we are measuring when we look at patterns of endpoint selection, or alternately the "range" of responses people allow themselves on an extended Likert scale item. One set of conclusions is inescapable when looking at items over time or whether we compare this behavior between items: the differences in how items behave can be stark. By question type, (1) feeling thermometer items in particular are far more prone to extremity than more contracted issue scales. This simple, and possibly intuitive, fact is more than simply an obvious fact, signifying that the construct itself which should be largely similar to other items that evoke ERS, are, in fact, quite strong in how they evoke ERS. (2) Yet even within feeling thermometers, a question set that varies only by a single referent, there are notable differences in how endpoint selection varies with respect to ERS.

The other set of conclusions are outward looking, and might serve as referenda on several epi-phenomena of politics we have come to observe. When contemplating the meaning of these results for contemporary politics, it may be useful to reiterate the literal meaning of

response style, then overlay what previous research and the previous chapters have found with respect to how ERS fits a model of cognitive simplification. Namely, the measure of ERS presented here, as in previous chapters is not simply a measure of extremity that parrots back to us what endpoint response frequencies dictate. The measure of ERS comports with the properties of ERS from multiple studies, spanning several decades – education, income, race, and political information (domain-specific ability) all run in the desired direction and remain consistent over these five election administrations. The multi-dimensional structure of the measurement insures that common variance contributed by both Political Evaluation (as identified by Feeling Thermometers in the political domain) and Social Welfare issues, the main substantive drivers of response in these items do not contaminate the ERS measure, and further, that the measure picks up a *pattern* of endpoint response that is common among all items which contain a scaled structure and not common among those items which are branched, or otherwise disallow endpoint response. While contamination of residual dimensions of substance is still a theoretical threat to validity, it does not come through in several cases for which there is a clear argument, in Abortion as a focal item, no statistically meaningful endpoint selection rise is evident.

It might be said that because a general pattern of endpoint response is rising, the question administration in the ANES could be deteriorating. Given the stability in these items and the history and consistency of the ANES, I find that explanation implausible. Another potential explanation for ERS is purely idiosyncratic, that respondents are employing judgment that is extreme on a case by case basis. It is plausible enough when we consider the presidential candidates, yet presidential candidates show little, if any variability in endpoint selection by item. As a pattern of response, one response in isolation is insufficient to identify ERS, which

requires, again, *patterned* extreme behavior. What is left in this measure, then, is a quantification of a pattern of response that is apparently independent of substance except primarily a tendency to choose endpoints with regularity, defined on a normal curve.

Why respondents behave like this can be debated. I have shown here and in previous chapters that the role of political information is influential, even when adjusting for education, income, and race. The role of cognitive simplification, once a side note in research, is emerging as one of the leading causes of response style. Yet cognitive simplification is not the whole story. It is yet unclear to what degree ERS represents a second structure in cognition, one in which a respondent requires a slightly more advanced baseline of information to form a rapid opinion, yet lacks a second, more deliberative component. Instead of "lack of awareness," one might call this tendency a "lack or refusal to deliberate on an object or issue." The latter seems a more palatable and nuanced definition, especially since it is possible, even desirable for respondents to economize in thought and judgment with respect to politics (Popkin 1992).

Whatever slice of definition one chooses, the increase in ERS with respect to issues and political groups is meaningful. Extremity as a response pattern is rising in these items, and it may not come as much of a surprise except that this extremity increase is sometimes very large and has little to do with question substance. And, as demonstrated in chapter 3, the influence of this pattern will modestly inflate correlations between items. The conclusion is inescapable: if non-substantive response patterns alone have become more influential in the past two decades, we will be overestimating the real relationships that are apparent in scaled data. Respondents will behave *as if* they have more consistent ideological impressions, but these impressions will be partly illusory and shallow. This research agenda, while potentially modest in scope, would

reinforce scholars who claim that polarization as a recent phenomenon in politics is not as worrisome on the margin as some may contend.

Stronger however, and foremost, these data strongly suggest that cultural referents and issues are much less likely to induce endpoint selection due to ERS. Abortion, Traditional Families, and Moral Standards are weaker than their other issue counterparts and Gays/Lesbians and Christian Fundamentalists as political objects are much weaker than their more staunchly partisan feeling thermometer items. This suggests two conclusions. First, that the mass public, and generally those who are giving the least thought to politics, are more keenly aware and cognitively involved in deliberating decisions about culture. This is not to say that culture is entirely deliberated across the public, but that it contains less of an automatic character. Secondly, the flip side of this observation is that endpoint selection among the most partisan objects, especially among feeling thermometers, are strongly influenced by ERS. By noting this, we can hypothesize that traditional group distinctions are the hardest for the least informed to grasp, and is possibly increasing. Importance or priority describes these results less well than concepts that are "easy" or "difficult," and the specific populations for which these easy of difficult references apply.

But it appears not to be simply an ease or symbolism alone that drives these main results. Symbolic cultural issues may be seen as easy issues, but it does not fit the strict definition of partisan or ideological groups to call these Democrats or Liberals a "difficult" concept to grasp. These labels fit at least two of the three conditions for easy identification, yet are the most difficult groups for this subset of the public to grasp: symbolic strength and prominence on the national dialogue, even while the ends of policy are sometimes multi-faceted and less than concrete. When searching for meaning in ERS toward variance in issues,

moderated and deliberated response fits the observations, a supposition I carry into chapters 5 and 6.

A 3.1: Latent Gold Syntax Example: Spending and Services

```
options
  algorithm
    tolerance=1e-008 emtolerance=0.01 emiterations=250 nriterations=50;
  startvalues
    seed=0 sets=10 tolerance=1e-005 iterations=50;
  bayes
    categorical=1 variances=1 latent=1 poisson=1;
  montecarlo
    seed=0 replicates=500 tolerance=1e-008;
  quadrature  nodes=10;
  missing  excludeall;
  output
    parameters=effect standarderrors probmeans=posterior profile bivariateresiduals;

variables
  dependent VCF0218o nominal, VCF0224o nominal, VCF0209o nominal, VCF0210o
nominal,
    VCF0211o nominal, VCF0212o nominal,
    VCF0424o nominal, VCF0425o nominal, VCF0426o nominal, VCF0427o nominal,
    VCF0809o nominal, VCF0839o nominal,
    VCF0886o nominal, VCF0887o nominal, VCF0890o nominal, VCF0894o nominal;
  latent
    CFactor1  continuous,
    CFactor2  continuous,
    CFactor3  continuous,

    Cluster nominal 1;

equations

  (1) CFactor1 ;
  (1) CFactor2 ;
  (1) CFactor3 ;

  CFactor1 <-> CFactor2 ;
  CFactor1 <-> CFactor3 ;
  CFactor2 <-> CFactor3 ;

  Cluster <- 1;
  VCF0218o <- 1 + Cluster + (a1)CFactor1 + (b1)CFactor2;
  VCF0224o <- 1 + Cluster + (a2)CFactor1 + (b2)CFactor2;
  VCF0209o <- 1 + Cluster + (a3)CFactor1 + (b3)CFactor2;
  VCF0210o <- 1 + Cluster + (a4)CFactor1 + (b4)CFactor2;
  VCF0211o <- 1 + Cluster + (a5)CFactor1 + (b5)CFactor2;
```

```
   VCF0212o <- 1 + Cluster + (a6)CFactor1 + (b6)CFactor2;
   VCF0424o <- 1 + Cluster + (a7)CFactor1 + (b7)CFactor2;
   VCF0425o <- 1 + Cluster + (a8)CFactor1 + (b8)CFactor2;
   VCF0426o <- 1 + Cluster + (a9)CFactor1 + (b9)CFactor2;
   VCF0427o <- 1 + Cluster + (a10)CFactor1 + (b10)CFactor2;
   VCF0809o <- 1 + Cluster + (a12)CFactor1 + (c2)CFactor3;
   VCF0839o <- 1 + Cluster + (a13)CFactor1 + (c3)CFactor3;
   VCF0886o <- 1 + Cluster + (c4)CFactor3;
   VCF0887o <- 1 + Cluster + (c5)CFactor3;
   VCF0890o <- 1 + Cluster + (c6)CFactor3;
   VCF0894o <- 1 + Cluster + (c7)CFactor3;

// constrain ERS
// set non-response equal to 0
// first category is zero (0) [x,1]

  a1[1,1]=0;
  a2[1,1]=0;
  a3[1,1]=0;
  a4[1,1]=0;
  a5[1,1]=0;
  a6[1,1]=0;
  a7[1,1]=0;
  a8[1,1]=0;
  a9[1,1]=0;
  a10[1,1]=0;
  a12[1,1]=0;
  a13[1,1]=0;

// missing are 0

  b1[1,1]=0;
  b2[1,1]=0;
  b3[1,1]=0;
  b4[1,1]=0;
  b5[1,1]=0;
  b6[1,1]=0;
  b7[1,1]=0;
  b8[1,1]=0;
  b9[1,1]=0;
  b10[1,1]=0;

// missing are 0

  c2[1,1]=0;
  c3[1,1]=0;
```

```
c4[1,1]=0;
c5[1,1]=0;
c6[1,1]=0;
c7[1,1]=0;
```

// fix extremity on ERS variables

```
a1[1,2]=1;
a2[1,2]=1;
a3[1,2]=1;
a4[1,2]=1;
a5[1,2]=1;
a6[1,2]=1;
a7[1,2]=1;
a8[1,2]=1;
a9[1,2]=1;
a10[1,2]=1;
a12[1,2]=1;

a1[1,9]=1;
a2[1,9]=1;
a3[1,9]=1;
a4[1,9]=1;
a5[1,9]=1;
a6[1,9]=1;
a7[1,9]=1;
a8[1,9]=1;
a9[1,9]=1;
a10[1,9]=1;
a12[1,7]=1;

a1[1,3]=a1[1,8];
a2[1,3]=a1[1,8];
a3[1,3]=a1[1,8];
a4[1,3]=a1[1,8];
a5[1,3]=a1[1,8];
a6[1,3]=a1[1,8];
a7[1,3]=a1[1,8];
a8[1,3]=a1[1,8];
a9[1,3]=a1[1,8];
a10[1,3]=a1[1,8];
```

```
a1[1,4]=a1[1,7];
a2[1,4]=a1[1,7];
a3[1,4]=a1[1,7];
a4[1,4]=a1[1,7];
a5[1,4]=a1[1,7];
a6[1,4]=a1[1,7];
a7[1,4]=a1[1,7];
a8[1,4]=a1[1,7];
a9[1,4]=a1[1,7];
a10[1,4]=a1[1,7];

a1[1,5]=a1[1,6];
a2[1,5]=a1[1,6];
a3[1,5]=a1[1,6];
a4[1,5]=a1[1,6];
a5[1,5]=a1[1,6];
a6[1,5]=a1[1,6];
a7[1,5]=a1[1,6];
a8[1,5]=a1[1,6];
a9[1,5]=a1[1,6];
a10[1,5]=a1[1,6];

a12[1,3]=a12[1,6];
a12[1,4]=a12[1,5];



// For testing new item, set endpoints to equality

a13[1,2]=a13[1,7];
a13[1,3]=a13[1,6];
a13[1,4]=a13[1,5];
```

Chapter 4. Acquiescent Response and Institutional Confidence: When Partisanship Informs Political Dissent

Having looked at Extreme Response Style (ERS) in previous chapters, this chapter modifies a modeling approach to attempt to capture Acquiescent Response Style (ARS), the tendency of a respondent to endorse agreement on a question independent of what a question is trying to measure. ARS has been identified in survey research as a bias to survey results, and consequently, most research is often weighted to prospectively correcting for respondent disposition in question construction.

More recently, however, studies have begun to see response style as reflective of both respondent disposition and content, the situational model. That situational model appears to bear out in chapters 1 and 3. Here I construct a model of acquiescent response using an item response model with World Values Survey Waves 4 and 5 data for American respondents (2001-2002 and 2005-2007). I then examine respondents' ARS agreement on fifteen questions about confidence in institutions. For institutions in the United States sample that are expressedly political, ARS is more influential than in the less political institutions – even where there is a greater percent agreement overall in less politicized institutions. Yet, when partisan identification is used, presumptively as a basis for disagreement, results show that reactions become less acquiescent. This is also true no matter what the leaning of the administration happens to be: in questions about Confidence about the Congress, the Government, or Political Parties. The implication for partisan identification reiterates political science research in that American political institutions induce more of a "kneejerk" acquiescence overall, yet this response is tempered by an affiliation with a partisan group. Partisanship, in its most basic function, appears to be one important way in which citizens may increase awareness from the

most minimally aware and engaged, to a threshold by which they can give at least some structured thought to government and politics response.

When confronted with a question that asks us to "Agree" or "Disagree," how seriously do respondents consider disagreement and how often do they agree without fully considering the alternative? This phenomenon of concern is commonly called *acquiescent response style* (ARS), the tendency for a respondent to agree with a question independent of the object of measure. An answer to these questions can be addressed by measuring acquiescent response explicitly and examining when and how we decide to agree on opinion surveys, and when we decide not to.

While it is true that ARS has intensely interested researchers, the analytical effort has largely been targeted toward prospectively fixing "bad" questions or assessing which questions contain response bias and why. Much research in survey construction is devoted to how to write questions that avoid acquiescent response by leading respondents to a conveniently favorable answer, the reverse coding of questions or otherwise inducing a respondent to carefully consider a response by disallowing neutral responses (Schaeffer and Presser 2003). Alternately, when identifying and measuring ARS, the focus is often geared toward assessing question bias, and less toward what this bias might say about respondents' situational use of acquiescence.

ARS is undeniably a nuisance, confronted across all disciplines that use agreement modes in survey scales as an important mode of evidence. And it would make less of a difference if response styles were empirically random, where consequently random deviations would be rightly categorized and modeled as measurement error. However if there is a

systematic association between personal attributes or specific content and ARS, point estimates using dependent variables contaminated with ARS may not simply have larger confidence bands, but be systematically biased (Green and Citrin 1994). Simply and formally, in factor analytic notation, we might view responses (x) as related through one or more factors (ξ_i) with loadings (λ_i):

(4.1) $x_{1i} = \lambda_{1i}\xi_i + \delta_{1i}$

(4.2) $x_{2i} = \lambda_{2i}\xi_i + \delta_{2i}$

Assuming here that cov(δ_{1i}, δ_{2i})=0, this simplified, unidentified example is measured without bias, yet if there is a positive correlation between the errors, we have a situation where λ_i will be inflated accordingly – situation where response style would be a pressing concern.

Little effort, however, has been devoted to why some questions may be easier to agree with than others and how they might provide insight into substantive political questions. Instead of informing bias in scales [cov(δ_{1i}, δ_{2i})>0] to better estimate λ, another factor may be introduced that has substantive meaning. To this end, I construct a model that looks to detect *situational* response, addressing not just who has a tendency to agree, but how that tendency is brought out by the content of a question. While some respondents might be more predisposed to agree than others (the *dispositional* thesis), questions should also vary in their ability to induce respondents to agree. The model in this inquiry examines whether survey content and respondent tendency are interactive by nature, respondents will have a tendency toward endorsing agreement, but this trait will be evoked through content to differing degrees. This study begins to assess whether it may be useful to use ARS as a means toward examining how we respond automatically to different survey questions, and what may play a part in this automatic behavior (Bargh and Chartrand 1999).

Formally, then, in factor analytic notation, we can introduce an ARS factor ξ_{ARS} to incorporate as a constructive remedy to bias that is inherent in equations (1) and (2), but in this analysis, to assess the factor itself.

$$(4.3.a)\ x_{1i} = \lambda_{1i}\xi_i + \lambda_{ARS}\xi_{ARS} + \delta_{1i}$$

$$(4.4.a)\ x_{2i} = \lambda_{2i}\xi_i + \lambda_{ARS}\xi_{ARS} + \delta_{2i}$$

In this way, we might be more confident that the other model parameters are more robust, but we might also glean the character of ARS from a well-constructed model, notably ARS that is measured independent of the other latent factors we intend to measure (ξ_i). This may also serve as a basic measurement model that might show how ARS (ξ_{ARS}) and politics (ξ_i) may inform each other, and when a respondent chooses to inject more considered judgment depending on political stimuli.

Thus, this chapter focuses on the substantive character of response style, a prerequisite of which is modeling response data in surveys of politics, revolving around two main themes: (i) the construction of a simple, yet more appropriate measure of acquiescent response than previous literature has devised and (ii) exploration of the connection of acquiescent response and the role of institutional confidence. The following sections lay out the nature and lineage of research in acquiescent response, the conceptual connections of political research and acquiescent response, and how we might more adequately and precisely measure this concept.

Because ARS does not always occur in survey data consistently or with enough strength to measure adequately (Baumgartner and Steenkamp 2001), I look for acquiescence in a survey where it is both highly likely to occur and be detected. I use the World Values Survey Waves 4 and 5 from 2001-2002 and 2005-2007 (n=1,200 and n=1,249 in the United States respectively)

to assemble a diverse set of questions consisting of scales that ask a respondent's agreement with a statement. The domain of the World Values Survey helps by asking questions about values, about which many respondents have low information, requires judgment with a high degree of abstraction, and toward content that is often ambiguous in meaning.

These questions, used to measure a common acquiescent response trait, are then used to evaluate agreement when respondents assess their own confidence in a number of prominent institutions in American society, governmental and non-governmental. If questions conform to ARS theory generally, I contend that respondents should evaluate institutions unfavorably when they have an immediately accessible reason to do so, and absent an immediate reason, will tend to acquiesce more often instead of perhaps abstaining from response altogether. This opens the door for hypotheses about how questions might differ in the strength of negative cues. Respondents may find these cues as part of their partisan social identity (Green, Palmquist, and Schickler 2002) or engage in left/right decision-making in a more or less automatic fashion. To the degree to which institutions are portrayed negatively according to a respondent's partisan identification, partisans may find more politicized institutions easier to disagree with, a finding which is more than just trivial. It is one thing to detect that partisans will disagree with their opponents more often, yet it is an entirely different matter to detect that partisans will "not acquiesce" to a question based on their partisanship *independent of content*. This chapter reinforces that part of the value of partisan identification is to help the mass public cue disagreement in accordance with their underlying values, not simply by registering an opinion that would be otherwise opposite, but by effectively registering an opinion *at all*. Thus, partisanship may mitigate acquiescent response bias, and consequently allowing for a formation of opinion / effective representation where there was previously none.

Acquiescent Response Style

Acquiescent response style has been a concern of survey statisticians for over a half century as part of a larger potential threat to the validity of survey results commonly called response styles (Jackson and Messick 1958) or response sets (Cronbach 1946, Cronbach 1950). The concern over response sets came from early observations of how respondents reacted to the survey situation and was reinforced by early statistical results. Attitude strength in surveys seemed to be incompletely, and erratically observed. Additionally, these response tendencies appeared to coalesce around certain response preferences that appeared independent of the survey question content (Paulhus 1991), part of preference instability that is well documented (Bassili 1996). For attitude research, these observations presented difficulties that came in the form of two initial questions - do response sets (or response styles) exist and who uses them?

For decades, the existence of response style proved to be the central debate. Some scholars were skeptical of response style from the outset (Rorer 1965), yet slowly the consensus has shifted (Messick 1991, Ray 1979, Ray 1983). Today there is little doubt that response styles exist, which opens the natural line of inquiry that delves into who uses them and when. Various theories have been developed about why people choose responses in a way that is content-irrelevant; in these theories, three major threads have emerged. One stresses the individual as the source of bias, exploring personal traits, usually cognitive ability and personality (McGee 1962, Meisenberg and Williams 2008, Marin et al 1992). Another speaks to cross-cultural causes, common approaches to survey situations that are shared due to common cultural understandings (Cheung and Rensvold 2000, Smith 2004, Johnson 2005). The last speaks to the interaction of the individual with the subject matter, noting that specific survey item formats and specific content areas served as a prerequisite for response behavior

(Baumgartner and Steenkamp 2001, Tourangeau and Rasinski 1988, VanBeselaere 2004). That is, if low cognitive ability is the individual cause of response style, questions that the respondents know little about will create the conditions for arbitrary response behavior.

Study after study show that ARS exists, yet the more difficult problem has been to identify exactly *why* it exists. Explanations for ARS are numerous and not always consistent, yet prevailing research has found several stable individual level indicators of response style which loosely center around the respondent's cognitive ability and engagement (Meisenberg 2008). By far, the most compelling theoretical argument for response behavior generally comes from the optimizing model of response that follows a more interactive approach. Ideally, when a subject responds to a survey, she optimizes: derives the intent of the question, searches available memory for attitudes and associations, reaches a judgment based on those attitudes, and translates that judgment onto an available survey response (Krosnick 1999). The tendency for a respondent is to use heuristics to decrease the cognitive load of a survey and circumvent optimized response. When a respondent is motivated to respond truly and has the wherewithal to complete this process, our surveys are minimally affected by measurement error or response bias. However, when response is complicated at any stage of optimization, as the cognitive load of the question increases, subjects will have the tendency to fall back on simplifying constructs. Ambiguity, content which evokes conflicted individual experience, and answer choices that do not match well with a respondent's subjective evaluation will prompt what has come to be called “satisficing” response behavior. When respondents acquiesce, they are telling us that they either do not know enough, or do not care enough to go through a more intensive evaluation of options and respond the way we might prefer. It is this understanding that is the first point of reference in this paper and follows from previous chapters’ results:

acquiescent response, like other response styles, is an artifact of agreement that, in part, comes from low information.

The implication of Krosnick's conception with regards to ARS is straightforward, but limited. Some complications arise in restricting the theoretical argument to knowledge and survey engagement in this manner. Namely, response disposition often carries a uniquely cultural component (Marin et al 1992, Cheung and Reensvold 2000). If we presuppose that ability and engagement are the primary drivers of ARS, artifacts of a common social approach to surveys may confound satisficing behavior and may exhibit significant pull within a given context (Meisenberg and Williams 2008). Gendered response is also a touchstone case of the idiosyncratic nature of acquiescence. Women have shown distinct response tendencies in some cases, highlighted by the tendency for women to provide no answer rather than guess, (Mondak and Anderson 2008) even controlling for overall knowledge. Research has now gravitated to the conclusion that context matters, whether it be cultural context, or situational. There is no doubt that respondent characteristics matter and question quality matters, however a contextual interpretation of survey response introduces the natural question: does content matter? Instead of assigning culpability to disengaged respondents, a cultural tendency, or inept question writing, can an otherwise "good" question elicit "bad" response due solely to content?

A different look at ARS emerges from this inquiry, one that marries both a respondent's tendency to endorse agreement and the question's ability to draw it out that forms the basis of a the justification hypothesis. Because disagreement requires more of the respondent in terms of knowledge and engagement, it is more costly than agreement. In order to reject a statement in a survey, the respondent needs to muster the cognitive resources and will to disconfirm an implicit attitude offered by the survey instrument. This tendency is conceptually related to

social psychological research concerning subjects who choose not to deviate from expectations even when the cost of passive compliance is high (Cialdini 1984). The tractable psychological rationale can also be viewed more aptly from the negative point of view, when a respondent disagrees, there must be an accessible reason to do so. If none exists, then the tendency is to comply with the statement as a default answer.

The immediate question that comes from a focus on response style is: "as a political scientist, why do I care"? Simply put, because we care about attitudes and the survey response is how we measure them. The first, and most obvious, reply revolves around the position of survey research in political inquiry. A vast body of inquiry on politics, political attitudes and opinions, and vote choice come directly from the tradition started with the Michigan studies around vote choice. Where we look for answers about the nature of American political thinking is, for better or worse, where we have the most information, and most of it comes from the closed-end survey. What Bartels notes about voting behavior is certainly true of why we look to surveys: because "that is where the best data are" (Bartels 1988). Second, political scientists care deeply about how to measure attitudes correctly. Research in political science has shown unequivocally that attitudes matter. And for many reasons they are central to our understanding of political behavior, not the least of which is that strong attitudes are key behavioral predictors (Miller and Peterson 2004).

But measurement work need not always be relegated to the status of contaminant. After decades of study, we know quite a bit about the nature of response style that it may be useful in and of itself. One way it may be useful is by furthering our understanding of political cognition. To the extent that cognition and engagement are shaping influences of response behavior, measuring these tendencies will be a revealed indicator of how respondents think, the

character of the environment in which they think it, and provide insight into the nature of the content itself. The classic framework for thinking about how the electorate chooses to engage has hinged on whether members of a polity possess the requisite skills and knowledge, possess the motivation, and are asked to do so (Verba et al 1995). For survey research, the opportunity prerequisite is always met: if you are being administered a survey, we are asking, and it is the norm in large scale surveys that we also spend a great deal of effort to get higher response rates. If response style is a measure of our knowledge of, and engagement with, the subject matter at hand, respondents should acquiesce in accordance with how knowledgeable and relevant questions are to them - *the situational model*. When we ask about confidence in society's institutions, then, we might expect parties to matter for partisans, churches to matter for the religious, and so on. We might expect certain questions to be more or less relevant to respondents, exacerbating or mitigating acquiescent response. If acquiescent response is situational and due to a respondent's knowledge of and engagement with the target of evaluation, how might we expect ARS to differ between questions?

When ARS fundamentally differs between questions, we might find that it is helpful in distinguishing certain subsets of questions from others in their response characteristics. I look at a subset of fifteen items devoted to evaluating confidence in various institutions in society. The items range from confidence in national government broadly as well as an array of prominent governmental and non-governmental institutions that are more removed and independent from partisan debate. Scholarly work in the area of confidence in government broadly is still developing expectations about what it means to have confidence in government and what this confidence is in context (Catterberg 2006, Nevitte 2003), yet a few properties of

confidence appear to be clear cut - namely the difference between evaluation of an institution's performance and an evaluation of an institution's legitimacy more generally.

When inspecting differences in the context of these questions, a key feature of these questions is their consistent question format and proximity in the survey - the questions are only different substantially in their object of measure. The expectations for how they might differ in how they evoke generalized, less thoughtful, automatic agreement follow:

First, we might expect ARS to be a part of the reason why respondents have confidence in institutions – that is, on some level, institutional confidence is a default answer to an ambiguous question. In theory, then, ARS *should be present in all questions that prompt agreement* (Weijters 2006) and this should be no different for a question of institutional confidence.[22] If questions contain a default value which is agreement, and information is needed to stray from this default value, the degree to which ARS is present in evaluating an institution will indicate how an institution evokes lack of awareness or engagement.

Second, it is unclear what we might expect from partisan affiliation, however it might be logical to expect several directions of influence, but, at least, *we might expect ARS to substantively differ with respect to the partisan nature of the object*. First, there is a facet of confidence that reflects performance evaluation in a current government regime (Hibbing 2001). This regime evaluation is separate from confidence in democratic or non-governmental institutions generally. Confidence in government, both internationally and in America, will

[22] There may be some doubt as to whether the text of the question as written in World Values is situated toward ARS. The argument hinges on "confidence" as a default answer, to which the possibility that it is not a default must be addressed. As both a hypothesis and validity check, to proceed with analysis, this must be true, and it holds in the results section of this chapter:

Question Text: I am going to name a number of organizations. For each one, could you tell me how much confidence you have in them: is it a great deal of confidence, quite a lot of confidence, not very much confidence, or none at all?

contain these two separate referenda about legitimacy in process and in performance evaluation. Thus, in the evaluative context, partisan messages and party identification should provide a more meaningful base for respondents to rationalize any deviation from a favorable response: "I am a Democrat, the current president is a Republican, therefore I disapprove." This hypothesis revolves around the relative accessibility of a rationale to agree and disagree.

To reiterate, two concrete expectations emerge that should make respondents more or less likely to express non-confidence in societal institutions that is indicative of *situational justification*. If a respondent wants to convey dissatisfaction, the respondent must have a rudimentary rationale to do so to overcome ARS, provided by the context of the question or by partisan cues. If ARS differs between questions, where does it differ? And where might ARS be stronger - for non-partisan institutions generally, partisan institutions? Should partisans agree more when evaluating an object that comports with their partisan identification and agree less when it disagrees with standing partisan identification, or is ARS exacerbated or muted by partisan identification more broadly?

Measuring ARS

The model I develop here is fairly straightforward in concept and reflects the factor analytic argument of equations (1-4). In any one question, it is impossible to separate preference from style. With no other data, it is antithetical to this effort to consider that an individual's agreement on a question is somehow false in and of itself. However, among a large number of questions about different, tenuously related subjects, we can say whether some respondents will tend to agree more with questions than disagree. From this tendency to agree, we can start to make inferences about whether the pattern of automatic agreement says something about institutional confidence.

Traditionally, the way to measure ARS has been to count agreement responses, but this does not perform as consistently as factor analyses and structural equation models (Cheung 2000, Billiet 2008). For a number of reasons, counting acquiescent responses may be adequate for inspection, but less than ideal measurement strategy overall. The main drawbacks of counting are the loss of precision and contamination. While acquiescence is a meaningful construct, it is a smaller portion of overall variance than constructs that are explicitly designed, making precision a very important issue to address. It is not the norm that questions are designed in scales that will have very little covariance, and, from the standpoint of ARS, this covariance invites contamination. Another drawback of counting / summation is the assumption that each question will contribute equally and automatically to ARS, which precludes a critical analysis of how the variable performs as a measurement construct. The alternative is to model ARS, not as a sum score of different items, but as an underlying trait that influences the response to individual questions. This opens the door not only to measuring an underlying common construct (ARS) more effectively, but to inspecting how each question relates to the common construct. More simply put, *the frequency with which a question elicits agreement will not necessarily mean that the question elicits a pattern of automatic agreement*.

Even with this approach, ARS is notoriously difficult to measure reliably. Several validity concerns make the task of isolating ARS complex. First, if the questions are related, the difference between agreeing on, say, 30 questions and 45 questions out of 50 is not a respondent's tendency to agree at all, but measures a common substantive association. For a respondent to agree that George W. Bush is an effective leader and that Dick Cheney is an effective Vice President does not identify ARS. This exemplifies the primary measurement concern: as much as possible when measuring ARS, we need to isolate the acquiescent response

component from survey content. The response style in this exemplar is hopelessly confounded with another construct – namely partisan politics. One solution for this problem is to "reverse code" half of these questions. This disconnects the content of the survey from a general agreement or valence of response. Yet this strategy is only available in secondary sources on occasion (the World Values Survey, as an amalgam of different approaches is not consistent in this regard). The alternate approach is the one I use here. I utilize the inherent size and heterogeneity of the World Values Survey as a means to identify acquiescence. In this survey environment, it is easier to assemble a set of questions to make the items, placed together, as devoid of content as possible - effectively allowing a statistical model to more easily break apart content and response. Conceptually, this means minimizing ξ_i as much as possible by design and validating that a common acquiescence is being extracted ($\lambda_{ARS}\xi_{ARS}$).

$$\text{(4.3.b)}\quad x_{1i} = \cancel{\lambda_{1i}\xi_i} + \lambda_{ARS}\xi_{ARS} + \delta_{1i}$$

$$\text{(4.4.b)}\quad x_{2i} = \cancel{\lambda_{2i}\xi_i} + \lambda_{ARS}\xi_{ARS} + \delta_{2i}$$

One concern that is readily apparent is how we might be certain that the residual influence of ξ_i, while minimized, does not have undue influence on the result. This concern requires some validation that the construct behaves as we might expect. Another concern about this approach revolves around an item's contribution to acquiescent response. It may not suffice to say that each question contributes to response style in the same way. Intuitively, we know it may be easier to comply with some statements than others. We may know more about some topics, and on others be reluctant to generalize. If we are to assess an individual's propensity toward evaluating political groups and people, it is not strictly correct to treat one's decision as

equally measured for every respondent by George W. Bush or Hillary Clinton as opposed to a larger class of "Young People" or "Catholics." For this reason, I deviate from the factor analytic approach presented thus far and allow each of these questions to vary according to the frequency of agreement (difficulty) in an Item Response Model, a factor analytic analogue for categorical data.

In sum, a comprehensive measure of acquiescence should, as much as practicable, (a) separate itself from content and (b) account for differences in each question in their frequency of agreement. The measurement approach I propose here is a close cousin to counting acquiescent responses, yet allows the questions to vary in a constrained way, including a process of purposive selection and requiring that each question be treated as equally indicative of ARS. These steps minimize the confounds of content and contributing to acquiescence using item response using items which are not substantially related. The end result is an estimate of acquiescent response on a standardized metric for each member of the survey which will serve as the base for the results presented here. To accomplish this in a way that attempts to purge the data of its associated content, I follow a process in which I first collect all measures with a valence of agreement from the World Values Survey and examine their content and pairwise correlations in order to reduce the confounding dimensions of substance.[23] By eliminating questions with high substantive covariance, we are left with a pool that attempts to attain the highest diversity of subject matter and lowest inter-item correlations.

Again, the World Values survey provides a large bank of questions from which to draw. The survey is fairly large, is adequate in its collection of demographic covariates (namely age, gender, education, and income to validate ARS), and contains a diversity of questions in

[23] Items with pairwise Pearson correlations above 0.40 in Wave 5 were excluded from analysis

content and measurement intent that make liberal use of items that ask the respondent to agree or disagree with a statement. To prompt ARS, the question must make a one-sided assertion that challenges the respondent to either affirm that assertion or find a rationale to defect from it. That is, the question provides a unidirectional statement and requires the respondent to either ratify the statement or deny it, with denial requiring the resources to do so. This narrows down the initial universe of questions to 43, winnowed to a smaller set of more diverse questions (23 in 2006 and 18 in 2001-2002) (Appendix 4.1). I divide the response options into three categories (0=general disagreement, 1=non-response, and 2=general agreement/acquiescence)[24] further restricting any substantive association. In a final effort to remove any residual content of the items, the acquiescent response strength is assumed to be equivalent between items. Regardless of content of a specific question, the acquiescence will behave similarly in these base items. The result will be a better identification of general ARS, which may aid in distributing any adverse effects that may result from residual dependence between questions. This diverse set of items constitutes a base of measurement for ARS, or *anchor items*, maximizing common variance based on agreement while minimizing the confounding effects of survey content.

Analytically, the nominal response model (Bock 1972) that is the basis for this inquiry consists of a (multinomial) logistic model which predicts the probability of selecting category *k* in item *i* given a latent dimension of ARS, θ_j. The probability of selecting category *k* in item *i* also relies on the item's properties λ_{ik}, commonly called the category slope, and ζ_{ik}, a constant intercept for each category. To identify the model, as in a categorical dependent variable

[24] Contrary to Meisenberg and Williams (2008) which allows acquiescence to be negatively associated with the extreme negative value, this model is based only on cooperation with the question or defection. Non-response is included as a valid category to account for its uniqueness and relative frequency, but not allowed to associate with ARS directly.

regression, the category boundaries are constrained: $\sum_k \lambda_{ik} = 0$ for each dimension and $\sum_k \zeta_{ik} = 0$. For agreement items, three unordered categories are used (agreement, disagreement and non-response.[25] Notably, without the category of non-response, the model reduces to the more common two parameter logistic (2PL) model, the more ubiquitous item response model (Lord 1952).

Nominal IRT Specification:

$$(4.5)\ P_{ik}\left(X_{ik}\middle|\theta_j\right) = \frac{\exp\left(\alpha_{ik}\theta_j+\zeta_{ik}\right)}{\sum_k \exp\left(\alpha_{ik}\theta_j+\zeta_{ik}\right)}$$

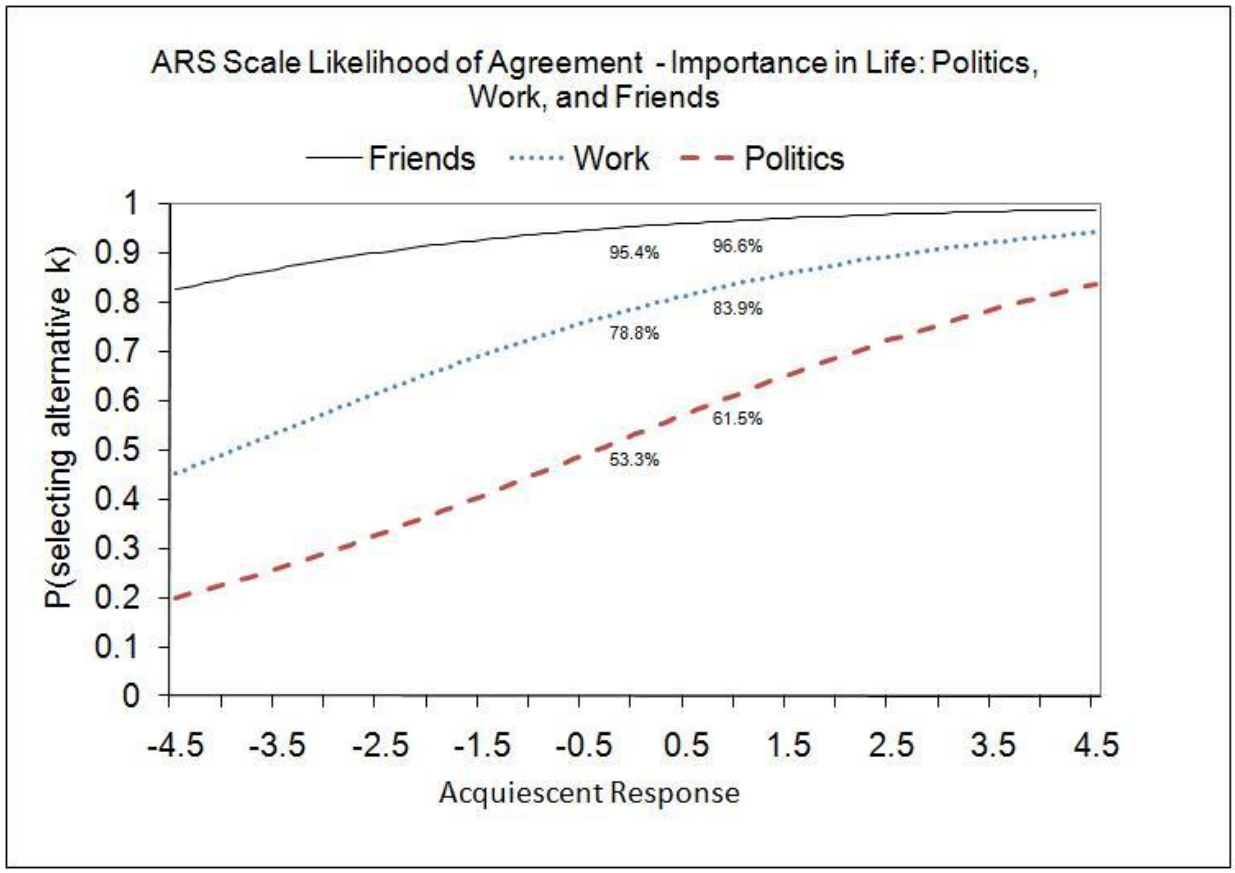
Intuitively, by modeling response this way, the parameters of interest (θ_j, α_{ik}, ζ_{ik}) communicate each respondents tendency to agree (θ_j, a standard normal measure), the common propensity for each question to provoke agreement α_{ik}, and an intercept ζ_{ik}, roughly the point at which a respondent is just as likely to choose the category relative to any other on a specific question at every level of the underlying ARS.

An example of how these effects are modeled appears in Figure 1. The questions of the importance of politics, work, and friends are plotted with respect to a latent tendency to agree (θ_{ARS}). From this figure, generally, friends are very important, work less so, and politics a distant third. The labeled percentages show the differential effects from this model for moving from an individual with an average level of overall tendency to agree ($\theta_{ARS} = 0$) to an individual with a 1 standard deviation greater tendency to agree ($\theta_{ARS} = 1$). The logistic ogive curves are parallel by design, yet the location of the curves gives the model more meaning in

[25] Nonresponse is allowed to occur, but is restricted to a zero slope, $\lambda_{i(NR)} = 0$, that is neither informative for ARS nor confidence in political institutions. This modifies our interpretation of each to be "conditional upon response," that is, those who choose to participate in the scaling exercise are evaluated only on items in which they choose to participate. Estimation of item parameters for this model were conducted in the Latent Gold software using Marginal Maximum Likelihood (Vermunt and Magdison 2005).

term of the modeled marginal effects (0.19 in Figure 1 for Wave 5, and 0.13 for Wave 4). The more acquiescent respondent will be only 1.2 percentage points more likely to agree on the question of friends, yet 8.2 percentage points more likely to agree on the question of politics with the same underlying change in propensity to agree overall.

Figure 4.1. Likelihood of Agreement as a Function of ARS



In sum, the model of ARS constructed in this paper seeks to isolate, as much as practicable, the common and positive loadings that response style possesses simply due to automatic agreement. The modeling involves estimating a number of different parameters, however, the parameters of interest in the base model – equation (5) – can be boiled down to a common propensity that is exemplified in Figure 4.1.

Assessing the Validity of Acquiescent Response

This approach to measuring ARS in this way has several advantages. A latent variable model such as this allows us to inspect how well the data comport with our conception of how ARS should behave in theory. Two expectations of ARS are examined here with respect to the base questions. The first expectation for this common factor is that ARS represents a construct that is roughly common to all questions of agreement. It is not identical, yet generally small and positive. The second is to validate this measure based on our conception of ARS - namely, in concordance with theory, that education, age, and income are negative predictors of ARS.

To the first question I inspect a rudimentary diagnostic principle components analysis on the dichotomized agreement variables in Table 4.1 (FA^1). The primary factor does conform to expectations, namely that it exhibits positive factor loadings on nearly all the variables included in the model. Some exceptions do occur, but at least on inspection do not disconfirm the theory of why response style might vary. The measures that load strongly on the common factor, revolving around being a member of the community and country, are at least intuitively difficult to disagree with, while the negative wording of *"marriage is an outdated institution"* is, by these models, perhaps an easier question to disagree with. At least on its face, the ARS assumption for the common factor does not appear invalid.

Table 4.1. Factor Analyses Using 23 Variables in Wave 5

Question	FA1^	FA2*	1	2	3	4	5
Eigenvalues			2.49	1.68	1.46	1.24	1.23
a002 Important in life: Friends	0.77	0.58	0.26	-0.13	-0.41	0.31	0.15
a003 Important in life: Leisure time	0.35	0.19	0.21	-0.22	-0.28	0.46	0.17
a004 Important in life: Politics	0.55	0.38	0.31	0.11	-0.41	-0.01	0.09
a005 Important in life: Work	-0.04	-0.02	-0.02	0.10	-0.15	0.42	0.36
b001 Would give part of my income for the environment	0.29	0.07	0.24	-0.28	-0.22	-0.42	0.40
b003 Gov't should reduce environmental pollution	0.20	-0.02	0.17	0.08	0.50	0.45	0.01
c001 Men should have more right to a job than women	0.08	-0.06	0.12	-0.36	0.12	0.30	0.05
c002 Employers should give priority non-immigrants	0.33	0.10	0.18	0.14	0.25	0.31	-0.17
c036 To develop talents you need to have a job	0.23	0.07	0.13	0.34	0.28	0.07	0.18
d018 Child needs a home with father and mother	0.58	0.32	0.28	0.51	0.04	-0.13	-0.11
d022 Marriage is an out-dated institution	-0.69	-0.64	-0.26	-0.22	0.34	-0.15	0.24
d054 Goals in life has been to make parents proud	0.74	0.35	0.38	0.26	0.02	0.04	0.35
d055 Make effort to live up to what friends expect	0.61	0.26	0.26	0.31	0.02	-0.11	0.44
d057 Being a housewife just as fulfilling	0.70	0.34	0.38	0.05	-0.18	-0.08	0.04
*d078 Men make better business executives than women	0.54	0.21	0.19	0.46	0.21	-0.11	-0.07
*d079 I seek to be myself rather than to follow others	0.96	0.37	0.42	-0.21	0.06	0.11	0.07
f104 People with strong religious beliefs in pub office	0.81	0.35	0.35	0.42	-0.08	0.08	-0.17
f105 Religious leaders should not influence government	0.10	-0.16	0.14	-0.42	0.36	0.01	-0.05
*g020 I see myself as member of my local community	2.43	0.94	0.66	-0.11	-0.03	-0.11	-0.27
*g021 I see myself as citizen of the [country] nation	4.51	1.54	0.71	-0.24	0.15	-0.03	-0.27
*g023 I see myself as an autonomous individual	0.46	0.06	0.32	-0.22	0.19	-0.20	0.12
e115 Political system: Having experts make decisions	0.11	-0.10	0.11	-0.01	0.44	-0.10	0.48
e117 Political system: Having a democratic political system	0.94	0.33	0.48	-0.22	0.04	-0.18	-0.07

Columns Labeled (1-9) represent and initial diagnotic PCA (Eigenvalues >1)

^FA1 represents a unidimensional factor analysis for categorical variables

*FA2 are the loading estimates for agreement for the unconstrained NIRT model

Other results in this table also point to the primary factor as ARS. The second column introduces an IRT framework, for display, which largely confirms the more simplistic factor analytical results from the first column. Alternate (though very similar) specification does not nullify results. Moreover, the next five factors show that some constructs in this question battery are idiosyncratic. They show a mix of positive and negative loadings that we might

expect is a hallmark of a nuisance factor, and not a factor common to all questions that we might call ARS.

All in all, the factor structure here is a very strong indicator that what we are measuring is, indeed ARS. Predictably, the loadings are not as strong as if we were attempting to measure an intentional factor, one explicitly measured as part of survey design. As a result, numerous dimensions may remain that may confound the analysis – as exemplified in factors 1-5 in Table 4.1. The eigenvalues in the principle components analysis indicate a primary factor with common variance of 2.49, which indicates in this simplified look, that slightly more than 10 percent of the variance in these items is attributable to the first factor. In this light, the model appears to have succeeded in pulling out a primary common theme, even if not strong. Confirming this consistency, a one-dimensional factor analysis for dichotomous dependent variables,[26] and an unconstrained item response model produces “loadings" in turn that show a similar pattern between items as our initial primary factor. Nonetheless, to counteract any potential bias in how these items contribute to ARS, even at the expense of some precision or fit, for simplicity and identification ease, I impose constraints on the questions so they are equal contributors to the latent dimension of agreement.

To the second question of validity, results are predictably more mixed. Several covariates are inspected that have been examined in previous ARS literature. The specific variables of education, income, age, and gender are inspected with similar questions in Waves 4 and 5 to compare with a previous study (Meisenberg and Williams 2008) and are simultaneously estimated with the measurement model for ARS by adding the regression equation to (5) for this basic set of variables:

[26] The factor analysis for categorical dependent variables conducted using MPlus (Muthen 2002)

$$(4.6.)\ \ \theta_j = \beta_0 + \beta_k X_k$$

If the measurement model is consistent with optimization ideas about acquiescent response, we will expect ARS to decrease in age (variables *agemid* and *agehigh*), education (variables *edmid* and *edhigh*) and income (variables *incmid* and *inchigh*), but not with gender (variable *female*) in most cases. Gender has intermittently been positively associated with response style, the strongest evidence for which has come when examining acquiescence.

Table 4.2. Modeling ARS - Determinants/Covariates of ARS

DV:	ARS Wave 4 (2001-2002)			ARS Wave 5 (2005-2007)		
Estimate	b	SE	p	b	SE	p
Intercept	1.473	0.777	0.058	0.850	0.288	0.003
female	0.392	0.143	0.006	0.332	0.100	0.001
age2	-0.078	0.168	0.640	-0.006	0.143	0.970
age3	-0.589	0.190	0.002	-0.468	0.140	0.001
edmid	0.389	0.207	0.061	0.028	0.105	0.790
edhigh	0.514	0.163	0.002	0.132	0.172	0.450
incmid	-0.263	0.172	0.130	-0.398	0.119	0.001
inchigh	-0.125	0.173	0.470	-0.455	0.149	0.002

The regression results in Table 4.2 do comport largely, but not completely, with what we expect to find in ARS. The traditionally cited expectations hold somewhat in direction and significance, however with female respondents exhibiting a statistical association and more educated respondents showing a positive association in Wave 4. And while these results might be an impediment to any following analysis, and certainly deserves further scrutiny, the evidence is more cautionary and less fatal to the subsequent analyses. First, if the factor structure still comports with the notion that these measure common agreement, there is still some face validity for a response factor. Second, by limiting the context of ARS to American respondents only, the cultural idiosyncrasies commonly cited as larger drivers of ARS may emerge, confounding the idiosyncrasy of group response with what we expect to see in the

individual analysis. If there is a cultural component to response, which there undoubtedly is, individual indicators may be pulled in different directions, yet still be measuring ARS. By restricting the analysis to the American context, idiosyncrasies of American response will be expected. Finally, the character of the respondents may also be a matter of differential survey response rates - low or high (Ray 1987). The least engaged and least informed, who we would expect to exhibit this tendency most, are also least likely to respond in the first place, ironically exhibiting greater ARS if finally convinced to respond. This creates a situation whereby response style is conditional upon response, watering down the cognitive and engagement influences we might expect in the survey.

The analysis proceeds with the knowledge that what is measured here may not strictly adhere to the optimization hypothesis, but that the lion's share of the evidence points to ARS as a common factor for these questions. The heart of the model still detects a factor that we might identify as common agreement, common agreement that has validity for respondents in terms of finding it easier to agree, and does not weight itself to any particular question in a way that indicates a strong a priori substantive factor. And, by viewing questions as easier or more difficult to justify based on context, a particular country's response (the American sample here) profile might (and should) drive individual effects on ARS.

Acquiescent Response and Confidence in Political Institutions

With this model of ARS as a base in equation (4.5), and the validity considered, I proceed to inspect fifteen indicators of confidence in social/political institutions (recoded to focus on agreement: agreement=1, disagreement=0) as dependent variables (Table 4.3). Each variable, simultaneously, is added as a dependent variable - a contextual agreement around the institution which is predicted by the demographic variables in equation (4.6) and dichotomized

partisan identification variables (Democrat, Republican, with Independent and No Affiliation as the base category). This acknowledges that standard demographic characteristics and partisanship will also influence a respondent's likelihood to agree on a specific question.

Analytically, this adds several (14 in 2001-2002 Wave 4 and 15 in 2005-2007 Wave 5) regression equations to (5) which simultaneously estimates the measurement model and inspects how each confidence question is predicted by ARS through a binary logistic regression.

(4.7) $P_{ik}(X_{i=1}) = \Lambda[\beta_0 + \lambda\theta_j + \delta_1(\theta_j * dem) + \delta_2(\theta_j * rep)]$

This set of estimates provides an array of effect estimates for each question and I restrict focus to three of these: the effect of ARS (general tendency to agree stylistically: λ) on each institution's confidence, the effect of ARS and identifying as a Democrat (δ_1), and the effect of ARS and identifying as a Republican (δ_2).

The expectations for λ, δ_1, and δ_2 are first, minimally, that a portion of institutional confidence is determined by an automatic positive response or acquiescence – ARS (λ). Second, that δ_1 and δ_2 – Democratic and Republican ARS – may either add or subtract from this automatic response which is an open question. We might expect several outcomes: (1) we may view partisanship as a kneejerk categorization in and of itself, such that those who might choose a partisan label also will tend to have confidence in government simply by agreeing (All partisan coefficients positive), (2) we might expect partisan identification to affect our reactions negatively to an opposition party in power, yet acquiesce to the respondent's own party (Partisan coefficients vary according to the presumptive partisan nature of the object), or (3) we may suspect that partisanship informs, minimally, all respondents in a way that negates the

influence of any possible kneejerk reaction regardless of the partisan loading or association of the object.

The Figures 4.2-4.4 show the coefficients and 95 percent confidence intervals of interest which are detailed in Tables 4.3.a and 4.3.b:

Figure 4.2.a. The Effects of ARS on Institutional Confidence (λ): Wave 4 (2001-2002)

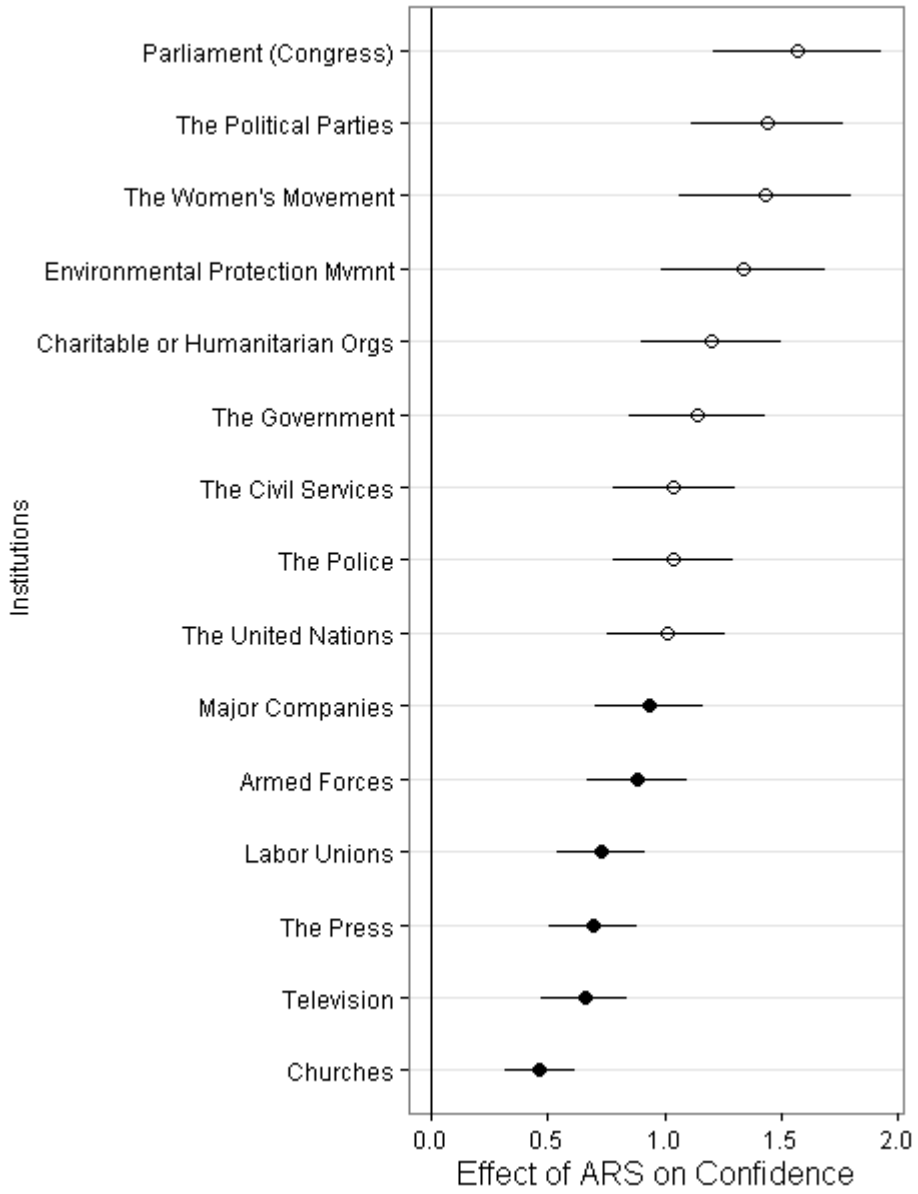
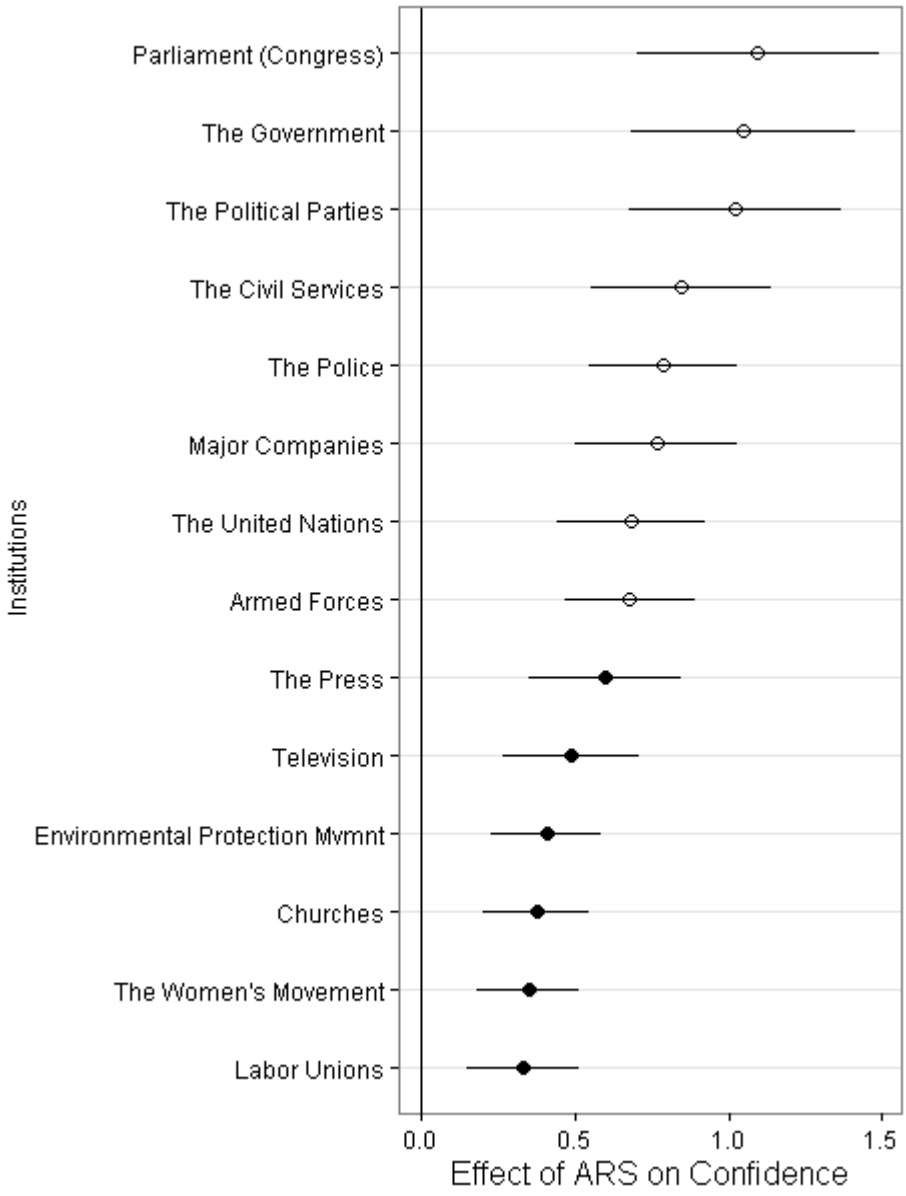


Figure 4.2.b. The Effects of ARS on Institutional Confidence (λ): Wave 5 (2005-2007)



Figures 4.2.a and 4.2.b confirm the first set of expectations, that the coefficients in both Waves 4 and Wave 5 are all positive. Again, this is both a validation of the modeling effort and a result. First, it validates that the confidence battery of questions are eliciting some sort of acquiescence in the slightly altered question format: those who may be in doubt, are less engaged with the question, or have a predisposition to agree will tend to have some amount of

"automatic confidence." That is, what is happening in these questions is not meaningful, but indicative of nothing more than ARS. Yet, ARS does have some variance in these models. Strikingly, the most political objects – Political Parties, Congress, and The Government are all among the top questions that elicit the most ARS (the top 3 in Wave 5). The magnitude of the coefficients are notable. The effect of ARS on institutional confidence is greater than 1 for all of these three political objects, indicating that the propensity to be confident about these politically charged institutions varies a great deal with automatic acquiescence, such that, as ARS increases, the confidence in particularly institutions also increases. Again, as modeled, this result occurs for no other reason that as a reaction to the question structure itself.

Aside from the most partisan and politically charged objects, there is little other variation in these coefficients that might qualify as particularly interpretable. For example, objects such as Labor Unions or Major Companies which comprise part of partisan coalitions are less influenced by ARS directly than the most associated with government. Churches (possibly conservative), The Press (possibly liberal), Environmental Groups (liberal), and Women's Movement (liberal) also do not exhibit consistent patterns of ARS that set them apart.

To the second expectation, *we might expect the effect of ARS to substantively differ with respect to the partisan nature of the object*, and the evidence is more nuanced. The expected direction of partisanship is uncertain, such that when partisan cues for disagreement are strongest, respondents may be less likely to acquiesce or that partisanship might identify a particularly strong mode of agreement. Also, when party cues offer a rationale for non-confidence, we might expect these cues to decrease the effect of overall tendency to agree by providing a negative rationale for confidence in an institution. I evaluate these sets of coefficients in turn for Democrats and Republicans.

Figure 4.3.a. The Effects of ARS on Institutional Confidence for Democrats (δ_1): Wave 4 (2001-2002)

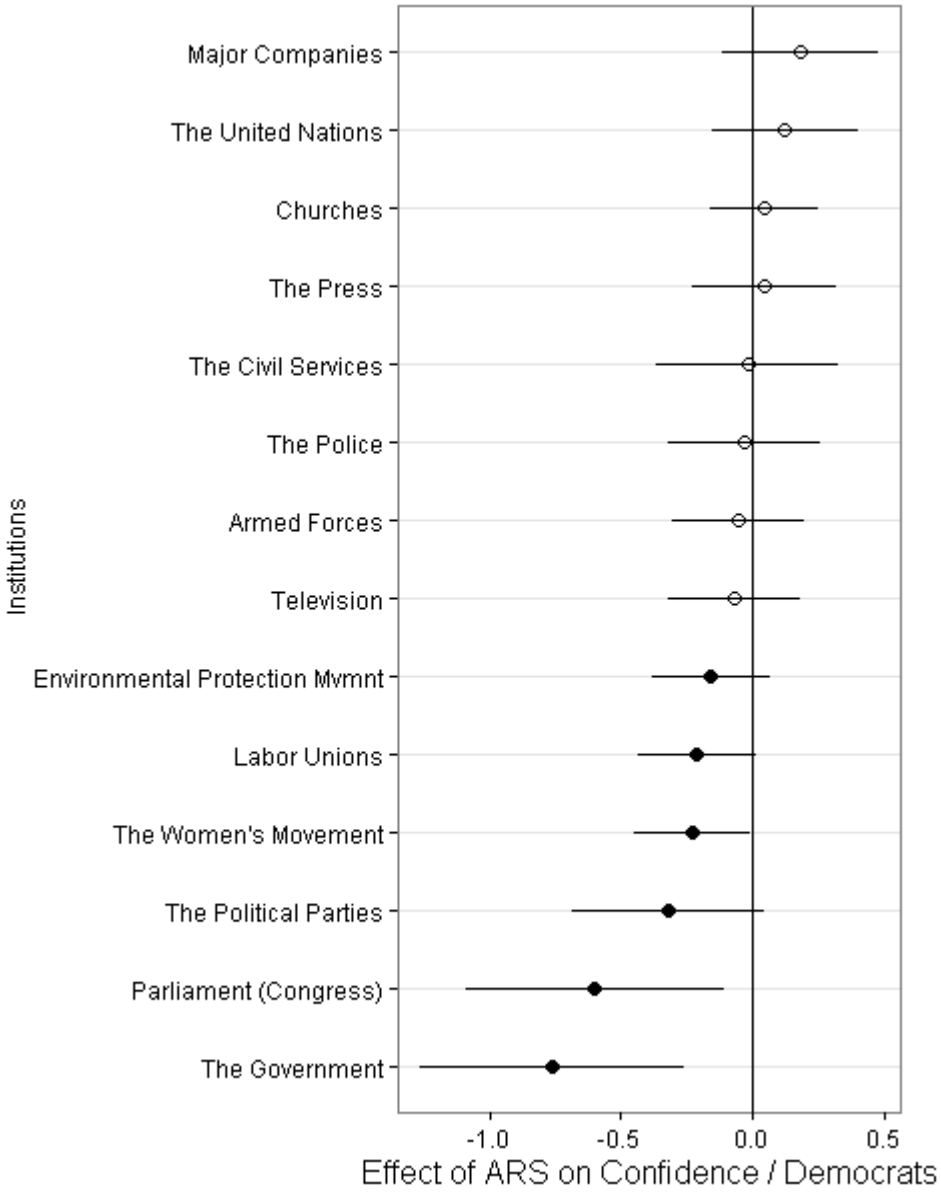
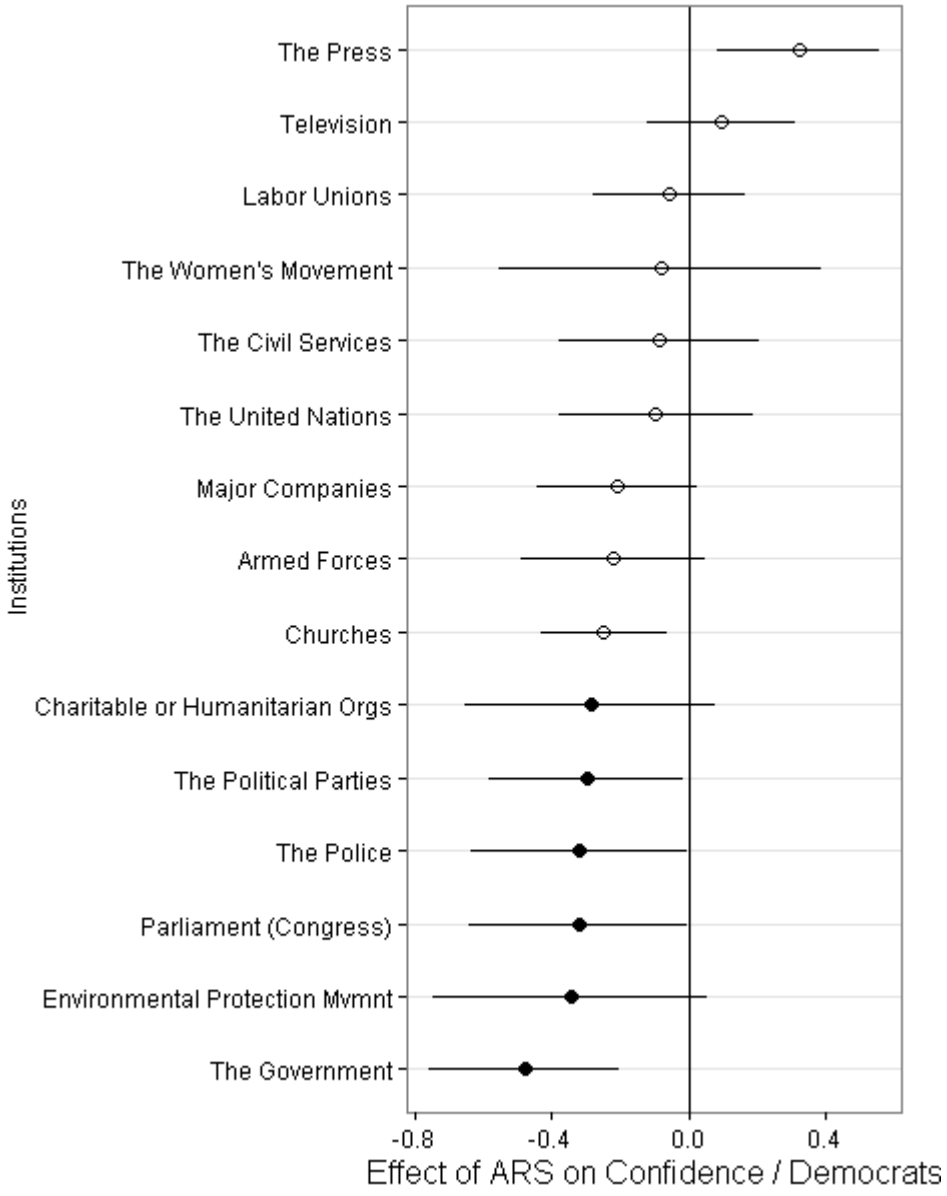


Figure 4.3.b. The Effects of ARS on Institutional Confidence for Democrats (δ_1): Wave 5 (2005-2007)



For Democrats, most coefficients are effectively zero, yet the exceptions are illuminating. There is a downward skew to the coefficients which are punctuates notably by Congress and The Government. Along with Political Parties, these are the same coefficients that are most influential in terms of ARS from Figures 4.2.a and 4.2.b, yet are now negatively related. For Democrats, this means that the same institutional confidence questions that are

very highly predictive in terms of overall ARS are less predictive when combined with affiliation with the Democratic Party. Substantively, direct questions about confidence in the Government or Congress may be more automatic responses overall, but affiliation as a Democrat actually negates this difference. The implication here is that being a Democrat actually informs people's opinions enough that they become just as likely to acquiesce to direct government questions as they would other questions. The Democrat effect δ_1 does not negate the full effect of ARS on institutional confidence, yet at least for Democrats, the effect negates what is a noticeable difference overall. All told, the information supplied by being a Democrat is enough to help a respondent evaluate direct government institutions in a less automatically favorable way.

These coefficients resolve the empirical questions about whether this one party inspection indicates an increase or decrease in acquiescence on more direct questions about government. For the most politically charged items, ARS decreases for Democrats. However, this is not technically true for many of the other questions. The interaction term says little to nothing for the other questions which include those institutions whose membership has historic partisan and governmental influence. In fact, the ordering of the other factors does not appear to have a particularly partisan ordering at all. What might constitute a slight "liberal acquiescent bias" in Wave 5 does not appear to hold in Wave 4. If there were a bias here by partisan affiliations, we would expect that more liberal Democrats would have more automatic, unthinking confidence in more liberal institutions, yet this kind of phenomenon is not corroborated by both years. Only the simple conclusion holds: being a Democrat appears to negate the difference in ARS in the most direct questions of confidence in government. The

only reason that Democrats appear to overcome ARS is due to increased awareness for the most direct items on governance independent of any partisan leaning.

Yet, does this happen for Republicans as well?

Figure 4.4.a. The Effects of ARS on Institutional Confidence for Republicans (δ_2): Wave 4 (2001-2002)

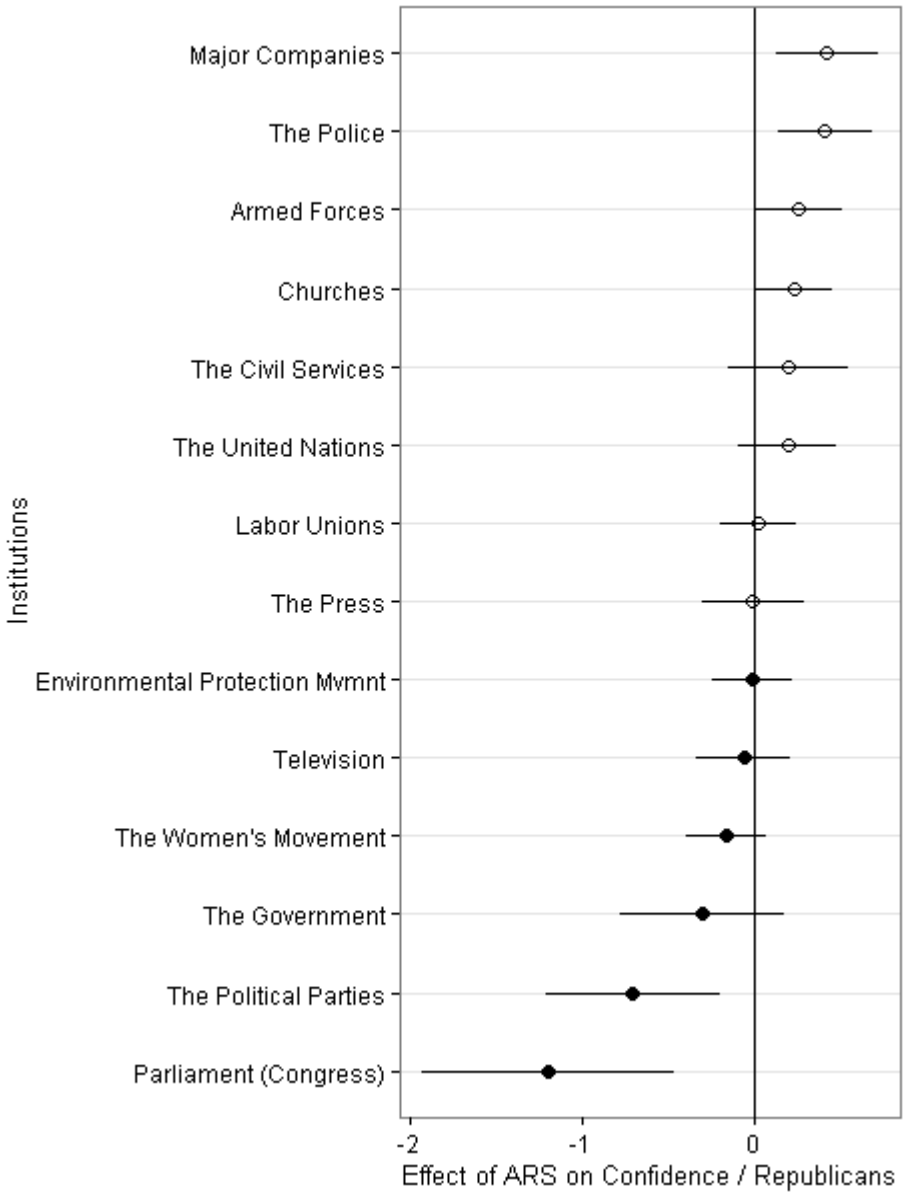


Figure 4.4.b. The Effects of ARS on Institutional Confidence for Republicans (δ_2): Wave 5 (2005-2007)

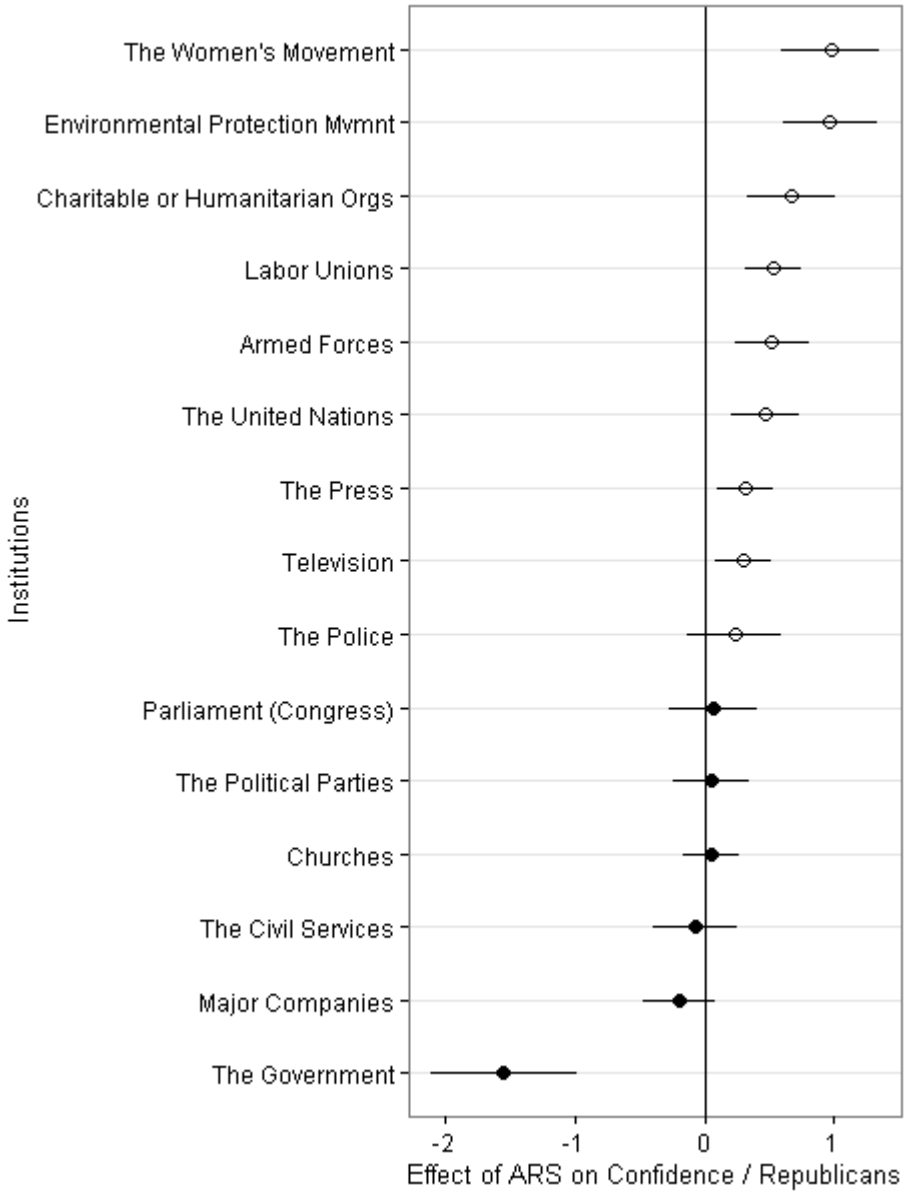


Figure 4.4.a-b demonstrate that the effect of partisanship is less direct, though questions of the Government (Wave 5), Political Parties and Congress (Wave 4), again exhibit a negative relationship. As with Democrats, in these areas, being a Republican appears to negate some of the effect of overall ARS for some of the most direct governmental confidence questions. For Republicans and the other questions as well, there is no discernible relationship between party

effect and ARS strength. When Republicans tend to have more confidence in institutions, they appear just as likely to exhibit a tendency to agree overall. The only difference is a mild tendency for Republicans to acquiesce compared to Democrats. Democrats, on the other hand, demonstrate a more consistent pattern, such that when partisanship appears to influence confidence more strongly, respondents are less likely to acquiesce. Republicans exhibit several more questions to which they would more likely agree automatically (ARS).

As far as the partisan ordering of acquiescence, there is no clear pattern. Like Democrats, if we expect there to be areas in which Republicans would acquiesce more due to their affiliation, it would be in objects that lean more conservatively, and that they might exhibit ARS less for objects with a liberal bent. Again, this might be the case for Wave 5, but the ordering is not corroborated by a similar ordering in Wave 4. The ordering and magnitude is not interpretable as showing a particular partisan bias. What it does show is a mild tendency for Republicans to evaluate groups and institutions both (a) in a more acquiescent way than Democrats overall, but (b) still less acquiescent in terms of direct questions of institutional confidence.

Put together, Republicans and Democrats show more similar patterns in how partisanship affects the way they may or may not truncate their evaluations of political institutions: the most directly political and partisan institutions are, the more partisanship (regardless of Republican or Democrat) appear to influence their confidence in the most direct questions of government, namely The Government, Political Parties, and Congress.

Discussion/Conclusion

The preceding analysis constructs a measure of acquiescent response and examines the properties of extreme response with respect to respondents' confidence in major societal

institutions. Several results here may enhance our understanding of how acquiescent response style behaves. While the measurement model performs adequately in identifying a common construct we might call ARS in both Waves 4 and 5, the nature of the ARS measure does not comport exactly with what we strictly would expect from optimization theory. While different in its characteristics from ERS, the different character of ARS is not entirely unexpected from previous research (Meisenberg and Williams 2008, Mondak and Anderson 2008). It may also carry some cultural components specific to the American sample. Response is more gendered, and does not strictly comport with education in how it behaves, even though it does comport in its effects by income. This said, the validity of the measure is bolstered by how a common ARS affects responses to government institutions generally (λ in Figures 2a and 2b and Table 3). Yet it seems plausible that some contamination may still exist in these measures that the model seeks to eliminate by design from equations (3) and (4). As such, it may be safe to assume that much of what is present is ARS, while some lingering connections may still exist with one or more nuisance dimensions. At the very least, the modeling effort improves on mere "counts" as implemented in previous analyses, and we might glean some very basic conclusions from it going forward.

One possible conclusion bolsters the theory that generalizes ARS as *situational justification*. Given the heterogeneity of ARS results in previous literature showing ARS as context specific, we might use a strong intuitive and theoretical understanding that positive response bias matters, and that it matters for different questions differently. When issues are at stake, causal stories matter and reasons matter (Bromley 2006, Stone 1989). When a statement is made, we may be hard-wired to accept a statement unless we have sufficient reason to reject it. The accessibility and prevalence of this reasoning will vary with cultural context, individual

knowledge, or be induced by content. This study emphasizes the contextual influence, and at least provisionally, there is some evidence that individuals vary substantially in their use of agreement by content - increasing in agreement for the most partisan political objects, yet decreasing in agreement when a partisan affiliation is present. These results suggest we must be open to the fact that true opinion about confidence in government may, and likely is, colored by an arguably false effect of agreement independent of content. There is also some indication that partisanship and the degree of partisan effect on confidence is a hallmark on increased awareness or willingness to respond – especially for the most obviously political institutions. For Democrats and Republicans, this may also be asymmetric in character, indicating the being a Democrat leads to less ARS than Republicans – though this conclusion is somewhat weak.

Even if there is some situational reasoning here, it is important to note what situational reasoning specifically. Notably, the more complex reasoning of knowing where one stands in the matrix of political associations does not appear to hold. Republicans and Democrats are not predisposed, in these models, to abandon ARS along lines that are partisan. Questions about the opponents' coalitions do not appear to be appropriately ordered in their associations. Only the most direct questions about government and partisanship appear to be affected in a way that might lead us to believe that partisanship helps reduce confidence in government by circumventing ARS.

This said, the precision of ARS measurement cannot tell us here whether variance in measurement is not substantially affected by the idiosyncratic nature of any survey, residual confounding effects, or real differential agreement. However, if we are generally aware that ARS bias exists, we do ourselves a disservice not to explore it scientifically, even if the area can be intractable. By looking at situational response style, as is likely the case with ARS, the

uncomfortable proposition remains that bias is not a clean function of the respondent or culture, but an interaction between the two, an interaction that is more aptly described theoretically by cognitive patterns in response to substantive content.

The measurement and analytical approach here is an improvement to existing practice, but by no means the only approach. The results are descriptive, which may call for a more inferential measurement - likely a question level "random effects" approach through a more flexible modeling strategy using MCMC. Additionally it is undoubtedly true that more thorough analysis of the sensitivity of results to measurement specification would bolster of disconfirm the nature of these results. Initial substitution of variables and creating unmodeled associations within the measurement model did not produce substantial differences, yet a more structured examination might be informative. And, without question, a richer set of covariates would be beneficial, yet not technically feasible. Data sets do not often exhibit the special properties present in the World Values Survey that provides a unique way of gaining traction on a notoriously intractable area of measurement.

If there is a tendency to agree, and if there is meaningful situational variance in this tendency, the substantive conclusions of analysis of survey results will be biased, and perhaps substantially so. Previous research and survey design texts attest to the first and this study suggests that we may reasonably expect this to be true. Acquiescent response, while not the information we asked for, it is the information we have. We might cast the information aside, but if we approach the survey response as containing meaningful information, this study also suggests that we would be remiss to ignore what ARS implies substantively - a potential window into political reasoning and rationale.

Where the bias exists is now more apparent. For partisanship, the results here reinforce the idea that partisanship is indeed the hallmark of a more developed approach to politics in the mass public. Where partisanship exists, bias in confidence is lessened. However put another way, where partisanship does not exist, there is a pronounced bias toward acquiescence in confidence in the most overtly political institutions. Partisan independents, perhaps counterintuitively, are most susceptible to ARS bias in their response to more politicized phenomena, ostensibly due to a lack of engagement or knowledge. It is the most politicized of institutions that will contain the most bias, and within this bias, it is the independent responses to confidence that will be affected most. It may be wise to remember in asking questions that prompt favorability in feeling or ideas in particular that, where ARS is at play, more exposure may reduce favorability in ways that have no bearing to any true change in disposition. In terms of stock "candidate favorability" questions, to the extent this ARS tendency applies: a favorable candidate is just an unfavorable candidate you haven't met, and if you're a partisan, you are just more likely to have met her.

These conclusions must also be recognized as valid in a limited way, and for a limited subset of respondents. The difference between knowing or caring enough to engage minimally in a survey is an extremely low threshold of participation in politics. While an important threshold to attain, it may not say much about the knowledge or engagement of subjects beyond a very rudimentary level. The difference between not giving a valid response at all to giving a slightly more valid response is a small, stylized piece of information which may or may not yield more generalizable results about which partisan affiliations or what strong partisanship may yield. Being an informed and engaged citizen, and determining who is or is not behaving in the correct normative way, requires a different analysis. On the other hand, it is the case that

we are often concerned with apathy or disengagement for the very (extremely) minimally engaged in politics. To this end, partisanship appears to play a role for those at the very cusp of caring. Partisanship, in its most basic function, appears to be one important way in which citizens may increase awareness from the most minimally aware and engaged, to a threshold by which they can give some structured thought to government and politics.

Table 4.3.a. Coefficients from Simultaneous Estimation of Confidence: Wave 4

QID	Confidence Question	Effect of ARS			Democrat			Republican		
		Beta	SE	p	Beta	SE	p	Beta	SE	p
e069_01	Churches	0.374	0.087	0.000	0.045	0.103	0.660	0.234	0.109	0.032
e069_02	Armed Forces	0.677	0.106	0.000	-0.059	0.125	0.640	0.265	0.126	0.036
e069_04	The Press	0.600	0.124	0.000	0.041	0.136	0.760	-0.005	0.150	0.980
e069_05	Labor Unions	0.331	0.090	0.000	-0.212	0.111	0.058	0.024	0.112	0.830
e069_06	The Police	0.789	0.121	0.000	-0.032	0.143	0.820	0.413	0.138	0.003
e069_07	Parliament (Congress)	1.096	0.197	0.000	-0.601	0.245	0.014	-1.201	0.366	0.001
e069_08	The Civil Services	0.848	0.147	0.000	-0.020	0.172	0.910	0.201	0.175	0.250
e069_10	Television	0.485	0.111	0.000	-0.069	0.126	0.580	-0.060	0.138	0.660
e069_11	The Government	1.046	0.182	0.000	-0.760	0.250	0.002	-0.302	0.241	0.210
e069_12	The Political Parties	1.022	0.174	0.000	-0.322	0.182	0.076	-0.706	0.253	0.005
e069_13	Major Companies	0.765	0.131	0.000	0.178	0.148	0.230	0.427	0.149	0.004
e069_14	Environmental Protection Mvmnt	0.406	0.089	0.000	-0.161	0.112	0.150	-0.008	0.117	0.950
e069_15	The Women's Movement	0.348	0.084	0.000	-0.230	0.109	0.035	-0.162	0.117	0.170
e069_20	The United Nations	0.683	0.121	0.000	0.121	0.139	0.380	0.195	0.143	0.170
e069_40	Charitable or Humanitarian Orgs	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

***Text (from the OECD version): I am going to name a number of organizations. For each one, could you tell me how much confidence you have in them: is it a great deal of confidence, quite a lot of confidence, not very much confidence, or none at all?**

Table 4.3.b. Coefficients from Simultaneous Estimation of Confidence: Wave 5

QID	Confidence Question	Effect of ARS			Democrat			Republican		
		Beta	SE	p	Beta	SE	p	Beta	SE	p
e069_01	Churches	0.463	0.075	0.000	-0.248	0.092	0.007	-0.053	0.110	0.630
e069_02	Armed Forces	0.880	0.107	0.000	-0.222	0.134	0.098	-0.524	0.144	0.000
e069_04	The Press	0.694	0.096	0.000	0.320	0.118	0.007	-0.317	0.107	0.003
e069_05	Labor Unions	0.729	0.095	0.000	-0.057	0.112	0.610	-0.537	0.107	0.000
e069_06	The Police	1.035	0.128	0.000	-0.319	0.157	0.043	-0.232	0.181	0.200
e069_07	Parliament (Congress)	1.567	0.181	0.000	-0.322	0.160	0.045	-0.066	0.171	0.700
e069_08	The Civil Services	1.041	0.130	0.000	-0.086	0.147	0.560	0.073	0.164	0.660
e069_10	Television	0.659	0.092	0.000	0.097	0.108	0.370	-0.301	0.105	0.004
e069_11	The Government	1.138	0.146	0.000	-0.480	0.139	0.001	1.546	0.280	0.000
e069_12	The Political Parties	1.438	0.162	0.000	-0.298	0.142	0.036	-0.053	0.148	0.720
e069_13	Major Companies	0.934	0.115	0.000	-0.209	0.118	0.077	0.190	0.139	0.170
e069_14	Environmental Protection Mvmnt	1.337	0.174	0.000	-0.344	0.200	0.085	-0.969	0.180	0.000
e069_15	The Women's Movement	1.432	0.185	0.000	-0.083	0.234	0.720	-0.975	0.188	0.000
e069_20	The United Nations	1.008	0.125	0.000	-0.096	0.141	0.500	-0.468	0.132	0.000
e069_40	Charitable or Humanitarian Orgs	1.202	0.149	0.000	-0.288	0.182	0.110	-0.665	0.170	0.000

*Text (from the OECD version): I am going to name a number of organizations. For each one, could you tell me how much confidence you have in them: is it a great deal of confidence, quite a lot of confidence, not very much confidence, or none at all?

Appendix 4.1. Winnowing the World Values Questions to those which Exhibit Lower Interitem Correlations

Many of the items in the World Values Survey measure ARS, but were intended to be scaled together to measure a common factor (the Schwartz series in particular). Others were intended to measure competing values (the A-series: valuing family, politics, friends, etc.). A complete list of the 43 items that fit the profile of ARS, asking for agreement, is given below. Only 23 of these fit the crierion that the Pearson correlations in Wave 5 (2006) that interitem correlations be below a certain threshold. Here, the threshold is r=0.40. To lend a more systematic process of elimination, the most correlated items were eliminated first, subject to a reasonable number of items being left in the analysis (greater than 20). The iterative process led to the final 23 items in Wave 5:

a002 Important in life: Friends
a003 Important in life: Leisure time
a004 Important in life: Politics
a005 Important in life: Work 20
b001 Would give part of my income for the environment
b003 Gov't should reduce environmental pollution
c001 Men should have more right to a job than women
c002 Employers should give priority non-immigrants
c036 To develop talents you need to have a job
d018 Child needs a home with father and mother
d022 Marriage is an out-dated institution
d054 Goals in life has been to make parents proud
d055 Make effort to live up to what friends expect
d057 Being a housewife just as fulfilling
*d078 Men make better business executives than women
*d079 I seek to be myself rather than to follow others
f104 People with strong religious beliefs in pub office
f105 Religious leaders should not influence government
*g020 I see myself as member of my local community
*g021 I see myself as citizen of the [country] nation
*g023 I see myself as an autonomous individual
e115 Political system: Having experts make decisions
e117 Political system: Having a democratic political system

*-Available in Wave 5, but not in Wave 4

Appendix 4.2: Code in Latent Gold (Vermunt and Magdison 2005)

```
model
options
  algorithm
    tolerance=1e-008 emtolerance=0.01 emiterations=250 nriterations=50;
  startvalues
    seed=0 sets=10 tolerance=1e-005 iterations=50;
  bayes
    categorical=1 variances=1 latent=1 poisson=1;
  montecarlo
    seed=0 replicates=500 tolerance=1e-008;
  quadrature  nodes=10;
  missing  excludeall;
  output
    parameters=effect  betaopts=szw
 standarderrors probmeans=posterior
    profile bivariateresiduals;
variables
  dependent A002b nominal, A003b nominal, A004b nominal, A005b nominal,
    B001b nominal, B003b nominal, C001b nominal, C002b nominal, C036b nominal,
    D018b nominal, D022b nominal, D054b nominal, D055b nominal, D057b nominal,
    D078b nominal, D079b nominal, E115b nominal, E117b nominal, F104b nominal,
    F105b nominal, G020b nominal, G021b nominal, G023b nominal,
    E069_01b nominal,E069_02b nominal,E069_04b nominal,E069_05b nominal,E069_06b
nominal,
    E069_07b nominal,E069_08b nominal,E069_10b nominal,E069_11b nominal,E069_12b
nominal,
    E069_13b nominal,E069_14b nominal,E069_015b nominal,E069_20b nominal,E069_40b
nominal;
independent rep dem;
   latent
    Cluster nominal 1,
    CFactor1 continuous;
equations
  E069_01b <- 1 + CFactor1 + dem + rep + dem*CFactor1  + rep*CFactor1 ;
  E069_02b <- 1 + CFactor1 + dem + rep + dem*CFactor1  + rep*CFactor1 ;
  E069_04b <- 1 + CFactor1 + dem + rep + dem*CFactor1  + rep*CFactor1 ;
  E069_05b <- 1 + CFactor1 + dem + rep + dem*CFactor1  + rep*CFactor1 ;
  E069_06b <- 1 + CFactor1 + dem + rep + dem*CFactor1  + rep*CFactor1 ;
  E069_07b <- 1 + CFactor1 + dem + rep + dem*CFactor1  + rep*CFactor1 ;
  E069_08b <- 1 + CFactor1 + dem + rep + dem*CFactor1  + rep*CFactor1 ;
  E069_10b <- 1 + CFactor1 + dem + rep + dem*CFactor1  + rep*CFactor1 ;
  E069_11b <- 1 + CFactor1 + dem + rep + dem*CFactor1  + rep*CFactor1 ;
  E069_12b <- 1 + CFactor1 + dem + rep + dem*CFactor1  + rep*CFactor1 ;
  E069_13b <- 1 + CFactor1 + dem + rep + dem*CFactor1  + rep*CFactor1 ;
```

```
  E069_14b <- 1 + CFactor1 + dem + rep + dem*CFactor1  + rep*CFactor1 ;
  E069_15b <- 1 + CFactor1 + dem + rep + dem*CFactor1  + rep*CFactor1 ;
  E069_20b <- 1 + CFactor1 + dem + rep + dem*CFactor1  + rep*CFactor1 ;
  E069_40b <- 1 + CFactor1 + dem + rep + dem*CFactor1  + rep*CFactor1 ;
  Cluster <- 1;
  A002b <- 1 + Cluster + (b1)CFactor1;
  A003b <- 1 + Cluster + (b1)CFactor1;
  A004b <- 1 + Cluster + (b1)CFactor1;
  A005b <- 1 + Cluster + (b1)CFactor1;
  B001b <- 1 + Cluster + (b1)CFactor1;
  B003b <- 1 + Cluster + (b1)CFactor1;
  C001b <- 1 + Cluster + (b1)CFactor1;
  C002b <- 1 + Cluster + (b1)CFactor1;
  C036b <- 1 + Cluster + (b1)CFactor1;
  D018b <- 1 + Cluster + (b1)CFactor1;
  D022b <- 1 + Cluster + (b1)CFactor1;
  D054b <- 1 + Cluster + (b1)CFactor1;
  D055b <- 1 + Cluster + (b1)CFactor1;
  D057b <- 1 + Cluster + (b1)CFactor1;
  D078b <- 1 + Cluster + (b1)CFactor1;
  D079b <- 1 + Cluster + (b1)CFactor1;
  E115b <- 1 + Cluster + (b1)CFactor1;
  E117b <- 1 + Cluster + (b1)CFactor1;
  F104b <- 1 + Cluster + (b1)CFactor1;
  F105b <- 1 + Cluster + (b1)CFactor1;
  G020b <- 1 + Cluster + (b1)CFactor1;
  G021b <- 1 + Cluster + (b1)CFactor1;
  G023b <- 1 + Cluster + (b1)CFactor1;

b1[1,2]=0;
```

Chapter 5. Constraint, Moderation, and Civic Education Instruction

When we talk about a knowledgeable and informed electorate, it is often unclear what we mean. We often value "fact-based" knowledge, we may value an awareness of the fault lines of dialogue, or we may value deliberative and circumspect judgment. How are these (normatively positive) traits transferred in practice? These traits matter for both scholars and political practitioners alike. And as I discuss here, they are also a concern to educators and the public. Partisans make claims consistently about the opposition, that they do not understand facts, how the world works, or are unthinking consumers of partisan information at the expense of critical thought. In this chapter, I use response style (ERS) as another tool toward characterizing what we might call moderation or deliberation that follow from results in chapter 3. The outgrowth of this definition here is an introduction of what we might call critical thought.

I use the 1999 IEA (The International Association for the Evaluation of Educational Achievement) Civic Education Database of 8th graders to assess the impact of instructional practice on three outcomes we might tie to political learning and response. The first is civics knowledge using a standardized test of civics, the second is a derived measure of belief cohesion (constraint) using the person-fit method of an item response model, and the third is a content-free measure of moderation in response using a nominal item response model.

The results show that what we might call constraint and moderation do play important roles and act independently of general civics knowledge and ability with regard to instruction. Furthermore, I find that modes of instruction that are less drill, memorization, and fact-based, do in fact show greater differences in how this 8th grade sample moderates response. The implication is that instruction that is issue-based, prompts discussion, and encourages opinion formation serves to decrease automatic response with respect to issues, the kind of unexamined

response that we might normatively discourage in a well-functioning democratic process. This study also reinforces a long-standing political science finding that belief constraint and learning go hand in hand, a point that is often suppressed in political dialogue. On the other hand, instructional practices that encourage opinion formation and civic knowledge also positively relate to moderation of response. The, perhaps counterintuitive, yet consistent conclusion is that instruction, and opinion instruction in particular, may both foster *use of ideology* and serve to *mute extreme opinions that are non-substantive*.

This chapter constructs measures of and examines the effects of issue constraint and issue moderation by instructional practice. There is a developed literature on what it is to learn about ideology and politics in the mass public, and some educational research that touches on civics coursetaking and political outcomes, but none that intersect dimensions of political learning and ideological thought with that of instructional practice. This may be due to the difficulty of measuring such concepts as belief constraint or moderation, or a reticence to branch out to consider ideological thought and consideration as a normatively positive outcome.

I use a US sample of 8th graders in the 1999 IEA (The International Association for the Evaluation of Educational Achievement) Civics dataset to inventory how different modes of instruction might affect civics learning in a multi-faceted way, one that may inform both political science and secondary instruction in civics. In addition to what we might call "civics aptitude" (general aptitude with a civics-specific component) measured through a standardized test of civics, I introduce two measures that we are particularly concerned with as political scientists. The first is a measure of "constraint," how strongly belief in any one issue becomes tied to another. The second is response "moderation," how belief in issues generally follows a more

muted, arguably circumspect route independent of content. Importantly, I examine how these two introductory measures act, and act independently of general aptitude given classroom instructional practice.

There are some limitations to the IEA Civics data set, notably a lack of covariates or other important variables that might give purchase on strict causal inference, however several results comes through in this piece. Notably that "constraint" and especially "moderation," as measured, do play important *independent* roles by instructional practice. That is, the general modes of instruction we might expect to operate to secure these outcomes of a more informed electorate are more detectably efficacious in securing these outcomes. This implies that a curriculum and pedagogy that encourages engagement with political discourse is more successful at teaching students about prevailing ideological lines and encourage moderated response, and those that stress memorization, are textbook-driven, or are fact-based material are less successful. And, to the larger debate, political science may communicate to those engaged in political rhetoric about what actually happens when students are taught about politics and belief. While messages in the classroom, like all other political messages, might be persuasive in either or any direction, I am more concerned about what instruction actually does to nascent learners overall in civics, and in which learning areas.

To this point, the popular press may take note that, first, civics education serves to increase belief constraint and cohesion, not to undermine it. Furthermore, there is some limited evidence here that some modes of instruction do, in fact, relate to belief cohesion more than others. Second, as far as curriculum and pedagogy undermining belief, this paper finds that, as students are presented information about opinions and encouraged to make up their own minds, they will tend to moderate their opinions, all else equal. Without regard to direction, practices in

civic education increase knowledge of civics, foster belief consistency, *and* encourage moderation of belief with respect to substance.

Constraint and Moderation

This chapter and the following take a normative stand on the merits of political reasoning, one that has its roots in a subjective nature of what political reasoning *should* be. This is stated up front, but it is not strange to imply that certain traits are more desirable than others. I will note that normative arguments abound about how certain qualities of the electorate are desirable, and others not – characteristics like deliberation (Fishkin 1997) and learning (Althaus 1998) are, for the most part, imbued with a sacrosanct "goodness." I make such a judgment here that there is positive normative merit in having such consistent and moderated belief.

Toward consistency and "constraint," I use a traditional view of political thinking in the mass public – namely borrowing the fact that more educated citizens connect issues in the same domain using ideological links (in the tradition spawned by Converse 1964). That is, that the more people know about the structure and substance of the broader American public dialogue, the more regular their patterns of response will be: their views on one issue will be more predictive of responses on another issue.

Furthermore, there is some consensus in ideological research that issues in similar domains should be connected in practice. This is the development of the awareness of contemporary fault lines of opinion that fit into the bailiwick of ideological reasoning along two poles, an acknowledgement that divisions are bi-polar artifacts of partisan competition which is some product of political structure - single-member districting, first-past-the-post, winner take all elections (Sniderman and Jackman 2002). That is, on a cluster of issues, a person should fit along a continuum from right to left, roughly Republican to Democrat.

On the other hand, in the political (and perhaps controversial) sense, I treat *moderation* as a desirable trait. These next chapters' view of reasoning again stems from a viewpoint about issue linkages and political reasoning and axioms of response dating to Zaller's *Nature and Origins of Mass Opinion* (1992). I propose that, in a competitive and ambiguous messaging environment, a hallmark of critical thought is a natural moderation of response. These persuasive messages are not always to clarify, but many of the messages serve to obfuscate or cross-pressure (Berelson et. al. 1958) individuals with arguments that seek to dislodge them from their stances and persuade them to select a different side, a different candidate, or otherwise dampen support (leading to abstention) (Franklin 1991). To measure constraint and moderation, it is a prerequisite activity to develop a frame of reference with which to examine ideological reasoning. Competitive parties have evolved and punctuated their stances in pivotal elections, some of which change the nature of party competitive substance, creating some difficulty in identifying relatively stable areas of difference in discourse. In modern times, however, the most pronounced and consistent of these are what we might term as Economic and Cultural Ideology (Social Welfare and Behavioral Norms) (Trier and Hillygus 2009). Of these, the most enduring, and empirically sound dimension is that of Social Welfare: how government is involved in national political life to redistribute wealth, economic outcomes, and economic opportunity (Claggett and Shafer 2010). Social Welfare continues to exhibit independent impact dating from the New Deal era, and continues to be the most forceful rift between parties currently. This stability gives it several natural advantages, not the least of which is the virtue of predictable measurement properties going back over half a century in America.

More exposure and acceptance of alternate arguments will lead to moderate response, as well as deep sampling of consideration. This stems from the dynamics of political information

both in its ambiguous nature and by virtue of Zaller's RAS (receive-accept-sample) model of reaction to political stimuli. The RAS model of political issue response posits that individuals are exposed to information, like metaphorical black and white ping-pong balls of the mind. They are bombarded with messages of both black and white, receiving all kinds of political stimuli – ostensibly persuasive messages from different parties which may be right-leaning in substance or cue, left-leaning in substance or cue, an independent mix of both, or an independent argument aimed at injecting (priming) a conflicting dimension. At the first stage of message receipt, a citizen may choose to accept or reject a message (the metaphorical ping pong ball). If accepted, it is called upon for storage. If not, it is turned away. Subsequently, when asked about a certain topic – on a survey or conversation, they will "sample" a response from memory.

At the very least, being exposed and accepting of alternate arguments would serve to qualify response, an exposure I inspect in this paper. We might entertain that no issue is inherently linked to any other a priori, and alternate (competitive) links might be fostered by new messages. Ultimately, this recognizes the connection between, say, abortion preferences and school prayer as essentially unrelated except through social construct and received argument. These constructs are cued by opinion leaders and elites and important and then manifest themselves as empirical phenomena, but in the end there is, only by argument and political interest, a linkage between the two. Ostensibly we would expect there to be a distinct relationship here between prayer and abortion, simply based upon what we know about the American political environment – a traditional vein and constituency drive the conservative Republican party toward limiting abortion and liberalizing restriction on school prayer. Yet, one may also have a legitimate libertarian viewpoint that favors allowing school prayer and abortion

decisions to be made by an individual, a viewpoint based not upon religious socialization, but an alternate received understanding of personal freedom and society.

So, to an essential progression of political learning, we may take two tacks. The first is constraint: a positive political level of awareness of prevailing arguments in which a student may find practical purpose in situating herself. This means listening to the public arguments/ reasoning and choosing a place on a larger continuum of issue connections and "constraining" them according to existing structure. A student looks at "what is," accepts the terrain of issues, and subsequently looks to arguments within them in order to find their voice and be able to express their citizenship within that structure. The second follows, then, insofar as moderation is present, individuals will receive information and accept a wider array of viewpoints and examine them (take them in as it were) regardless of a standard of belief that is internally or externally dictated. Thus, when issue knowledge is assembled, constraint is developed. And when sampling is undertaken, the requisite theoretical result for a critical thinker will be to moderate response, all else equal. More different viewpoints will be internalized, greater substantive thought will result, and a greater sample gathered when called to respond.

A logical, and critically thinking, student of political dialogue, then, will moderate response. It is with this understanding that a defined or undefined moderation in social policy choice is a product of critical thought, and "examined" belief.

> "The art of explicating, analyzing, and assessing these 'arguments' and 'logic' is essential to leading an examined life" (Paul 1990, p. 66).

> "…a self-sufficient person is a liberated person...free from the unwarranted and undesirable control of unjustified beliefs" (Siegel 1988, p.58).

Having said this, the alternate must be considered. Namely what if a rejection of structure comes with non-moderation? It is partly this idea that can be both proposed and examined in the

context of learning in civic education, yet for a couple reasons, I largely dismiss this general tendency. If one were to favor guaranteeing a minimum income for citizens, but not jobs, there is certainly a logical line that supports it. However, it is not "the" line of association, and no *popular* (or politically effective) line of argument supports it.

Lest this preamble get too abstract: in sum, I argue that both issue/belief constraint and moderation are normatively positive outcomes of political learning, both (a) grounded in the merits of learning and adopting ideological self-placement based on prevailing discourse and (b) taking in various ideas and taking a larger sample of considerations respectively – a hallmark of critical thought. To understand political dialogue is to be able to both place your belief in it and qualify it.

Educational Practice and Political Outcomes

With these concepts in hand, I turn to educational practice and encounter a relatively blank slate. Educational research, while replete with other outcomes, has not particularly valued those that are expressedly political, and when they do, it is has started in the form of a generic civics coursetaking pattern on standard fact-based political knowledge (Langton and Jennings 1968, Niemi and Junn 1998), finding null and positive effects. A more circumspect list of outcomes is found in Gimpel, Lay, and Schuknecht (2003), including effects of coursetaking on political discussion and participation, and Hillygus (2005) effects on adult political engagement.

> "Literatures in education, public policy, and economics are replete with studies evaluating the effect of a given educational policy, practice, or context on a social, educational, or economic outcome of interest. Very few works, however, analyze the effects of educational policies, practices, or contexts on political outcomes" (Carlson 2012, p.13)

The IEA Civics dataset comes with several strengths toward evaluating instructional practice. First, this survey revolves around the sample of US 8th graders in 1999, to frame a

nationwide population of middle school aged children. The age group is an important point, where many students are gaining a nascent grip on issues in society, having developed in their learning and being exposed (sometimes initially) to the divisions and problems that are essential to social and ideological cleavages. The types of questions asked is another asset, containing a multiple choice battery on social welfare ideology, a civics standardized battery to assess knowledge, and a wide array of instructional practices reported. Of these instructional modes, I choose eleven (11) (Table 5.1) that are specifically targeted at teaching.

Table 5.1. Instructional Practices Examined

Instructional Practice
254.2 Students are Encouraged to Make up Their Own Minds*
255.3 Teacher Respects our Opinion*
256.4 Great Importance on Learning Facts and Dates
257.5 Feel Free to Express Opinions in Class*
258.6 Requires Students to Memorize Dates/Definitions
259.7 Discuss Political/Social Issues*
260.8 Teacher Presents Several Sides of an Issue*
261.9 Brings Up Current Political Events*
262.10 Memorizing Dates to Get a Good Grade
263.11 Teacher Lectures and Students Take Notes
264.12 Students Work on Material Through Textbook

*-Encourages opinion or an opinion-based practice

The rudimentary question from these eleven is fairly simple: I expect practices that teach and encourage opinion to affect opinion-based measures, "constraint" and "moderation." Do these practices succeed at teaching what is measured? Questions are put to 8th graders, who are prompted to respond whether civics instruction "certainly does not do this," "probably does not do this," "probably does do this," or "certainly does do this" which I code as either 1= "civics instruction contains this practice" or 0= "civics instruction does not contain this practice." The questions are chosen from a single battery in order to contrast practices that are encouraging of opinions, beliefs, or otherwise seek to instruct based on presenting belief systems. The remainder of practices does not preclude opinion, but are characteristic of a more fact-based, passive

instruction which include lecture, textbooks, and memorization. Therefore these do not necessarily translate into either belief cohesion or qualification.

Measuring Constraint and Moderation with Social Welfare Ideology

To measure constraint and moderation, it is a prerequisite activity to develop a frame of reference with which to examine ideological reasoning. Competitive parties have evolved and punctuated their stances in pivotal elections, some of which change the nature of party competitive substance, creating some difficulty in identifying relatively stable areas of difference in discourse. In modern times, however, the most pronounced and consistent of these are what we might term as Economic and Cultural Ideology (Social Welfare and Behavioral Norms) (Trier and Hillygus 2009). Of these, the most enduring, and empirically sound dimension is that of Social Welfare: how government is involved in national political life to redistribute wealth, economic outcomes, and economic opportunity (Claggett and Shafer 2010). Social Welfare continues to exhibit independent impact dating from the New Deal era, and continues to be the most forceful rift between parties currently. This stability gives it several natural advantages, not the least of which is the virtue of stable and predictable measurement properties going back over half a century in America. Consequently, I rely on a battery of nine (9) items that represent what we have come to know as this "social welfare" ideology, which serves to identify how much government should intervene to produce normatively beneficial outcomes in society.

Toward measuring this trait, I use the IEA civics battery of nine questions that ask about opinions on government intervention in society:

Should Government:

- GUARANTEE A JOB FOR EVERYONE?
- KEEP PRICES UNDER CONTROL?
- PROV. BASIC HEALTH CARE FOR EVERYONE?
- ADEQUATE STANDARD OF LIVING FOR OLD PEOPLE?

- INDUSTRIE WITH SUPPORT NEED TO GROW ?
- ADEQUATE STANDARD FOR UNEMPLOYED?
- REDUCE DIFFERENCES IN INCOME AND WEALTH?
- PROVIDE FREE BASIC EDUCATION FOR ALL?
- ENSURE EQUAL POLITICAL OPPOTUNITIES FOR MINORITIES?

These questions measure a latent trait that models common variance: guaranteed jobs, price control, healthcare, standard of living, industrial support, unemployment, income equality, education, and opportunities for the disadvantaged or minorities. Furthermore, the responses indicate the degree to which government should be involved in these areas (securing these outcomes).

- 0 don't know
- 1 definitely should not be
- 2 probably should not be
- 3 probably should be
- 4 definitely should be

One result of these modeling efforts are scales of social welfare ideology, N~(0,1) with liberals on the negative (left) end and conservatives, yet it is not these traits that are of direct interest, but the following measures developed as a byproduct of deriving them.

Toward measuring constraint, political science literature has not ventured too far past simple correlations (Aachen 1975, Ansolabehere 2008). To remedy this, I make use of an item response "person-fit" statistic from a two parameter logistic model (2PL) of social welfare ideology. Quite simply, this statistic looks at response patterns and calculates an individual's deviance from expectation given an array of item responses and a latent social welfare trait. The result is a measure of likelihood of a response pattern given a trait: how "traited" is an individual along a social welfare dimension?

We often look at how each "question" fits an overall construct: how factors load on a given latent variable. The person-fit statistic uses the latent variable to assess how strongly a

"person" adheres to a latent construct (Meijer and Sijtsma 2001, Drasgow et al 1985, Levine and Rubin 1979, Reise 1990). Here, the construct is social welfare ideology and the fit statistic is calculated for students, and placed on a continuous scale.

Constraint here is measured by fitting a two parameter logistic (2PL) item response model that estimates a latent trait of social welfare ideology and examining how well individuals fit the overall ideological model: not to evaluate the "question" as is normally done, but to evaluate how well the data fit "each person." In some traditional practical work with this model, the person-fit measure looks to see whether there is something important in an individual's response pattern that deviates from what is being tested (a measure of cheating, copying, or multidimensional traited behavior in psychological testing). This is an important jump from traditional correlation item analysis, but makes use of an estimate of how well our data fit the model for persons: person analysis. That is, if the response for item *i* fits with an overall measured trait, the response on item *i+1* will correlate with a certain probability. If the deviance of the predicted probabilities of these models with respect to a respondent's ideology θ_j is aberrant to an unacceptable, designed threshold, the person is subject to being flagged for "misfit."

The regression analogue of person-fit is roughly an individualized Pseudo-R-Squared measure, where an individual's pattern of responses is evaluated with respect to their latent ideology and their probability of responding positively to any specific question. If a person is evaluated as a "conservative," their likelihood of responding, for instance, that they "support guaranteed jobs and a standard of living" should be fairly low. Each item is evaluated as such, where a less than ideal ideological fit in the overall model (as expressed in log distances) indicates that an individual is weakly "traited" on social welfare. This yields a continuum of fit

with those who have tightly coupled responses yielding positive values and those who are loosely coupled on economics have negative values (Using "R" package ltm, "personfit" method.) (Meijer and Sijtsma 2001, Drasgow et al 1985, Levine and Rubin 1979, Reise 1990). Usually, we are concerned with how well questions relate to an underlying construct (item fit), yet this mode of analysis conveys how people view the set of questions: does it mean something to not have a clear view of how questions are "supposed" to relate?

For our purposes, a limited number of factors theoretically contribute toward an examinee's ideological score being spurious or otherwise less than meaningful: careless responding, creative responding, and random responding (Meijer 1996). Careless responding occurs when the examinee, answers certain items without regard to ideology because of haste or inattention. Some examinees with high ability engage in creative responding, where they obtain "incorrect" responses to certain easy items, because they interpret these items in a unique, creative manner.

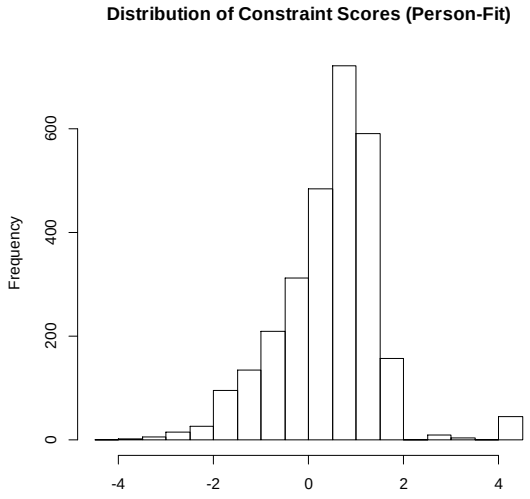
Finally, random responding refers to the situation where the examinee randomly chooses the multiple-choice option for each item on the test (Sijtsma et al 2005). Additionally there may be specific political science literature that bears upon why "traitedness" and specific, non-ideological response is present – the situation by which advanced politicians come to realize that the ideological associations in the mass public are somewhat malleable according to elite cues and national (world) events by which alignments are made and broken over many years. This line of reasoning has led to a proliferation of work in political science, but much of it acknowledges that this phenomenon is slow and evolutionary in nature. And, in any case, the current issue alignment in social welfare has not changed substantially in over half a century and

remains the dominant rift in politics. In this area, traitedness in ideological terms has been nearly unanimous.

Practically, the constraint is measured using the L_0 statistic (Levin and Rubin 1979) based on the likelihood of response for each item, making the L_z "standardized" statistic meaningful at 0, where the base for L_0 is the natural log of the likelihood of the response for each respondent summed over all items:

$$L(\theta_j) = \sum_i^n x_{ij} \ln\left(P_i(\theta_j)\right) + (1 - x_{ij}) \ln(1 - P_i(\theta_j)) \quad (5.1)$$

Figure 5.1. Frequency Plot of Social Welfare "Person Fit" - Constraint



To account for the spike of respondents at the top, which may be influential in an analysis for reasons unrelated to style, I follow convention for person fit and dichotomize the variable where index values less than 0 are coded to 0, otherwise to 1 as designation of persons who fit the model. "Don't know" responses are dealt with by penalizing answers in the proportion which they appear. The new variable of constraint follows a conventional distinction between those who are roughly using social welfare to guide response, and those who are not.

To gauge moderation, I obtain a measure of extreme response style, a tendency to select endpoints independent of the object of measure (Paulhus 1991), based upon work that aims at measuring a dimension of interest using a nominal response IRT model and inspecting residual endpoint selection patterns. This method has been used to gauge a respondent's propensity toward kneejerk or automatic extreme response (Bolt and Johnson 2009, Bolt and Newton 2011). In previous literature, this trait has been associated with different cultural approaches across countries, but in terms of a similar culture: lower income respondents, minority groups, and from this thesis, demonstrably, low political knowledge. This notably borrows from earlier chapters in the same intuitive way and with the similar model as follows. To do this, I rely on a multi-dimensional item response specification that "nets out" any social welfare ideology and provides a measure of an independent second dimension of extreme response (Bolt and Johnson 2009, Bolt and Newton 2011).[27] As with constraint, social welfare ideology is estimated and then "controlled for," resulting in a content-free measure of social welfare moderation (non-extremity) (Figure 5.2).

The content-free tendency toward extremity indicates, in survey behavior terms, a truncated and polarized thought process. In terms of Krugman's analysis of the Republican assault on critical thinking, the kneejerk response fits well. Do you hold strong opinions, or are those opinions unduly exaggerated for reasons that are purely idiosyncratic – all else equal?

As with constraint, social welfare ideology is estimated and then controlled for, resulting in a content-free measure of social welfare moderation (non-extremity) (Figure 5.2). Again, the Nominal Response model is used, but with parameters that conform to equation (3) where θ_{1j} is

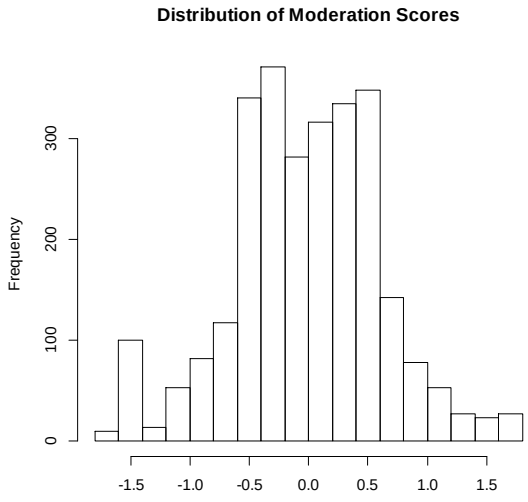
[27]The nominal response model is estimated using Latent Gold (Vermunt and Magdison 2004, 2005) The model also nets out "don't know" responses by modeling as a unique dimension, useful in evaluating constraint and sidestepping the critique that extreme (non-moderated) response may be overestimated (or incorrectly constructed) by ignoring "don't know" values (Böckenholt 2012).

a latent ideology, and θ_{2j} is a latent propensity toward extreme response explored in previous chapters, the negation of which is called "moderation."

Formally then, the specification is a familiar one, again as follows, where $P(X_{ik}|\theta_{1j}, \theta_{2j})$ is the probability of selecting a response category given a substantive latent variable, θ_{1j}, and an extreme response latent variable, θ_{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 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 ($a_{i(NR)1} = 0, a_{i(NR)2} = 0$), which provides the interpretation of ERS conditioned on selecting a point in the scale.

(5.2) $$P\left(X_{ik}\middle|\theta_{1j}, \theta_{2j}\right) = \frac{\exp\left(a_{\mathrm{ik1}}\theta_{1j} + a_{ik1}\theta_{2j} + c_{ik}\right)}{\sum_{\mathrm{i}}^{\mathrm{K}} \exp\left(a_{\mathrm{ik1}}\theta_{1j} + a_{\mathrm{ik1}}\theta_{2j} + c_{ik}\right)}$$

Figure 5.2. Frequency Plot of Social Welfare Moderation

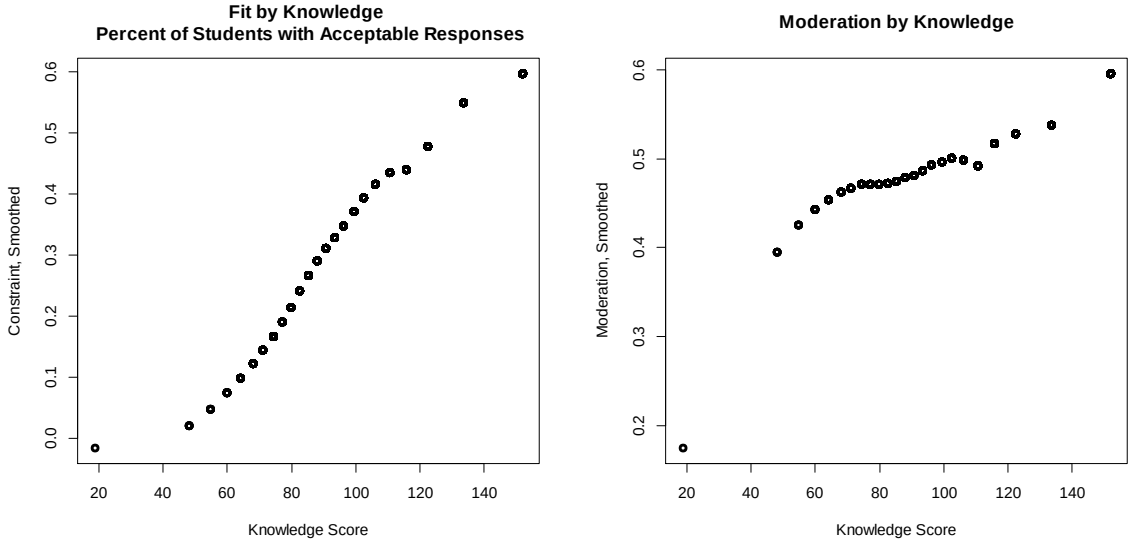


Again, the figure shows some inconsistency with an ideal normal distribution. On the far left are a cluster of extreme respondents with most respondents fitting in the middle, between -0.5 and 0.5. A casual glance at the distribution shows mild bimodality as well: there are fewer respondents in the exact middle, which implies very weakly that responses may gravitate to either "typically moderate" or "typically extreme." However, compared with the fit score, the measure does appear to center on zero adequately. As with fit, to create a more interpretable result, the normalized score was dichotomized (0=more extreme, 1=more moderate). The indicator may then represent "whether a student responded more or less moderately than average."

Results

The following puts together the three measures of knowledge, constraint, and moderation first to look at whether they behave as expected, and then adding instructional practice variables to inspect the effects they may have. To gain some understanding of the relationship between general knowledge and constraint/ moderation I start with some simple bivariate plots (Figure 5.3).

Figure 5.3. Fit (Constraint) and Extremity (Moderation) by General Civics Knowledge



*-loess plot of percent who exhibit "traited" behavior (fit) and those who exhibit moderate response by their MLE Civics scores [N(100,20)]

The plots run in the expected direction. That is, as knowledge increases, both "constraint" and moderation" increase. Constraint, in particular, increases dramatically and confirms the dominant notion that education and constraint are related. For constraint, for respondents who are very low on the civics knowledge scale (more than two standard deviations), there is virtually no evidence of fit. Five percent or fewer of these low information respondents conform to a reasonable fit to the social welfare model, while the highest abilitied students fit the social welfare model reasonably at higher than fifty percent. This is the kind of relationship political scientists have come to expect, that the more educated in the electorate will connect their responses along a left-right continuum more closely, and the less educated will respond more idiosyncratically.

Moderation, however, rises more modestly overall. For the least knowledgeable, a few respondents are quite extreme, but the two standard deviation range (~60 to 140 on the civic

knowledge scale) stays from 45 percent moderate at the bottom to just over 50 percent moderate at the top of the range. This restricted movement implies some association, but not stark. Given the theoretical expectations of this paper, the movement is consistent with a broader exposure and encouragement of opinion given the RAS model: more knowledgeable students will tend toward moderation.

What does not come through necessarily in these plots is a high dispersion of points. While the trends go in the correct direction and exhibit linear significance, the amount of variance explained is quite low, some of this may be attributable to measurement error in these indices. On one hand, this might be a cause for concern, yet it might also be evidence that these measures are not entirely capturing the same thing, namely general ability of the student.

Next, I present a table of effects (Table 5.2) detailing instructional practice and their linear effects on the three knowledge areas of question, using OLS for knowledge and linear probability for constraint and moderation. On Civics Knowledge, every practice except "bring up current events" and "feel free to express opinions" performed in class had some significant effect and none exceeding a half of a standard deviation. One conclusion is that most instructional practices have an effect on civics knowledge. This is a plausible notion on one hand – instruction works. Yet, unsatisfyingly perhaps, different modes of instruction, from the more rote to the more opinion-based appear to be similarly effective. The difficulty with the idea that nearly any instruction is efficacious is spurious correlation, such that there is an omitted influence that selects students into coursework that follows certain instructional practice, driving results. With this caveat set aside, the hypothesized (*) practices do not seem to act differentially in bolstering civics performance. In fact, "Learning Facts and Dates" showed a very much higher than anticipated difference in civics knowledge.

Table 5.2. Instructional Practice Differences by General Civics Knowledge, Constraint, and Moderation

Instructional Practice	Effect by Knowledge (SE)	Effect by Constraint (SE)	Effect by Moderation (SE)
254.2 Students are Encouraged to Make up Their Own Minds*	8.48 (1.16)	.183 (.027)	.082 (.028)
255.3 Teacher Respects our Opinion*	6.23 (1.03)	.099 (.024)	.072 (.025)
256.4 Great Importance on Learning Facts and Dates	9.14 (1.13)	.143 (.026)	.060 (.027)
257.5 Feel Free to Express Opinions in Class*	1.89 (1.06)	.080 (.024)	.060 (.025)
258.6 Requires Students to Memorize Dates/Definitions	8.01 (1.10)	.124 (.026)	.035 (.027)
259.7 Discuss Political/Social Issues*	4.94 (0.95)	.093 (.022)	.040 (.023)
260.8 Teacher Presents Several Sides of an Issue*	6.21 (1.07)	.132 (.025)	.077 (.025)
261.9 Brings Up Current Political Events	0.29 (0.88)	.054 (.020)	.018 (.021)
262.10 Memorizing Dates to Get a Good Grade	1.90 (0.94)	.048 (.022)	.014 (.022)
263.11 Teacher Lectures and Students Take Notes	3.42 (1.07)	.050 (.025)	.019 (.026)
264.12 Students Work on Material Through Textbook	6.82 (1.25)	.129 (.029)	-.005 (.030)

*-Encourages opinion or an opinion-based practice

By contrast, constraint performed more closely to expectations, though clearly falling short in some areas. Teaching and encouraging opinion should work more closely to secure outcomes of constraint, by theory. Yet emphasizing facts and dates (+14.3 percent), textbook work (+12.9 percent), and memorization (12.4 percent, 4.8 percent) all had higher than anticipated effects for areas that should not necessarily influence opinion formation. Perhaps it is the case that constraint behaves like general civics knowledge, where a "rising tide lifts all boats." It is helpful, in this light, to inspect where the effect of constraint differs from general civics knowledge. Several noteworthy observations come trough when looking at deviations. Strikingly, those who "feel free to express their opinions in class" gained in civics knowledge only slightly, but were 8.0 percentage points more likely to constrain their responses adequately. Similarly, when students are "encouraged to make up their own minds," civics knowledge is associated, yet constraint is very highly associated (18.3 percentage points higher). This implies that constraint may respond to different instructional modes, but not in the way this paper anticipates – perhaps behaving like civics knowledge in some ways, and more like opinion in others.

More strikingly, moderation appears to more closely conform to opinion encouragement and instruction, with or without comparing effects to general civics knowledge. These effects come out in opinion expression (7.2 percentage points greater in moderation) and presenting both sides of an issue (7.7 percent), and encouraging students to make up their own minds (8.2 percent). The size of the effects are more muted, but their significance overall and their unique ordering of instructional effects do imply

Because there is some relationship between knowledge and constraint/moderation (both in the theory I've outlined and given associations found in literature), it is possible that the effects found in instruction are merely byproducts of general civics knowledge and not indicative of real effect. The last question looks at the association of constraint and moderation independent of general civics knowledge. Another reason for doing so is to control for any effects that "tracking" different students may have on educational experience. One critique of an uncontrolled result is that higher or lower ability students are given different types of instruction. Table 5.3 presents these effects.

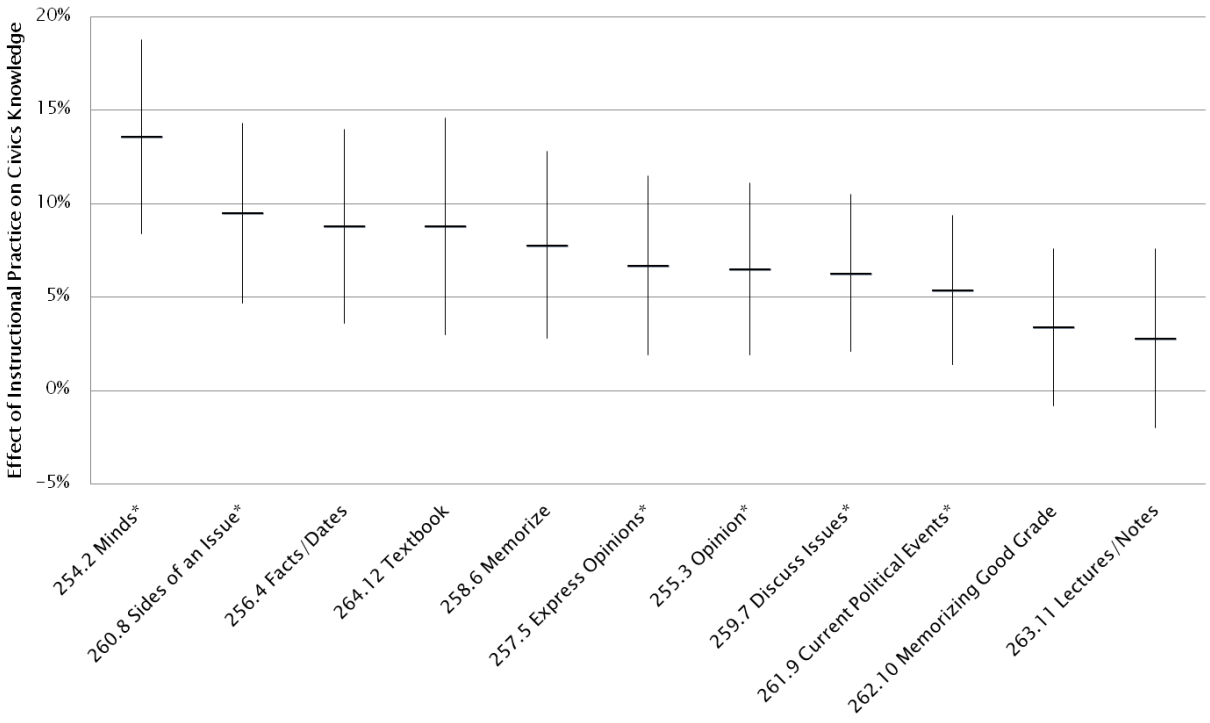
Table 5.3. Instructional Practice Differences by Constraint and Moderation, Controlling for General Civics Knowledge

Instructional Practice	Effect by Constraint (SE)	Effect by Moderation (SE)
254.2 Students are encouraged to make up their own minds	.136 (.026)	.073 (.028)
255.3 Teacher respects our opinion	.065 (.023)	.062 (.025)
256.4 Great importance on learning facts and dates	.088 (.026)	.047 (.028)
257.5 Feel free to express opinions in class	.067 (.024)	.055 (.025)
258.6 Requires students to memorize dates/definitions	.078 (.025)	.024 (.027)
259.7 Discuss Political/Social Issues	.063 (.021)	.033 (.023)
260.8 Teacher Presents Several Sides of an Issue	.095 (.024)	.069 (.026)
261.9 Brings Up Current Political Events	.054 (.020)	.017 (.021)
262.10 Memorizing Dates to Get a Good Grade	.034 (.021)	.012 (.022)
263.11 Teacher Lectures and Students Take Notes	.028 (.024)	.013 (.026)
264.12 Students Work on Material Through Textbook	.088 (.029)	-.017 (.030)

*-Encourages opinion or an opinion-based practice

The results from Table 5.3 confirm that the effect of constraint and moderation are still present, even when controlling for general civics knowledge. They are generally not as strong – perhaps implying that some of the effect of constraint and moderation work through general knowledge – yet the effects are still positive and significant in most cases. Figures 5.4 and 5.5 present these independent associations' point estimates by size of the association with 95 percent confidence intervals.

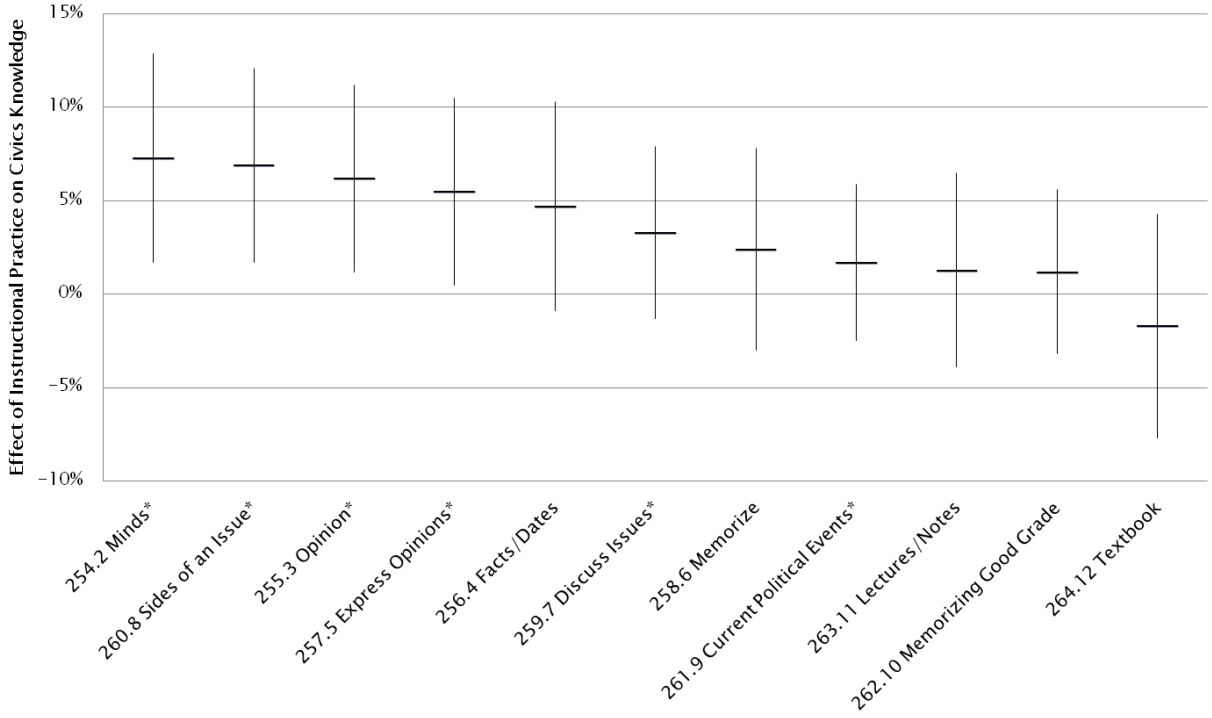
Figure 5.4. Constraint Net General Civics Knowledge (Effect of Instructional Practice on Percent of Students Fitting Social Welfare Model)



Arraying the results in this way helps demonstrate how different instructional practices may work to secure outcomes. For constraint, the two most influential practices (encouraging students to make up their own minds in particular) and the bottom two (insignificantly positive, showing below 5 percentage point difference) practices conform to the notion that instruction that includes opinion as a component relates to the cohesion of social welfare responses. The

intervening modes of instruction, however, do not follow suit in a clean way: facts and dates, textbook work, and memorization all show slightly more difference than their more opinion-based counterparts.

Figure 5.5. Moderation Net General Civics Knowledge (Effect of Instructional Practice on Percent of Students Who Generally Moderate Their Response)



Moderation is very different. By ordering the effects from largest to smallest (Figure 5.5), it is clear that instructional practices that encourage belief expression or examine opinion gravitate to the top of the list of instructional practices. In general, these practices, as hypothesized, encourage moderation in belief, and practices that generally do not encourage examination of opinion, do not. The four most influential practices show a relationship of above 5 percent more moderate, and these practices are all of the opinion-fostering variety. The effect is smaller than constraint, but meaningful, especially given the higher variance we might expect in such a measure of moderation as a measurement construct, "residual selection of more

moderate values independent of belief," which may understate actual effects. Discussing issues and current events aside, all effects are positive and statistically significant for these practices, while the others are generally positive, but not significantly different from zero.

Discussion/Conclusion

This paper starts with two open questions. First is how belief cohesion and even-handed opinion, dubbed constraint and moderation, might serve to augment "fact-based" or "general" knowledge as important, beneficial outcomes tied to political learning. The second point is one of instruction and how different educational practices might serve to influence these different educational outcomes.

I have produced two separate measures in order to augment the use of fact-based knowledge. The measures of constraint and moderation, based in how we respond to survey stimuli, are unique in how they assess the constructs of "constraint" and "moderation." Constraint is constructed by inspecting how students' responses "fit" a latent social welfare ideology and moderation is constructed by examining extreme versus moderate response tendencies. These measures appear to reflect the intended constructs, showing a strong positive association of general civics knowledge and constraint, and a modest general relationship between civics knowledge and moderation. Furthermore, they all appear, like civics knowledge, to be influenced by nearly all instructional practices.

Toward instructional practice, it is noteworthy that different instructional practices are related differently between these three domains of civics knowledge, constraint, and moderation. Civic knowledge appears not to be affected much by differences in practice that might touch on opinion, constraint shows some mild differences, yet moderation being influenced most specifically by opinion-related practice. One narrow point can be based on these results, with the

standard caveats that might accompany statistical analysis. If one were to accept that all of these outcomes are desirable and measured adequately, we might argue that opinion education and pedagogy that fosters consideration of opinion should be part of a core secondary school curriculum, or at least that this connection is convincing enough here to merit further study.

A broader point, however, might concern political science when confronted with popular political discussion about facts and beliefs. The general rhetoric of learning in politics is reinforced with a journalistic anecdote which is not atypical of how learning and belief are addressed. At large, it might be argued that opinion instruction erodes consistency of belief: it might also be argued that consistent ideology is not a hallmark of an informed learner. Both of these possible, yet implied, arguments in practice are fallacious. Political learning and belief have been studied for half a century with respect to ideological consistency, and the repeated result is one that shows the more educated an individual is, the more they tend to use ideological reasoning to connect multiple issues. The results in this paper merely parrot this result in a new setting: the 8th graders in this sample show higher ideological cohesion as civics knowledge increases and different modes of instruction tend to positively influence belief cohesion.

Where ideologues may have concern is when we attempt to measure a different concept: that of well-considered belief, even-handed consideration of issues, qualified belief, critical thinking, or "moderation." However this latent concept is most appropriately named, when students are taught about civics they will tend to moderate their responses. Yet this conclusion needs to be qualified itself. When accounting for general civics knowledge, moderation is independently influenced by practices that specifically target the encouragement and teaching of opinion. In this area, consistent with Zaller's theory of survey response, moderated response is more common as opinions are overtly examined in civics instruction.

Chapter 6. Political Discourse and Religion: Belief Constraint and Moderation

In this chapter, I look at two ways in which response can help add to our knowledge of how issues behave that shows how ideology can be augmented to inspect the character of discourse, and particularly discourse and its religious associations. Chapter 5 looked at content of curricula in nascent 8th grade civics instruction and how it may vary with respect to Constraint and Moderation and found that the extension of meaning to ERS which characterizes it as engagement with material, deliberation, and critical thinking is an instructive one.

This chapter uses the same concepts to look at religious groups and their possible value of teaching in another way. The important aspect of this chapter is the exploration of connection of ideas, and how these ideas ostensibly mature in the electorate for religious groups at varying places on an ideological spectrum. Until now, whether you are a conservative or a liberal was something that was roughly incidental to any of the cognitive strategies employed. This chapter takes a different tack, in addition to naming merits and infractions of political thinking, shedding light on religion in terms of which side of the ideological spectrum these cognitive infractions occur.

"Enough already about ideology" (Kinder 1982) has been a clarion call for the scads of research that has enveloped the ideological space. What is ideology? What makes people believe the way they do? How does ideology differ from partisanship, or work with partisanship, to secure electoral outcomes? How do we measure ideology through mass opinion, through Congressional vote, or Supreme Court rulings? If ideology is independently important, as well, how does one appropriately measure it, in how many dimensions? There are many threads of research here, too many to do justice to in a simple narrative, yet a compelling criticism remains implanted in mass ideology and its measurement. The roots of this are in Converse (1964) that

argues, at the very least, that ideological thought is not uniformly practiced, even as he may lambast the public's knowledge and use of ideological concepts.

Response style can be used as more than a mere curiosity and artifact of a particular person's response. It might also be used substantively to investigate how what drives response can illuminate our conception of political issues. This chapter demonstrates another such application, by using the nomenclature in the previous chapter (Constraint and Moderation), and using response to elaborate on the ideological and religious underpinnings of contemporary modern political cleavages in a unique way. Ideology has been discussed in terms of Social Welfare and Cultural Values, and religious placement therein, with the social groups who identify as "religiously observant" or "evangelical" occupying a uniquely right of center political topography, while other groups (Jews and Non-Christians, as well as the more centrist Catholics) fall in different realms of the spectrum. However, some criticism and queries about this work (and of mass ideological placement scales in general) comes with the strength of "traited" behavior and moderation of self-identified social groups. Knowing where respondents fall on the ANES survey of issues tells us little about how strongly aligned these same groups are, or how moderately inclined they are. The confluence of ideology, constraint, and moderation can tell us a more nuanced story of the religious right and left than opinion locations alone.

I construct two Item Response models from the 2008 American National Election Studies to complement the analysis of Economic or "Social Welfare" component of ideology. The first is aimed at creating a measure of social welfare belief consistency (person-fit) and the second targets a content-free measure of moderation. The results from this effort hint at a more nuanced interpretation of ideology and religious identification; a chance to outline the contours of devotion and religious group in terms of what facets of learning and deliberation are present and

a chance to reach a more nuanced understanding of an instructive (selective) nature of religion in politics and go beyond the Op Ed caricatures that may accompany the description of religion in political life.

A common criticism revolves around the question: "if one is a (capital M) 'Moderate,' or in the middle of the ideological spectrum, how do we really know whether or not this is *true* moderation or simply a lack of understanding?" It can be an acceptance of true self-placement or it can be an insidiously content-free middling response. Once this question has been posed, the follow-up asks whether the same might be said of those to the right or left in the political spectrum. How does one really know if self-placement represents a real ideological trait or whether someone's right or left views are somehow a product of truncated, or "other" reasoning? Certainly partisans make these claims consistently about whether another's ideology is correct.

This chapter is an exploratory use of two measures of ideological "traitedness" in relation to ideology, religious affiliation, and religiosity – whether one ideological "side" or the other uses their ideology more or embodies more considered belief. I examine how denomination and religiosity might play a role in relation to these characteristics of belief. In particular, does either the religious right or left follow a more consistent line of belief on Economic issues? Second, does one side or the other follow a pattern of "kneejerk" belief or are these beliefs nuanced and well thought out? I go beyond ideology to inspect not just where respondents place themselves, but characterizes the nature of belief in broad facets of coherence and thoughtfulness.

The concepts on which I elaborate presently: the measures here serve as independent tests of such behavior: constraint and moderation of belief based – not on self-reports - but on observed survey response patterns themselves. The question here is not just whether more "born again" Christians occupy the right of the ideological spectrum (on many issues they do), or

whether the irreligious occupy the left (they do), but whether it's possible to say something about the cohesion and even-handed nature of beliefs that make up these areas. I find that, with these tools, developed and meaningful for response, we might find evidence that religion does, indeed shape ideological issues. Importantly, it is not simply the traditional vs modernist ideological issues that are at play, but the very core of social welfare ideology, the "role of government" that has been a stalwart measure of political division in America that has not yet been superceded since the New Deal Era formed in American Politics.

Using Constraint and Moderation to Evaluate Political Discourse

These measures, again, show a tendency to follow a line of given connection between issues (Constraint measure by "Person Fit") and a tendency to select endpoints independent of the object of measure (Moderation measure by "ERS"). I argue again that both issue/belief constraint and moderation are normatively positive outcomes of political learning, or at least variables of interest, both (a) grounded in the merits of learning and adopting ideological self-placement based on prevailing discourse and (b) taking in various ideas and taking a larger sample of considerations respectively – a hallmark of critical thought. *To understand political dialogue is to be able to both place your belief in it and qualify it.*

The repetition of the mode in chapter 5 is deliberate. As teaching facts and democratic values is to 8th graders, we might develop a similar parallel to advanced instruction present in the socialization and moral teaching present in our faith-based life. The theme of teaching, knowledge, and persuasion in faith give rise to journalistic allegations that one side or the other is captured and suspended in a fact-free belief. A recent Op Ed by Paul Krugman seems to give a voice to this from a derogatory and one-sided place that is all too familiar. In his condemnation of Republicans as against learning and critical thought:

> "Last year the Texas G.O.P. explicitly condemned efforts to teach 'critical thinking skills,' because, it said, such efforts 'have the purpose of challenging the student's fixed beliefs and undermining parental authority.'" (Krugman 2013)

The implication is that there is a school of thought, and an eddy of political effort, that questions how students and adults are taught. "I don't want liberals (conservatives) teaching people what to think about government and politics" is certainly a part of this.

> "The truth is that America's partisan divide runs much deeper than even pessimists are usually willing to admit; the parties aren't just divided on values and policy views, they're divided over epistemology. One side believes, at least in principle, in letting its policy views be shaped by facts; the other believes in suppressing the facts if they contradict its fixed beliefs." (Krugman 2013)

Such rancor about "belief" and "knowledge" is commonplace. From the left, the right is lauded underhandedly for their values in action, yet chided for a perceived lack of scientific, economic, social prowess, or critical thinking, and from the right, simply derided as being "devoid of belief" or similarly tied to righteous causes that have no similar standing in fact. Yet, there is some intuition we might use here as an entry to study knowledge, consistency, or even-handed behavior that we might consider a hallmark of a "good citizen."

This caricature is illustrative, then, of two points that we might give voice to political scientists when confronted with these kinds of opinion pieces. One is a call to critically examine these components, critical thinking and belief strength independent of learning, both as nascent learners and as adults. In study, we often value the idea of opinion-free knowledge to the exclusion of other beneficial civic learning and behavior, yet political argument involves multiple dimensions of civic learning.

Measuring Constraint and Moderation

Having set forth why we should care about belief constraint/consistency and moderation previously, I turn to how we might reliably measure these constructs here. The measures of

constraint and moderation adopted in this paper come from two separate Item Response Models run on the same data. As a proof of concept, I use the ANES in 2008 to calibrate both models (n=1,146 are asked all questions) and with the same questions. These questions tap the main dimension of Economic or "Social Welfare" ideology.

Consequently, I rely on a battery of nine (9) items in the ANES that represent what we have come to know as this "social welfare" ideology, which serves to identify how much government should intervene to produce normatively beneficial outcomes in society.

- (3 point scale) Education Spending
- (3 point scale) Social Security Spending
- (3 point scale) Child Care Spending
- (3 point scale) Welfare Program Spending
- (3 point scale) Poor People Spending
- (7 point scale) Guaranteed Jobs
- (7 point scale) Spending and Services
- (7 point scale) Healthcare
- (7 point scale) Affirmative Action

One result of these modeling efforts are scales of social welfare ideology that resemble those that have been constructed routinely by analysts of ideology, roughly N~(0,1) with Republican and conservative on one side and Democrat and liberal on the other with shades of belief designated in between. Yet it is not these traits that are of direct interest, but the following measures developed as a byproduct of deriving them. Figure 6.1 shows the frequency distribution of the scale estimated in this paper from conservative to liberal. It uses a Nominal Response model (Bock 1972) which has been used more frequently in educational and psychological measurement to derive measures of an underlying individual trait of ideology, commonly notated θ_j, using the item response characteristics of multiple items where an item-specific propensity slope and intercept are estimated, maximizing the available category information toward the most nuanced estimate of ideology from these data.

Figure 6.1. Social Welfare Ideology: 2008 ANES

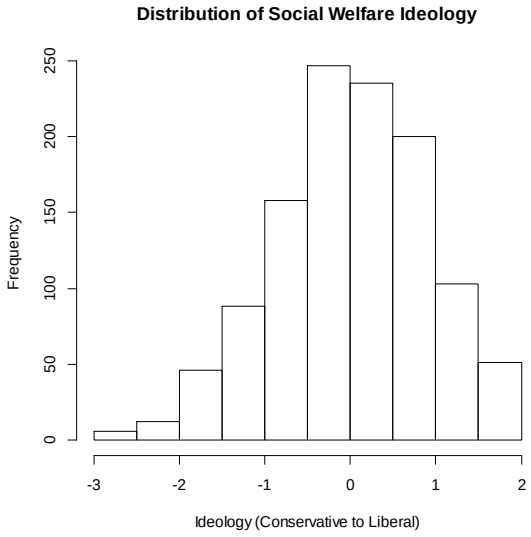


Figure 6.1 shows this relatively neat progression in ideology with slightly more extreme tails leading the more conservative (here left) side, but otherwise very close to what we might expect. Empirically, however, a couple measurement issues might be present, but are likely not too concerning. First, the fit data are skewed such that the bulk of responses exhibit reasonably good fit and less comfortably fitting responses tail to the left (Figure 6.2), making use of this metric as a continuous measure slightly difficult to interpret. This is routinely handled by setting the threshold by which aberrant-responding examinees are identified and was the original purpose of the measure, such as 0 which has been standard. Second, clusters of responses have their fit overstated by the presence of "I Don't Know" responses (notably the three rightmost bars in Figure 6.2). There is no specific guidance on how to proceed in the case of self-reported "misfit" (I Don't Know responses), but I make a requisite adjustment in order to make the typical response pattern absent modest DK or Refusal to be still positive and cohesive, a penalty of .3-.5 (.5 in this analysis).

Figure 6.2. Frequency Plot of Social Welfare "Person Fit" - Constraint

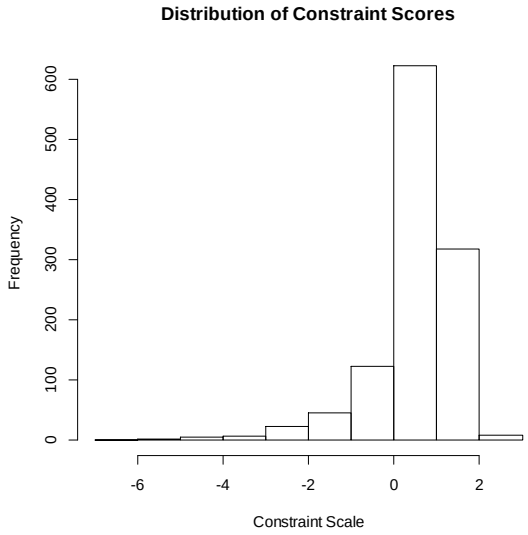
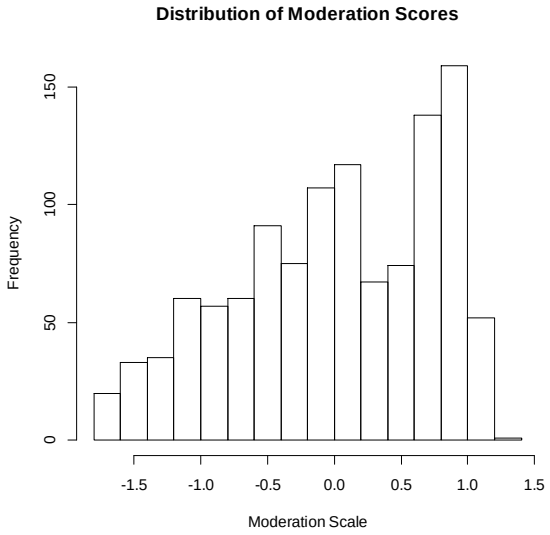


Figure 6.2. shows what the "standardized" metric looks like, with a negative skew, typical of the transformed L_0 metric, and most respondents showing reasonable fit to social welfare ideology. Compared to Figure 5.1, some of the same traits are evident, common to typical person fit distributions. Again, to account for the skewness, more pronounced here, and to create a more interpretable measure, I create a dichotomized variable, where index values less than 0 are coded to 0, and "don't know" responses are penalized proportional to the frequency with which they appear.[28] The rest of the reasonably fit responses are coded to "1" as respondents who exhibit reasonable ideological fit to the model (Meijer and Sijtsma 2001). The result is a measure that might be interpreted as "whether a respondent is approximately using social welfare ideology to constrain issue responses."

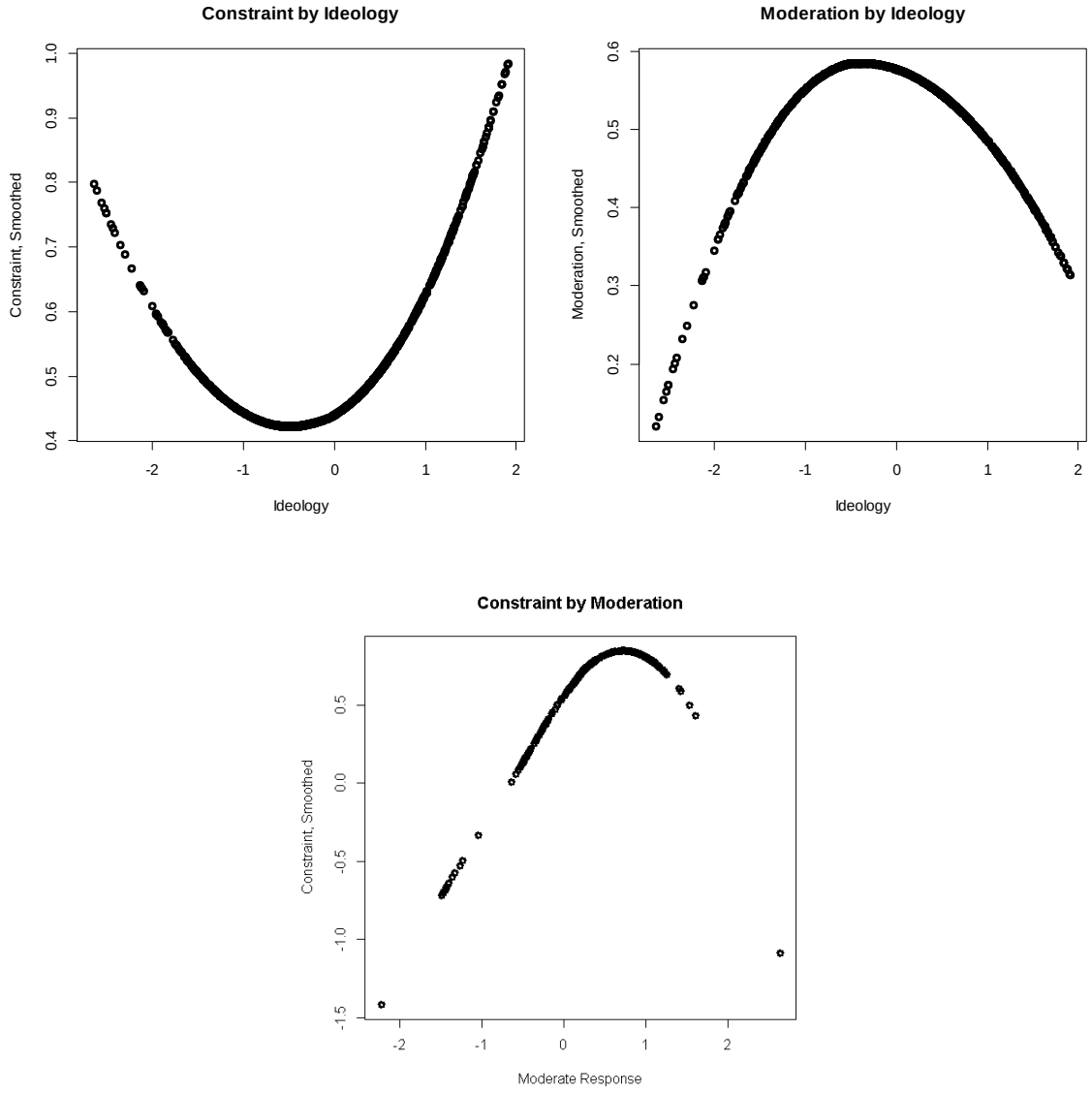
[28] The subsequent moderation model includes a robust "don't know" index used to discount the constraint of those who choose to abstain.

Figure 6.3. Frequency Plot of Social Welfare Moderation



Again, the figure shows some inconsistency with an ideal normal distribution, a bit more marked than its analogue in figure 5.2. The bimodality ranges from the middle and at a modal response of slightly less than 1.0, which implies very weakly that responses may gravitate to either "typically moderate" or "typically extreme." Again, by design, zero can be roughly considered average extremity, though not as cleanly as in figure 5.2 and the ERS score is again dichotomized (0=more extreme, 1=more moderate) in order to retain the previous interpretation: "whether a subject responded more or less moderately than average."

Figure 6.4. Constraint and Moderation by Social Welfare Ideology



What becomes clear on inspecting these relationships (Figure 6.4) is that constraint and moderation are related directly to a level of ideology as a measurement construct. As ideology goes from conservative to moderate (the left panel), the proportion of respondents in the population who exhibit firm ideology drops from a substantial majority to a minority of respondents. Those who are moderately conservative to middling are less likely than not to adhere to social welfare ideology, while the more ideologically inclined, and particularly liberal

respondents, show a consistent ideology in a U shape. The right panel shows the opposite pattern for moderate response, where majority the middle of the ideological spectrum moderates their response (between -1 and 1) and the minority of the more typical ideologues moderates response. This confirms the intuition that those in the middle appear to be a mix of respondents who are (1) truly moderate, (2) those who are not constrained in their response (do not detectably use ideology to evaluate social welfare issues), and (3) those who moderate as a response "rule."

While Figure 6.4 represents loess point estimates (the figure omits the dispersion of points), it does tell a story of a relationship that exists, yet is not traditionally linear. While there is little linear relationship to speak of, the responses of those in the political center show a distinct parabolic shape here (and not unanticipated) (Reise 1990, Bolt and Newton 2012). The political center shows less constraint, and more moderation, solely based on their ideological position. This is an artifact of the response patterns we might anticipate by the models – all else equal, conservatives and liberals on social welfare will have more evidence of constraint and moderate their responses less, solely based on ideological positioning.

In the bottom panel of Figure 6.4, I present Constraint as predicted by Moderation to round out the display. This panel shows both standardized metrics of Constraint and moderation where constraint is a function of moderation. The plot shows that the trend of respondents who are moderating their response show a constraint level above zero, and that extreme responders also tend not to have a great deal of constraint. This shows a level of consistency of our underlying measure where we might expect examinees who are not aware of ideology to also exhibit extremity, low levels of learning (or abstention) is associated with low levels of deliberation. By contrast, moderately high levels of learning show more diversity in their moderation.

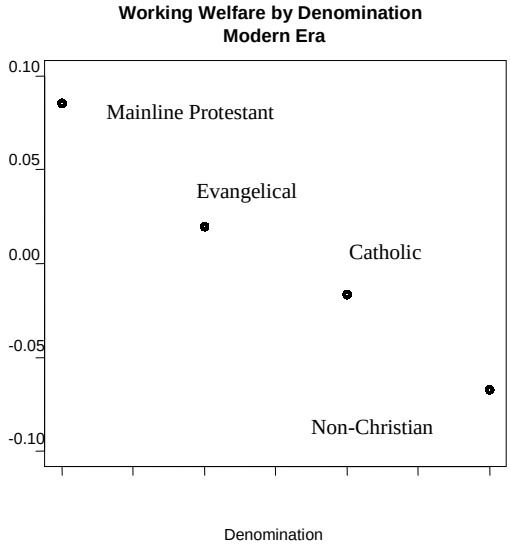
These plots show us that constraint and moderation are behaving largely as we might expect and provide valuable information to take into any further analysis – both as a function of ideology and in relation to each other. Yet the non-linear relationships present some difficulties in showing how these latent variables behave. In order to remedy this ideological "dependence" and attempt to test hypotheses broadly, I examine constraint and moderation measures *(a) with respect to the underlying levels of ideology and (b) with respect to overall population averages*. In graphical terms, this means I inspect the deviations of each group along the entire ideological continuum in a localized polynomial regression (loess Linear Probability Model) framework and deem "significant" those areas which exhibit statistically meaningful difference from average along any interval that is predominantly liberal or conservative. That is, if conventional significance ($\propto \leq 0.05$) on the loess model is significantly greater at any point along the ideological scale on either side of the spectrum, the results will be deemed significant.

Formulating Expectations of Religious Constraint and Moderation

Having derived measures of constraint and moderation, then, what do these social welfare contours look like for the dominant religious groups and their observance patterns? Not just their ideologies overall, but the constraint and moderation of these ideologies for different religious groupings and devotional profiles?

Hypotheses may abound, but I start with the notion that we are generally aware of the densities of these respondents and where they are positioned. Evangelical Christians and Mainline Christians occupy a predominantly rightward stance on social welfare whereas Catholics are a mix of right and left (a swing ideological constituency, if you will), while "Other" Non-Christian respondents generally cluster leftward on economic ideology (Claggett and Shafer 2010).

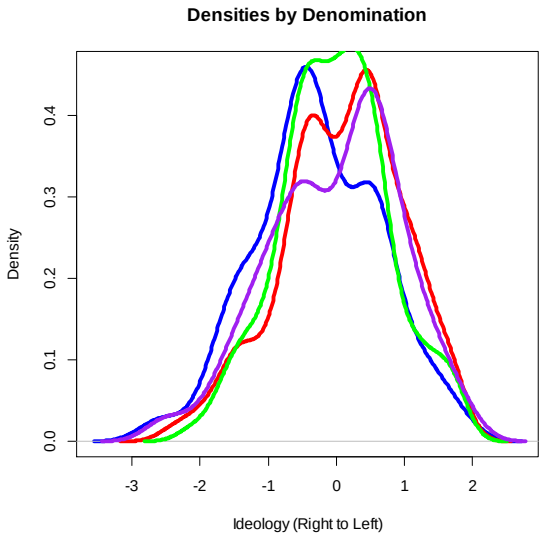
Figure 6.5. Views on Working Welfare: ANES 1992-2008



Measure from Claggett and Shafer 2010. Mainline protestants include denominations: Congregational, Episcopalian, Lutheran, Methodist, Presbyterian, and Unitarian.
"Evangelical" Protestants are all others.

In this chart, on a standard normal metric of social welfare (Claggett and Shafer 2010), we might observe the divisions in social welfare by denomination. The chart show a general, though weak, tendency for mainline protestants to veer to conservative, evangelicals to be slightly conservative, Catholics to tend toward the liberal end, and non-Christians to be most liberal. What may defy some intuition is the relative modesty of the conservative stance overall of those protestants dubbed "Evangelical." While, on social issues, this group may lean heavily conservative, on social welfare issues, they are more "middle of the road." This may be correctly placed, but may also be an artifact of some evangelical groups who tend to identify with social groups who are more disadvantaged, low income, lower educated, or identify with a racial or ethnic group that values equality (African Americans most significantly, but increasingly Latinos).

Figure 6.5. Views on Social Welfare Ideology by Denomination: ANES 2008

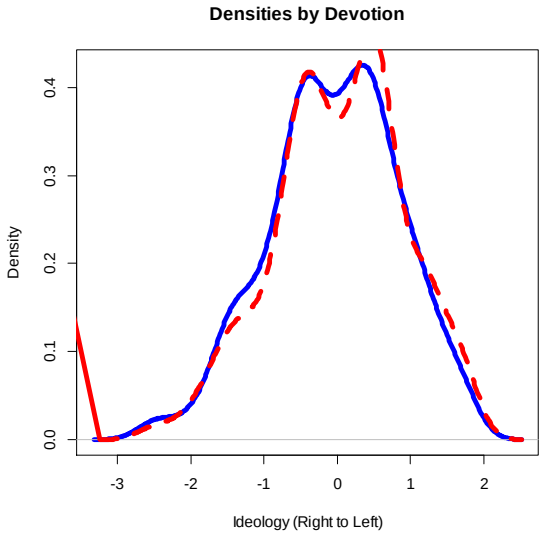


Blue=Mainline Protestants, Red=Evangelical (Other) Protestants,
Green=Catholics, Purple=Other Non-Christians

Figure 6.5 represents a density chart of these religious groups in the 2008 ANES sample. In this 2008 ANES sample I use in this paper, the conclusion that "Evangelicals lean conservative" is further eroded – or alternately may need to be subsequently qualified. While the blue line of Mainline Protestants is generally conservative leaning, other Protestants tend to lean here in a liberal direction – or at least maintain a bimodality that has a primary density that leans liberal. I will retain the hypothesis that these other Protestants will tend to be more constrained and less moderate despite this fact. Catholics here are the predominant centrists of these religious populations and Non-Christians predictably anchor the denominational left. Another artifact of this measure of ideology is the density blips around the zero point. With the possible exception of Catholic respondents, each of these denominational categories *decrease to the zero point*. The exact "capital-M" Moderates are not showing up in religious group typically. This smaller observation may be taken too far, yet it is notable that liberals and conservatives separate here, revealing a tendency for measurement to show that this population is what we might expect – divided on ideology such that the exact zero point is not a modal point estimate. In fact, it

comports with Figure 6.4, in that the zero and near zero points are comprised of more moderates who moderate because they are unconstrained or habitual moderate responders, and not because of pure ideological location. The result is a true density of ideology that is effectively (but mildly) bimodal by nature.

Figure 6.6. Views on Social Welfare Ideology by Devotion: ANES 2008



Blue Solid=Every or Almost Every Week,
Red Dashed=All Others

As far as devotional status (Figure 6.6), there is also little stark difference in ideological densities. Those respondents who practice their faith weekly or approximately weekly are slightly more conservative, but not markedly so. The blip that may be of some limited interest is that where the less devout tend to cluster slightly more on the modestly liberal end and are perceptibly, though again not substantially, more likely to be on the far left and less likely to be on the far right. And again the division of ideology has "zero" as a less frequent point estimate than we might hope in a strictly normal distribution, yet not inconsistent with division on ideology that we might expect in the population.

Knowing where these groups position themselves in the 2008 ANES, we might go beyond the more base partisan accusations of "kneejerk belief" vs "beliefless" and enter in a more nuanced assumption that the social welfare ideologies embody beliefs that are taught (implicitly or explicitly) religiously, either in the pulpit, or through social identification. I graphically assume that these ideologies are exogenous, for some simplification purposes, though they need not be interpreted as such. They may be a product of social group, self-selection into congregation, or mediated/moderated by some other process. Nonetheless, it is instructive that the relationship in other studies has been relatively strong, and that Economics in America has some intellectual roots: anecdotally, the "social gospel" in Evangelical and Mainline Protestant denominations, and the Catholic vein of social justice. Add to this a normative teaching of equality of opportunity versus equality of outcome; efficiency (an anecdotally, perhaps caricatured protestant ethic) versus mandated equality. However this is framed, there is some version of teaching and learning that happens and happens in a way where religion is a possible, normatively important covariate of separation.

The following Tables 6.2 and 6.3 sets up some of what we might expect, holding partisan rancor aside and coming from the realm of positive social science. A first expectation revolves around the locations of where respondents are: namely that the coherence of economic ideology is most observable where the bulk of respondents place themselves. That is, there is a dominant teaching or identification such that "where the bulk of respondents are" is where the ideological coherence lies. In Table 6.2, Mainline and Evangelical Protestants will exhibit ideological coherence (constraint) on the right, where the left will exhibit cross-pressure on their religious affiliation. I might dub this kind of expected effect as *home constraint*. That is, where

denominations and devotions are likely to find their ideological homes are where they are most likely to foster deeper connections.

The second, weaker expectation revolves around that where the most coherent ideology exists by religious affiliation is where automatic extremity (moderation) will be most visible – namely that evangelical protestants on the right (a) will adhere to a program of belief, but exhibit more extreme responses independent of their ideology – they will exhibit not just constrained thinking of ideologues, but exhibit an independent behavior of extremism above and beyond ideology. Yet this expectation is weak, and the weakness of this expectation comes from its colloquial nature. Adherence to an ideology does not necessarily imply that they are "independently" extreme, just that their preferences align. In a more derogatory way, this set of hypotheses goes to a bi-partisan intuition which derides each ideology's extremes as substanceless and unthinkingly adherent to an elite program. In real terms, this expectation tests this in terms of response pattern. I might dub this kind of expected effect, on the other hand, as *cross-pressured moderation*. That is, where an ideological home is not is exactly where qualification would be most likely to occur. In ideological homes, by contrast, we might expect to see Krugman's Op Ed effect where ideology is more set in and dogmatic.

Table 6.2. Constraint and Moderation by Religion and Ideology

	"Ideological Right"	"Ideological Left"
Mainline Protestant*	C↑M↓	C↓M↑
Evangelical Protestant*	C↑M↓	C↓M↑
Catholic	--	--
Jewish / Other	C↓M↑	C↑M↓

*-Mainline protestants include denominations:
Congregational, Episcopalian, Lutheran, Methodist, Presbyterian, and Unitarian.
"Evangelical" protestants are all others

The devotional expectations run between the observant versus the less observant (Table 6.3) and can be thought of in a few separate ways. First, we may expect that the devout have a

"home" on the right and calibrate our expectations like the above where home constraint and cross-pressured moderation run accordingly. However, the alternate hypotheses I might introduce here imply that more observant will simply have an increased consistency of response across the ideological spectrum. This is an expectation that comes from what we might think of as teaching and learning of social values (Social Welfare) that happens either by design by leadership, or in a group with similar values.

Table 6.3. Constraint and Moderation by Devotion

	"Ideological Right"	"Ideological Left"
High Devotion (Every or Almost Every Week)	C↑ M↓	C↑ M↓
Low Devotion (Seldom or Never)	C↓ M↑	C↓ M↑

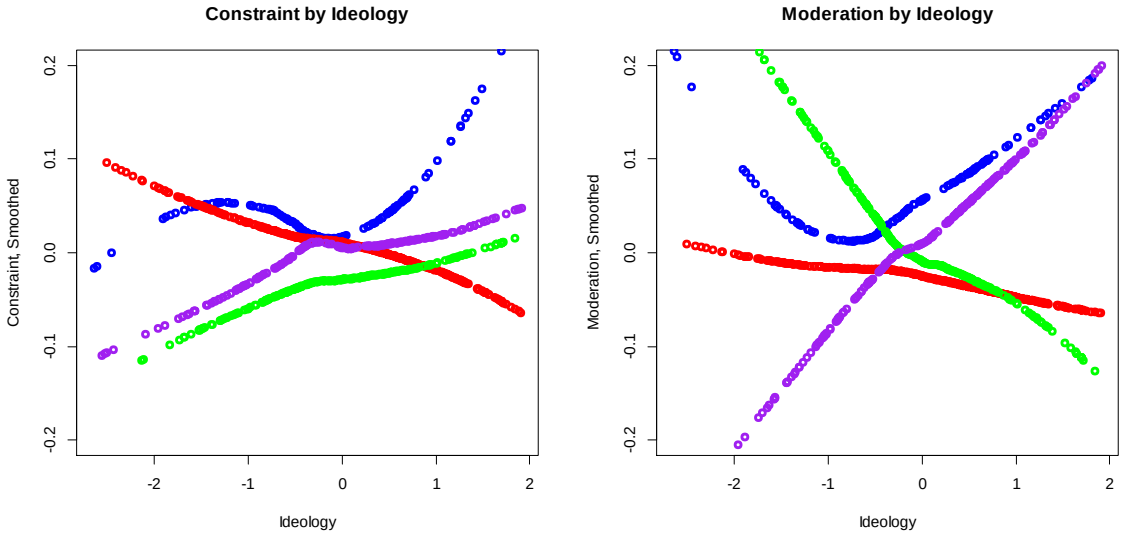
These hypotheses in hand, it might be noted that testing on the whole generally suffers from limitations here. First, the analyses are simple bivariate ones. Other covariates may be at work here that may bias interpretation: income, education, race, and gender are a few. With such a weakness, definitive conclusions are weak. Second, with variables as imprecise as these and with a calibration sample of just over 1,000 will tend to hold Type I error well in check, yet exclude potentially meaningful results unnecessarily. With this in mind, while hypothesis testing is the standard, I go beyond this testing to interpret the results somewhat liberally, and also make some suggestive inferences about what these data may say by pattern and what they do not imply. As a pilot, this piece may be limited in a scientific sense, but serve as a more descriptive indicator of what might be said about different ideologies, religious groups, and devotion. With this said, punditry and general discourse often make the same errors in misspecification, so to a broader point: even if naïve, the larger point remains in civic discourse of whether we are being "true" to religion and devotion when we talk about even basic measures of belief cohesion and even-handed response.

Denominational Constraint and Moderation on the Ideological Right and Left

Now we venture into the realms of questions: at each level of ideology, from conservative to liberal, how do denominations behave with regard to how they adhere to a social welfare ideology and how they moderate their response, and for which levels of ideology? We might expect that, from political science research, in a denomination's ideological "home," they will exhibit more constraint and less moderation and that by devotion there will be an element of belief cohesion regardless of ideology.

Figure 6.7 shows local polynomial regressions (loess) of the different denominations in the ANES and how the measures of ideological constraint and moderation deviate from the population average. The interpretation of the y-axis is an indicator of "net difference in constraint," that is, how much more often do respondents fit a model of social welfare ideology? Do they adhere to an ideology at that location more than average?

Figure 6.7. Constraint and Moderation by Denomination



Blue=Mainline Protestants, Red=Evangelical (Other) Protestants,
Green=Catholics, Purple=Other Non-Christians

Table 6.4. Results: Constraint and Moderation by Religion and Ideology

	"Ideological Right"	"Ideological Left"
Mainline Protestant*		
Evangelical Protestant		
Catholic	M↑** Interval: (-1.61,-0.92)	
Jewish / Other		M↑** Interval: (1.05,1.09)

*-Mainline protestants include denominations:
Congregational, Episcopalian, Lutheran, Methodist, Presbyterian, ad Unitarian.
"Evangelical" protestants are all others
**Interval designates the minimum and maximum values of ideology for which the LOESS LPM model is statistically significant at 0.05

For constraint (Figure 6.7, left), the results are modest and only mildly suggestive. In statistical terms, no denomination is significantly more constrained in belief (Table 6.4), at any interval, than the population average. However, the results do suggest a possible simple division between protestants and other respondents on the right – namely that protestants hold more consistent conservative views than Catholics or Non-Christians in this sample. From roughly -2 standard deviations (to the conservative side) to -0.5, protestants of all stripes are 5 percentage points more likely than average to adhere to belief consistently between questions and the others are generally 5 percentage points less likely than average and the difference between them is significant – though caution must be exercised in this hypothesis. Also, suggestively, the protestant division on the right of the chart (left of the spectrum) seems to break down. Evangelical (or "other") Christians cluster with Catholics and Non-Christians while Mainline Protestants, the few that remain tail higher. In sum, however, the interpretation here is generally one of non-result, but where there is a hint (albeit post-hoc) result, the direction is of Protestant constraint for conservatives.

Moderation, however (Figure 6.7, right panel), shows more stark relationships. Significant results do include the Catholic sample for conservatives as being significantly moderate in their response (strictly on a moderate to high conservative interval) while, on the liberal side, Non-Christians tend to moderate their response significantly (strictly on a moderate to high liberal interval). Again, however, there are some conclusions that present themselves immediately that go beyond testing of significance. First, form these data, is the relative lack of support for any notion that more extreme response is connected to consistent belief. In this sense, results are mixed enough to avoid any conclusions whatsoever. To wit, there is no plausible simple link between a denomination's belief consistency and its level of moderation. Second, and similarly, Evangelical Christians show very little deviation in moderation from the population at large. This is true at every ideological level. The, perhaps less than obvious, non-result follows: at least in the realm of social welfare ideology, Evangelicals are no more or less extreme than any other religious group. Overall, as well, there does not seem to be strong evidence of high constraint and low moderation in places that represent each denomination's ideological "home." The only hint of this kind of "home constraint" phenomenon is in the Protestant proclivity for constraint on the right. Likewise, the moderation hypothesis of "cross-pressured" moderation does not hold. Catholics on the right may be somewhat cross-pressured on social welfare, but Non-Christians on the left are certainly not.

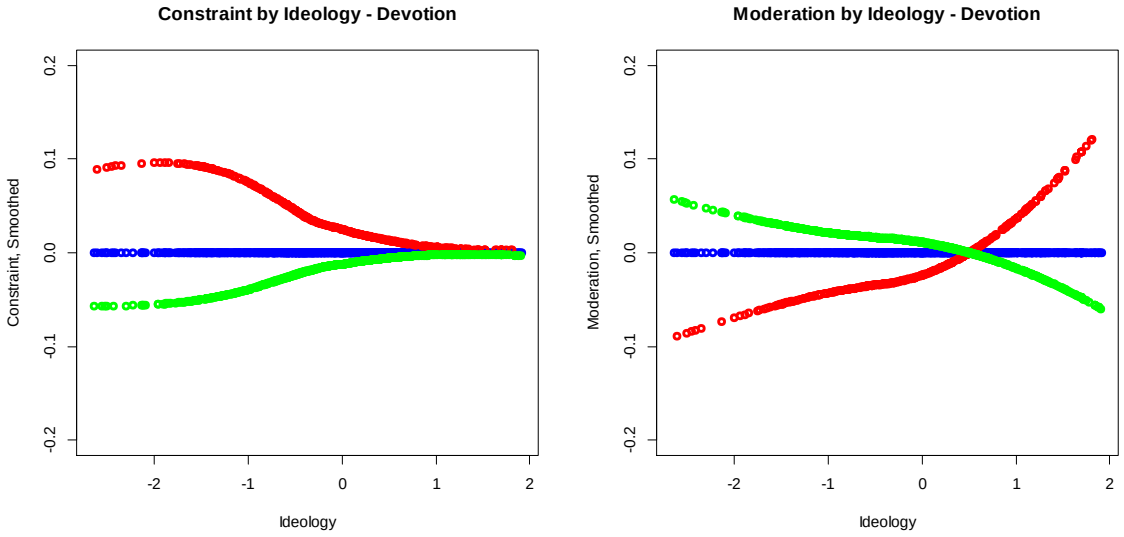
In both right and left of Figure 6.7, it becomes clear that we might consider a linear association of difference. Instead of inspecting the ideological continuum for areas of relative constraint or moderation, we might consider, for each denomination, as liberalism increases (decreases), is there a general trend apparent toward (against) moderation. The results for linear fit parrot those of Table 6.4, namely that no linear association exists by denomination between

ideology and constraint difference and, for moderation, the negative relationship in Catholicism and the positive relationship among Non-Christians hold.

Devotional Constraint /Moderation on the Ideological Right and Left

To the next set of hypotheses: does devotion increase constraint and decrease moderation? Is there something apparent in simple association between the degree of devotion and the extent to which social welfare ideas "hold together"? And does devotional frequency decrease moderation?

Figure 6.8. Constraint and Moderation by Devotional Groups



Blue=Population Average, Red=High Frequency of Devotion (Every Week/Every Other Week), Green=Low Frequency of Devotion

Table 6.5. Constraint and Moderation by Devotion

	"Ideological Right"	"Ideological Left"
High Devotion (Every or Almost Every Week)	C↑** Interval: (-1.19,-1.08)	
Low Devotion (Seldom or Never)		

**Interval designates the minimum and maximum values of ideology for which the LOESS LPM model is statistically significant at 0.05

Figure 6.8 and Table 6.5 speak to these results. In terms of constraint and moderation, there is little statistically significant evidence from these data that indicate devotion has a uniform effect. The lone significant effect is on the conservative end of the ideological spectrum, where devout conservatives show more ideological cohesion than average. Suggestively, however, there is an avenue for future study in that, when viewed from the lens of devotion, belief constraint does not appear to comport with moderation. Though not significant in moderation, there is a general rise in moderation to the left of the ideological spectrum for the observant, and a fall to the left of ideology for the non-observant. The observation follows that, by simple association, the frequency of observance is associated with the (normatively positive) use of social welfare ideology toward creating a coherent belief system and the (normatively negative) use of extreme response. Thus the "instructional model" that devotion increases constraint and decreases moderation uniformly does not hold for all levels of ideology, and similarly, the home constraint and cross-pressured moderation hypotheses do not hold either. In a devotional sense, the left and right of the political spectrum appear to be acting independently. That is, how devotion behaves seems to depend on where on the political spectrum you happen to be. Some patterns appear to occur on the conservative end, but no discernible pattern is occurring in social welfare ideology on the left whatsoever.

Discussion/Conclusion

This chapter has proceeded in two parts. First is in measurement: how belief cohesion and even-handed opinion, dubbed constraint and moderation, might serve to augment political ideology as important indicators of nuance of mass political belief. Second is the question of how these indicators perform in practice and relate to religion in both denomination and frequency of practice. Common dialogue and punditry speaks often of right-leaning Evangelicals or the left-

leaning Jewish or secular populations and notes how religion and belief systems intersect. Editorials and partisans use these caricatures of traditional zealotry or the godless modernists to then characterize the thoughtfulness, knowledge base, or even-handedness of their own beliefs and those of others.

I also look at creating respondent measures that serve to represent our colloquially expressed "depth of knowledge." I have produced two separate measures in order to augment the use of social welfare ideology. The measures of constraint and moderation, based in how we respond to survey stimuli, are unique in how political scientists measure latent traits that pundits and experts might reference offhand and colloquially, but have not sought to measure explicitly to evaluate these expert references. Constraint is constructed by inspecting how responses "fit" a latent social welfare ideology and moderation is constructed by examining extreme versus moderate response tendencies, independent of the object of measure, and indicating a moderation that can inform our views of ideology, yet not be, itself, a product of it.

These results signify a mass public that does not cleanly follow the rules of popular dialogue, if at all. By denomination, values on the right may signify a coherence of belief for Protestants of all denominations, but weakly, and that Catholics on the right and Non-Christians on the left display moderate response tendencies. By devotion, the more frequently observant are more cohesive in their ideology on the right (significantly) and less moderated (though weakly).

A broader point, however, might concern political science when confronted with popular political discussion about facts and beliefs. The general rhetoric of learning in politics is reinforced with a journalistic anecdote which is not atypical of how learning and belief are addressed. At large, it might be argued that opinion "instruction" erodes consistency of belief: it might also be argued that consistent ideology is not a hallmark of an informed learner. If we

believe political scientists, both of these implied arguments are fallacious in practice. Political learning in belief have been studied for half a century with respect to ideological consistency, and the repeated result is one that shows the more educated an individual is, the more they tend to use ideological reasoning to connect multiple issues. It is by understanding that we connect issues. The results in this paper merely use this result in a new setting: the 2008 ANES sample shows higher ideological cohesion for Protestants and for the devout on the conservative right. While this may be an artifact of some other variable, education or income perhaps, stating it like this prompts us to temper public thinking when engaged in partisan debate.

Again, it is by understanding that we connect issues, and, I argue it is a hallmark of understanding that we then moderate our response to them. To this, I measure a different concept: that of well-considered belief, even-handed consideration of issues, qualified belief, critical thinking, or "moderation." However this latent concept is most appropriately named, in this study it does not cleanly finger any particular group for criticism on its face. Catholic respondents appear to significantly moderate their responses from the conservative end, and Non-Christians tend to be most circumspect at the liberal end. Suggestive evidence also exists that, for the devout, circumspect/moderate response increases toward the liberal end of the spectrum and for the less devout, moderation decreases (though weak).

None of this detracts from what is *not* shown in these analyses. Namely that what we might call the "religious right" is somehow fostering traits that are negatively associated with what we might view as normatively negative. Whether or not the significance of all Christian groups on the right hold, what is definitely true is that, in social welfare terms, Mainline Christians, Evangelicals, and Catholics on the right of the political spectrum are not fostering traits of ideologically inconsistency and extremity that I argue we should value. And I argue that

I am not alone in this value. Constraint, while pejoratively might be looked upon as an unthinking adherence to a dogmatic point of view, is not a uniformly reviled trait. It is, in fact, one trait whose absence we had once lamented. Much like Responsible Party Government, when absent it is a problem for democracy and representation, but when present is also seen as a liability, especially that the electorate and elites are ideologically divided (read polarized) into two camps.

To polarization and religion, we might hypothesize that being of a certain religious background or choice, or that being devout fosters these deep divisions, or short of fostering these divisions is a reflection (covariate) of these divisions that threaten democracy. This analysis hints that this is also not the case. Denominational rifts in "moderate" response, response which has a more circumspect and qualified character, do not indicate that critical thinking is notably absent in these populations. In terms of devotion, we might observe that the religious right is more prone to more extremity and less moderation, but the magnitude of this association is still not conventionally significant. The hint is that devotion and the ideological right behaves much differently than denomination, and there is much nuance to uncover that does not fit what we believe is the journalistic narrative about these groups. Again, it cannot be interpreted from these results that any group is particularly, or consistently uninformed or non-deliberate in a Conversian sense. Yet it is also true that no group is running particularly afoul in creating the soft characteristics here that enable a polarized environment.

It is also remarkable that the simple hypotheses I construct using basic intuition grounded in political science research and popular dialogue do not hold cleanly for the political right and left. Namely, the ideas grounded in *home constraint* and *cross-pressured moderation* do not hold either for either the political right or left. It similarly does not hold that the constraint (and

accompanying lack of moderation) due to devotion behave consistently on the right and left of the political spectrum. This kind of result throws some caution into how analysts and observers of politics view belief, and how ideological belief and religious belief intersect in politics. The picture is more nuanced and contextual than these blunt hypotheses might uncover. Broadly, a rearrangement of thought may need to occur where ideological values in social welfare exhibit very different characters on the right and left, where conservative home constraint and less moderation may be the mode impressionistically (but not deterministically) and liberal soft traits that of ideology are more idiosyncratic.

Extension of this research would, no doubt, benefit by looking at a deeper array of covariates, a more pointed traditional versus modern political cleavage than social welfare, and sharpening the measure by borrowing from different years and questions that I sidestep here for purposes of depth and hypothesis generation. This effort would, necessarily, require a larger sample that can stand up to disaggregation more broadly. This analysis reflects one year of data from the ANES, which, at the precision of these variables, does not provide the kind of sample that will show small differences well. Additionally, I have picked social welfare ideology purposively, in order to see what can be seen by the most stable and enduring ideological cleavage in the electorate. To gain some insight into traditional versus modern ideological cleavages, much less robust questions will need to be chosen. Additionally, in my previous research, these traditional ideological divisions are slightly more considered and qualified than their social welfare counterparts, making more precise measurement a greater challenge. There is much yet to be seen beyond the window here.

Yet, at the very least, by looking at how these relationships unfold, we might begin to structure a more subtle way of viewing belief, the intersection of social and political with the

religious, not through the beliefs themselves necessarily, but *how* they are held, which is the larger point of this chapter. We might do well to listen and test the more subtle arguments we make in politics – that one group is thinking more deeply and/or is smarter about their beliefs than another. Positive political science tells us that believing one thing or another is not in itself bad or good, and it often tells us that the formulation of belief along a partisan political spectrum is normatively good for democracy and representation. Considering the alternative, how might a group of representatives act together to form a majority of opinion if that opinion is indecipherable? A prerequisite for representation in an American system must be formulation of a belief along a spectrum that makes it possible to be represented by one of two "teams." The measures of constraint and in this paper shows that we can, indeed, measure this basic formulation, and measure it better by looking at *how* respondents formulate their beliefs. Constraint lets us know if there is a baseline of informed belief, and moderation may help us find out if these beliefs have gone too far. Are they cohesive to the point where critical thought and circumspect, measured consideration are sacrificed? A cohesiveness of belief coupled with unmoderated response is the normative risk. By these measures, we might delve into whether or not these normative conditions hold, and for whom.

At least when it comes to religion, only the devout on the right come even close to meeting this threshold which leads to negative and polarized dialogue, and even then, not strictly for denominations that we might expect. Divisive political argument in the press conditions our expectations of who is most at-risk of running afoul of sullying democratic discourse by creating conditions of ignorance or behavior that counteracts deliberation. These arguments may be grossly misleading. Op Eds make the point that the godless left and the fanatic religious right are prone to making errors of belief. At the very least, students of politics might do well to critically

examine self-interest in these characterizations, and, by this research, look elsewhere, or deeper, for the threat to America's potential discursive ills.

Conclusion

In this thesis, I have attempted to answer a few key questions about response style in political surveys. These main questions include:

(a) Does response style exist? Is it possible to verify response style experimentally?

(b) What are some of the potential biases we might expect to see?

(c) If there is overt meaning in response style, what might it be, and what might it tell us about political questions, political opinion, and formative learning in the electorate?

(a)Existence and Verification. The first question might strike some as a bit pedestrian. No doubt some have even been bored by it. With a long history of response style research, much of which has lived and is developed outside of politics, there is really very little to doubt about the subject. Extreme response style exists and when we use the specific historical situations where these questions are experimentally manipulated, we can detect it. However, there are still vestiges of thought about public opinion that maintain that false opinions do not really exist. By demonstrating they do, repeating how and under what circumstances they do, and creating a short inventory of likely bias goes some way to reinvigorating our skepticism about some of what we think we know about public opinion. We need to be skeptical of an assertion that stylistic response does not exist, because it is as close to consensus that there is little truth to the assumption that these types of question do not contain bias, and do not inflate error due to question format alone. By addressing skepticism with a strong measure and with experimental data for causal inference, the groundwork for the rest of the this work is laid, and avenues begin to open up for further study.

(b) Potential Biases. More interesting than simply manufacturing solid causal estimates, is whether response style might be troubling and whether we should be concerned in our more quantitative behavioral survey work. The answer here is contextual, and I shed light on some of these questions throughout this project. On one hand, for a wide swath of questions and respondents, the answer has to be "no." There are those who are obeying the implicit survey rules of the road, according to our implicit expectations, and we might have faith that they are being truthful inasmuch as they choose, are aware of detail in their preferences, and can communicate them on a variety of topics well. It is also true that surveys do not typically employ less than optimal questions intentionally, though some legacy questions persist which largely creates grist for this dissertation, even though the questions may be retained simply for the sake of historical comparison, and not for their technical qualities. Some questions may also be engaging enough (see Abortion in chapter 3) and unambiguous enough that strength of preference is not typically or detectably present, even when people may be prone to response style and a Likert scale stimulus is presented.

Yet, what comes through time and again in this thesis and in research on response style as a whole is that we need to be mindful of the populations that are less engaged, when they are at risk of bias, and under what conditions these biases occur. Unfortunately, these conditions are met more than we might care to admit, in no small part because it is of theoretical importance in democracy to learn what moves voters who are not particularly involved or aware in politics, and those voters will typically intersect with those racial/ethnic minorities, lower income, lower educated, and less politically aware individuals who engage more heavily in stylistic response. Academics worry about them because these less engaged voices may not carry the weight they should and that their absence, or the "manipulation" of these voices by elite actors might carry

with it referenda on the health of democracy. We would like to know what marginally engaged voters think, and yet some survey measures are less well equipped to handle these opinions.

If we can measure response style well, which is likely in the case of ERS (and manageable, yet tenuous with ARS), it is more compelling to get to a place where we can identify what the likely magnitude of the bias, which motivates the inspection of a simpler case that considers true political extremity. By invoking actual, observed political behaviors, it might be demonstrated if, and to what degree, response style is meaningful. A consistent, and intuitive, foil for testing impact in its raw form is this: true opinion will influence behavior, whereas untrue opinion will not. And the results here are very clear. Extreme respondents are no more likely to engage in political activity on balance than anyone else, and, in fact, are typically less inclined to vote than respondents who hold less extreme views, controlling for partisan political evaluation. Contrast this with those who are true ideologues, and the difference becomes even more meaningful, where those with genuinely strong views are almost certain to vote and much more likely to participate in other political activity. If we mistake true belief for stylistic tendency, then, analysts do themselves a disservice which is magnified when we consider subpopulations that are identified using contaminated measures.

I also demonstrate that, to a lesser degree, response style introduces bias even outside of the immediate context of response style, influencing survey results where it should have little direct bearing. If we are concerned about true attitude and its connection to behavior, which we are, we will get a less than true picture of the behavioral consequences of opinion. In this case, one that waters down the effect of true opinion (true extremity).

More than that, by demonstrating an effect outside of the context of the response itself, the implication is that survey opinions may carry a degree of bias that does not simply reflect a

lack of thought on a particular item, but may extend to other areas of political reasoning for substantive reasons we may not fully appreciate. It must be considered that response style is itself an influential trait that has meaning outside of its original context. We could conclude simply that responses on "bad items" would be suspect, and subsequently that scale relationships will also be affected where these items are combined with others. Yet even in relationships that are separate from the immediate response, we might observe a small bias. Added to this, a bias appears to exist independently of its documented demographic covariates. The correction procedure, inverse probability weights, appears not to add much by weighting based on observable characteristics that are associated with response style such as race, income, and education. When we extract response style, it not only cleans the data that are left behind, but on its face it seems to have influential characteristics, and characteristics that do not follow necessarily from whether it is associated with an identified set of demographics that we might capture in these analyses. It becomes apparent that *something* is at play, and that what is at play is at least consistent with a cognitive strategy that chooses to answer questions incompletely.

In tackling and verifying the existence of response style and its potential bias, not only can we see that response style exists, but just as importantly, it confirms that *it is possible to measure this trait in practice* and measure it well enough to be able to inspect what it might be and what use it might have as an independent construct. So the revelation of the first two chapters in this light is far from pedestrian. We have found an impurity in our data, we can identify its properties, pick it out, and isolate it the best we can. That way, it is not merely helpful for analyses that intend to use data that is left over when response style is taken out, but to critically examine what this measure has produced. Once we have the impurity in hand,

inspect its properties, and catalogue potential bias, what might we learn by using it? Can it be used? Can we attribute meaning to this bit of survey refuse?

(c) The Meaning of Extreme Response Style. The answer is: "yes." When we look at gun control as an issue in chapter 1, for example, there seems to be little in the way of response style. Yet other questions seem to bring out response style more starkly which opens the door to asking why this is the case. If we entertain the notion that response style is ill-considered response, and that this ill-considered response has meaning for the individual and the content we hope to analyze, curious patterns appear that do not conform wholly to our ideas of what is important and engaging in the electorate, how this reflects on our view of institutions in society, and how we learn and process political belief and knowledge.

Chapter 3 tackles the situational and definitional process by inspecting extreme response as both situational, as an independent indicator of truncation of thought and a satisficing answer. Therefore I might plausibly use this variable as an indicator of engagement with the subject matter, and possibly indicative of issue importance. To this end, an exploratory analysis of five election years in the ANES, reveals some regularities that we might attribute to how we might think differently about political issues. First and foremost, there is a marked difference between issue clusters such as Defense and Spending and Services, which demonstrate stronger relationship to ERS (in the analogue to factor loading), and Abortion and issues evoking Traditional Values, which are more muted in how they evoke ERS. All this says on its face is that respondents are choosing endpoints more for Defense and Spending and Services, and less Abortion and issues evoking Traditional Values, for no discernible and detectable reason.

Yet if we entertain these measures as meaningful, we might take the next step and entertain the notion that Abortion and issues evoking Traditional Values are more well-considered, more engaging, and less ambiguous than Defense or Spending and Services. Making this small leap of inference highlights something that may be uncomfortable in popular dialogue, that we may be divided in our thinking about cultural issues, but we seem to be taking them less lightly than other categories of issues. Or, more pointedly, those who are more prone to be less engaged on some issues, do so less frequently in matters of cultural values. As Carmines and Stimson note, cultural values are not difficult to process. They are "easy," and the easy moniker appears to come through in response style, easy enough that those who do not otherwise show nuanced response are capable of tackling the material in a nuanced way. And not only are they easy, but they show consistent easiness over time.

The temptation to call issues "easy" or "difficult" is alluring. It fits a notion of political awareness and lack of engagement. Certainly abortion as an issue is symbolic, engaging, popular enough and specific enough that it will cause more thought to occur. Yet this "easiness" is not transferred to the symbolic and widely-known partisan and ideological labels. In that way, engaging, lack of thought, importance, and priority do not fit well as descriptors. *However, of the descriptors that do fit the behavior of extreme response style, and how it agrees with these results is a certain connotation of deliberation.* Partisan and ideological objects, by this standard are not "low priority" or "low importance," but certainly not deliberately examined.

(d) Extensions. ARS. The rest of this thesis explored other avenues and interpretations of response style, where those styles might have meaning in their respective survey contexts. Chapter 4's distillation and discussion of Acquiescent Response Style is a bit different since it

involves a different kind of default response and the measured political trait is in a different analytical space: confidence in several institutions in society and whether partisanship helps provide a rationale for disagreement. Its contribution to this thesis is another possible way of looking at response style and cognition. In this case, similar to extreme response, it can be said that knowledge and engagement play a key role. This time, however, more knowledge and engagement is facilitated by partisan affiliation and helps gravitate responses away from the default answer of acquiescence to consider and alternate, negative response. All else equal, a default, positive response is resisted under conditions where respondents have accumulated knowledge, opinion, or engagement with the object/institution in question. Partisanship, at a very low level, facilitates, where appropriate, transition to a negative response just as knowledge facilitates transition to more nuanced indication of opinion in the ERS case.

From ARS, we can see a number of similarities among response styles. First, there seems to be a strong pull to a default response in both, and some demographic characteristics overlap (gender being a notable exception with ARS). What appears different is the kind of thinking it appears to take to dislodge a respondent from this default location. For ARS, partisanship seems to be that motor that drives a respondent out of a default response, into considering negative judgment. Oddly enough, however, ERS requires more than partisanship to dislodge, where partisan feelings appear to be a strong indicator of residual endpoint selection. The possibility exists that there are several steps, both a pre-awareness of a survey stimulus, and nuance, a level of cognitive development with the stimulus in order to extract nuanced response.

(e) Extensions. Constraint and Moderation. Chapters 5 and 6 take the interpretation from chapter 3, exploring opinions and names ERS specifically in order to explore:

- whether acquired knowledge at a young age is more indicative of a knowledge and acceptance of where political divisions lie, a more critical examination of these opinions lie, and what kind of instruction appear to facilitate these traits, and
- whether acquired knowledge in a religious division is more indicative of a knowledge and acceptance of where political divisions lie, a more critical examination of these opinions lie, and for which sides of the political spectrum.

The names I give are "constraint" and "moderation," where constraint is the degree to which a respondent fits a measured trait of ideology, which in these cases is political ideology of two types. Moderation, then, building upon observations in chapter 3, is what we might call a tendency to more nuanced, deliberative, engaged opinion through ERS.

The results apparent in the IEA 8th grade dataset are informative / reiterative of a positive relationship of fact-based measurement of knowledge and both constraint and moderation. The normative conclusion is instructive if we consider that recognition of the prevailing political divide and "correct" placement of self inside this proscribed spectrum is somehow indicative of growth. Alongside this growth, then, comes a more moderated, considered, and deliberative way of thinking. It is difficult, at times, to reconcile these two findings in practice. Popular dialogue misconstrues education, at times, as somehow threatening. It also doubts that, in some way, an accumulation of knowledge will result in normatively positive outcomes or creates more extremity. The implication for opinion at a nascent age and understanding proposes a different line of thought, where adding information in the form of instruction actually transitions young learners into the popular dialogue, and does so in a way that converts uninformed opinion into informed, logically consistent, and moderated opinion. The other side of this uncomfortable tension is that we might shy away, specifically, from exposing young learners from learning

activities that are less rote, and more accepting of expressing and valuing expression of opinions in a classroom setting. On its face, the results show a mild benefit of this formal instruction using opinion expression and encouragement toward moderating opinions.

Finally, with the same ideas of constraint and moderation, a different kind of instruction might be inspected through extreme response style, one of religious cleavage. With the same kind of popular rhetoric, different sides of the ideological spectrum might characterize one another as less than deliberative in their approach to political issues. Different stereotypes emerge of right or left-leaning Evangelicals, Catholics, or non-Christians in society based on where different voting blocks are likely to be. If we formalize hypotheses as I have, where "constraint" is more likely to be evident where voters are most concentrated (to the right or left), or moderated more where they are less likely to be concentrated, perhaps more "cross-pressured," these hypotheses do not seem to explain much about religious division itself. It appears as if greater "instructional forces" are at play, or simply there are compositional biases that trump religion. At the very least, this opens the door to a more comprehensive political narrative surrounding who the unthinking masses are, where they are situated on the political spectrum, and why. The only stereotypical finding would lie on the political right, where social welfare constraint is higher. In addition to the catalogue in chapter 3, chapter 6 suggests a still greater host of possibilities when we think about the meaning of ERS, how issue positions might move together, and the locations of those political ideas for which these traits are manifest.

Put together, chapters 5 and 6 tell a story of how cognitive attributes might be distilled and used to bring cognitive evidence to bear on practical issues.

Having demonstrated only a few extensions in practical application of response style, there are several ways in which this line of research might plausibly branch in effective future work with response style generally, in particular those that are most effectively measured in mid-point selection and extreme response selection. These extensions highlight a number of practical directions: (a) methodological innovation and (b) use of ERS batteries going forward, (c) re-analyzing existing data, and (d) in public polling.

(a) Methodological Extensions. More methodological points of response style are also interesting for our disciplinary study and theorizing about the nature of cognition in politics. Measurement literature in psychology is deeply concerned with our internal, cognitive order of operations when evaluating tasks like scaled survey items, and at what point we might deviate from this order. Current discussion revolves roughly around stages of decision. That is, there is a mental branching construct that occurs which considers an initial categorization (say, Democrat, Republican, or Independent), and a second construct in which strength of feeling is communicated. For response style, one prevailing theory is that response style occurs at this second stage. Implicitly, branching constructs should still contain extreme response bias. In the first chapter, experimental evidence, throws some doubt on that theory, since branching questions are less susceptible to response style than an open scaling, Likert response option. What is happening here is not consistent with an isolated misattribution of strength of feeling. Instead, the chapter shows that this behavior is more closely aligned, theoretically, to what we might call unconscious, habitual, or automatic non-response. It is much more likely, from chapter 1, that extreme response style does not mainly occur at this second stage, but as a response to a combined question at the outset. When asked first about direction, and then

following with strength of opinion, in a second branch, the measures of extreme response style are largely flat, indicating non-existent or very weak influence of response style. If this were the end of the story, then there might be few follow-up questions to ask, yet there are some indications that there might be more nuance to these results than at the extremes. Midpoint response appears in some branching items (sometimes starkly), which may substitute one style for another, or appear to influence near or far extremes in a different way for different measures.

Some other measurement strategies may benefit the identification of those who are extreme respondents. Some of those fit the mold of branching structures, nested item response models that take into account direction and intensity. And while the analyses in this thesis do not seem to support this notion, other studies have found response style by using these strategies of parametrizations. More productive, I believe, would be another series of models that more keenly separate the process of ERS where it is meaningful, toward respondents who more completely espouse this extreme response behavior. When considering extreme response in particular, one potential strategy may aid in representing a truer version of a truncated thought process that acknowledges complexity (Samejima 2000). An example here is based on a 2PL model for a scale θ_j, where a_i and b_i are standard discrimination and difficulty parameters:

$$\Psi(\theta_j) = \frac{\exp\left[a_i(\theta_j - b_i)\right]}{1 + \exp\left[a_i(\theta_j - b_i)\right]}$$

$$P(U_{ij} = 1|\theta_j) = [\Psi(\theta\_j)]^{\xi_i}$$

The introduction of ξ_i, however, gives some interpretation to the complexity of an item, and consequently helps to delineate scale points that are more of interest and facilitates "acceleration" of effects in ERS that are not captured as well by a logistic function alone. This may facilitate (a) a better scaling approach in modeling endpoint selection as a complex trait that

is mostly relevant to practitioners at the very high end of the scale, and (b) an interpretation of ξ_i by item to look at item subprocesses and complexity. Still other popular approaches may be appropriate as well, separating clusters of respondents through latent classes.

Another branch to take from this thesis is furthering a measurement effort toward less obvious targets generally, shifting from self-reports to patterns of behavior. Extreme response is one of possibly many ways in which we might be able to use non-primary dimensions of reasoning, or difficulties in measurement, to gain purchase on questions in politics. Introducing person-fit measures starts expanding what we might know from not ascribing to a traditional left-right dimension of conflict. If we believe that political moderates are intentionally centered in a spectrum, or if we believe they are of an intentionally opinionated type which results in centrality, or if this centrality is the result of disengagement, we might be very concerned. What it means to be moderate can mean strong, conflicting views, true moderate opinions, or a way to communicate what is not important to an individual. Person-fit, therefore, may tell us what is happening of interest in that part of the population that does not, or refuses to subscribe to the outlines of contemporary conflict, and may be a more productive by-product to find that do not depend on unreliable questions.

(b) Prospective Use of ERS Batteries. For practitioners and survey researchers, the implication of response style contains some irony. If surveys are constructed with a very keen eye toward preventing response style, the use of response style as a measurement becomes a non-factor. Yet, for those surveys where response style might present itself, there is unexpected benefit, and possibly fruitful discoveries to be made. In surveys that contain otherwise unidimensional batteries of questions that contain response style, this may muddy the waters if

we consider doing away with questions such as thermometer items in the ANES. To the extent that this provides something meaningful to work with, there may be another argument to keep these otherwise suboptimal items in the same format, other than the current need for continuity in measures.

If we speculate further however, what might be the benefit of inserting a battery of questions that seek to induce response style traits in respondents? The cost-benefit calculus of such an approach may well need to be evaluated, but as it stands, it makes little sense to further introduce less than reliable measures with the sole purpose of reconstituting unreliable data into a cognitive measure. What does make sense, as an alternative, might be to augment the ANES in a way that proposes specific, limited questions which may give some information and piggyback, as it were, on existing bad items. That might consist of a limited battery, or experimental introduction, of feeling thermometers that include policy propositions. A spiraled form of overlapping feelings on issues might produce a number of insights, adjudicating between quantity of response options, subject matter, and symbolic content. Or, as in chapter 4, a set of endorsements may be proposed to which agreement is primed. Methodological research, combined with the political, may prove fruitful in furthering both our substantive understanding of cognition in politics, but provide referenda on the psychology of response. It happens that, in the specific subject of politics, the structural competition and ambiguity present may be an unexpectedly fertile environment for researching response modes under uncertainty. Another proposition that these data might support is an oversample of certain groups of respondents by observable characteristic. Specifically those interviewees who are less aware, have lower educational attainment / incomes, or otherwise might engage in problematic response behavior.

(b) Retrospective Use of Existing ERS Batteries. Even though future prospects for new response style data may not be expanding overall, historical data provide enough material to refine and explore new models and delve into different areas. Chapter 3 highlights that there is room presently for a more thorough catalogue of response style than can be tackled in a handful of exploratory models. A few immediate options are available for extending research. In particular, I have looked at modeled estimates of ERS on one dimension, when a more fluid analysis may be interactive. I tackle ERS *given a politically moderate respondent*, but more interesting observations may come from tacking that respondent right of left, or inspecting the "surface" of extreme response on two dimensions. For each of these instances of Differential Item Functioning (each issue, candidate, or group), then, a whole new set of hypotheses may emerge. Another strategy may be to "anchor" certain issues over time, which we might call theoretically constant between years, and use the power of the entire two decade sample toward a unified analysis of ERS between issues; the opposite proposition, too, might consider leveraging issue constraints toward finding, and documenting, changes over time with greater power. Chapter 3 is a "jumping off" point for exploration, with branches I have demonstrated in chapters 5 and 6.

Another set of ideas that might benefit from the introduction, capture, and independent use of stylistic response is that of *political distance*. In scales, we estimate and describe the contours of political debate and place people in frameworks, as I have, which define their overall ideological and partisan positions. Knowing that scales are likely biased where response style exists, it is a first step to wonder what biases it contains, and whether it is meaningful where people have response style. Are voters conflicted, or are they hardened, for which ideas? Given the question of polarization, and the susceptibility of scales to bias, can we tell if the weight of

deliberative versus non-deliberative opinion is actually, or more superficially separated. And even further afield, if extremity is non-deliberated opinion, do campaigns and media messages create the conditions, or entrench existing conditions, that encourage the formation of opinion (chapter 4), but not its deliberation?

Some projects are simply too applicable to response style not to follow, and with political distance, there is a swath of low-hanging fruit in this regard. Without having to search too far for an application, Paul Sniderman and Edward Stiglitz in *The Reputational Premium* (2012) explore exactly the use of scaling and feeling in a partisan environment with an experimental dataset. And, while each specific case or model of style might require special constraints and adjustments, isolating political distance for those who have little information is central and identifiable. In chapter 4, acquiescence appeared to diminish with partisanship, while in chapter 3 extremity appears to increase with partisanship. Experimental and ERS research may intersect here again to get to the bottom of this kind of empirical dilemma, informing how we see party and ideology, and further defining the meaning of response styles.

(d) Polling and Public Opinion. In an "honorable mention" among applications for response research, it may be impossible to escape what response style might mean for what is perhaps the most high profile concern of vocational political scientists: public polling. No doubt there are polls that will appear a year, months, weeks, or days in advance of any meaningful voting decisions that communicate to candidates and the public who is ahead, behind, strong, weak, and with which segments of the population. An open secret about these polls is that they are based in populations that prospectively describe who is likely to vote: typically registered

voters or likely voters. By linking opinion with action, extreme response tells us that there is a class of respondents who believe in things strongly, yet are not any more likely to vote. These may, in fact, be more likely to be moderates, the sacred prize in any campaign which often houses the median voter. These may also be, in fact, be more likely to be among those who have less information or have not given matters serious thought, and possibly susceptible to influence. Whether there is a workable practical effect in gleaning response style in a structured way, it is hard to tell. The best guess from this analysis is that by not accounting for whatever response style reflects, whether it is deliberativeness or goes by another name, there is probable bias in likely voter models, and that bias is probably translated to vote choice as well. I have demonstrated that this bias lives partly outside of basic demographic covariates, and while small, is consistent.

Whether in polling, or simply part of a broader survey effort, we ask the public questions, and our survey samples tell us what they think. Some of these people in our sample are thoughtful and resourceful, deliberative, and obedient to the survey construct. Others less so, yet they count the same in our survey results, and we take them seriously, and more seriously than they might deserve depending on context. It is true that in a democracy, one person, one vote rules the day. In this light, it is difficult to reconcile that with the survey environment, where one deliberative statement is not equal to a non-deliberative one, and one opinion, one strong opinion, is not equal to another. In ideological and partisan political reasoning, sometimes what we are most interested in is not the sample universe of equal people, but a sampling of discourse and weight of opinion itself, and subsequently the weight of action that follows.

References

Achen, C. H. 1975. Mass Political Attitudes and the Survey Response. *The American Political Science Review*, 1218-1231.

Adams, Greg D. 1997. Abortion: Evidence of an Issue Evolution. *The American Journal of Political Science*, 41(3): 718–737.

Althaus, S. L. 1998. Information Effects in Collective Preferences. *American Political Science Review*, 92(3):545-558.

Alvarez, R. M. 1998. Information and Elections. University of Michigan Press.

Ansolabehere, S., J. Rodden, and J. M Snyder. 2008. The Strength of Issues: Using Multiple Measures to Gauge Preference Stability, Ideological Constraint, and Issue Voting. *American Political Science Review*, 102(02): 215–232.

Bachman, Jerald G., & Patrick M. O'Malley. 1984. Yea-Saying, Nay-Saying, and Going to Extremes: Black-White Differences in Response Styles. *Public Opinion Quarterly*, 48(2): 491–509.

Baldassarri, D., & Gelman, A. 2008. Partisans without Constraint: Political Polarization and Trends in American Public Opinion. *AJS; American Journal of Sociology*, 114(2), 408–446.

Banaji, M. R., K. M Lemm & S. J. Carpenter. 2001. Automatic and Implicit Processes in Social Cognition. Blackwell Handbook of Social Psychology: Intraindividual Processes p. 134-158.

Bargh, J. A & M. J Ferguson. 2000. Beyond Behaviorism: On the Automaticity of Higher Mental Processes. *Psychological Bulletin*, 126(6):925945.

Bargh, J. A & T. L Chartrand. 1999. The Unbearable Automaticity of Being. *American Psychologist*, 54(7):462-479.

Bargh, J. A. 1997. "The Automaticity of Everyday Life." Advances in Social Cognition 10: 1–61.

Bartels, L. M. 1998. Electoral Continuity and Change, 1868-1996. *Electoral Studies*, 17(3):301-326.

Bassili, J. N. 1996. Meta-Judgmental Versus Operative Indexes of Psychological Attributes: The Case of Measures of Attitude Strength. *Journal of Personality and Social Psychology*, 71:637-653.

Baumgartner, Hans & Jan-Benedict E. M. Steenkamp. 2001. Response Styles in Marketing Research: A Cross-National Investigation. *Journal of Marketing Research*, 38(2):143-156.

Berelson, B. 1954. Voting: A Study of Opinion Formation in a Presidential Campaign. University of Chicago Press.

Berg, I. A., and J. S. Collier. 1953. Personality and Group Differences in Extreme Response Sets. *Educational and Psychological Measurement*, 13(2): 164.

Billiet, J. B & E. Davidov. 2008. Testing the Stability of an Acquiescence Style Factor Behind Two Interrelated Substantive Variables in a Panel Design. *Sociological Methods & Research*, 36(4):542.

Bock, Darrell R. 1972. Estimating Item Parameters and Latent Ability when Responses are Scored in Two or More Nominal Categories. *Psychometrika*, 37(1): 29–51.

Böckenholt, Ulf. 2012. Modeling Multiple Response Processes in Judgment and Choice. *Psychological Methods*, 17(4): 665-678.

Bohrnstedt, G. W. 1983. Measurement. Handbook of Survey Research. 69-121. Academic Press.

Bolt, D. M, and T. R Johnson. 2009. Addressing Score Bias and Differential Item Functioning Due to Individual Differences in Response Style. *Applied Psychological Measurement*, 33(5): 335.

Bolt, Daniel M. and Newton, Joseph R. 2012. Multiscale Measurement of Extreme Response Style. *Educational and Psychological Measurement*, 71: 814-833.

Brady, Henry E. 1990. Traits versus Issues: Factor versus Ideal-Point Analysis of Candidate Thermometer Ratings. *Political Analysis* 2(1): 97-129.

Brady, Henry E., and Paul M. Sniderman. 1985. Attitude Attribution: A Group Basis for Political Reasoning. *The American Political Science Review*, 79(4): 1061-1078.

Bromley, D. W. 2006. Suffcient Reason: Volitional Pragmatism and the Meaning of Economic Institutions. Princeton University Press.

Carlson, D. 2012 Chapter 3. The Effect of Educational Policies, Practices, and Context on Civic Knowledge and Skills, from unpublished doctoral dissertation *Out of the Classroom and into the Voting Booth? Analyzing the Effects of Education on Political Participation*, University of Wisconsin-Madison.

Carmines, E. G. & Stimson, J. A. 1980. The Two Faces of Issue Voting. *The American Political Science Review*, 74(1), 78–91.

Carmines, Edward G. & James K. Stimson. 1986. On the Structure and Sequence of Issue Evolution. *The American Political Science Review*, 80(3): 901-920.

Catterberg, G. & A. Moreno. 2006. The Individual Bases of Political Trust: Trends in New and Established Democracies. *International Journal of Public Opinion Research*, 18(1):31.

Chen, C., S. Lee, & H. W Stevenson. 1995. Response Style and Cross-cultural Comparisons of Rating Scales among East Asian and North American Students. *Psychological Science*, 6(3): 170–175.

Cheung, G. W & R. B Rensvold. 2000. Assessing Extreme and Acquiescence Response Sets in Cross-cultural Research Using Structural Equations Modeling. *Journal of Cross-Cultural Psychology*, 31(2): 187.

Chun, K. T, J. B Campbell, and J. H Yoo. 1974. Extreme Response style in Cross-cultural Research: A Reminder. *Journal of Cross-Cultural Psychology,* 5(4): 465.

Cialdini, R. B. 1984. Influence: How and Why People Agree to Things. William Morrow & Co.

Claggett, William J. M. & Byron E. Shafer. 2010. The American Public Mind: The Issues Structure of Mass Politics in the Postwar United States. Cambridge University Press.

Clarke III, I. 2001. Extreme Response Style in Cross-cultural Research. *International Marketing Review*, 18(3): 301–324.

Converse, Philip E. 1964. The Nature of Belief Systems in Mass Publics. Chapter in *Ideology and Discontent*. London: Free Press of Glencoe pp. 206-261.

Cronbach, L. J. 1946. Response Sets and Test Validity. *Educational and Psychological Measurement*, 6: 475-494.

Cronbach, L. J. 1950. Further Evidence on Response Sets and Test Design. *Educational and Psychological Measurement*, 10:3-31.

de Jong, Martijn G, Jan-Benedict E.M Steenkamp, Jean-Paul Fox, and Hans Baumgartner. 2008. Using Item Response Theory to Measure Extreme Response Style in Marketing Research: A Global Investigation. *Journal of Marketing Research*, 45(1): 104–115.

Delli-Carpini M, Keeter S. 1996. What Americans Know About Politics and Why it Matters. New Haven, CT: Yale Univ. Press

Downs, Anthony. 1957. An Economic Theory of Democracy. New York.

Drasgow, F., Levine, M. and Williams, E. 1985. Appropriateness Measurement with Polychotomous Item Response Models and Standardized Indices. *British Journal of Mathematical and Statistical Psychology*, 38: 67–86.

Eagly, A. H & S. Chaiken. 1993. The Psychology of Attitudes. Harcourt Brace Jovanovich College Publishers Fort Worth, TX.

Emons, Wilco H. M., Klaas Sijtsma, and Rob R. Meijer. 2005. Global, Local, and Graphical Person-Fit Analysis Using Person-Response Functions. *Psychological Methods,* 10 (1): 101-109.

Erikson, R. S., M. B. MacKuen & J. A. Stimson. 2003. The Macro Polity. Cambridge University Press, New York

Fabrigar, L. R & J. A Krosnick. 1995. Attitude Importance and the False Consensus Effect. *Personality and Social Psychology Bulletin,* 21:468.

Fiorina, M. P., Abrams, S. J., & Pope, J. C. 2005. *Culture War?* New York: Pearson Longman.

Fishkin, J. S. 1997. The Voice of the People: Public Opinion and Democracy. Yale University Press.

Franklin, C. H. 1991. Eschewing Obfuscation? Campaigns and the Perception of US Senate Incumbents. *The American Political Science Review,* 85(4):1193-1214.

Gimpel, James G., J. Celeste Lay, and Jason E. Schuknecht. 2003. Cultivating Democracy: Civic Environments and Political Socialization in America. Washington, D.C.: Brookings Institution Press.

Green, D. P., B. Palmquist & E. Schickler. 2002. Partisan Hearts and Minds: Political Parties and the Social Identities of Voters. Yale Univ Press.

Green, Donald P. & Jack Citrin. 1994. Measurement Error and the Structure of Attitudes: Are Positive and Negative Judgements Opposites? *American Journal of Political Science*, 38:256-81.

Greenleaf, E. A. 1992a. Improving Rating Scale Measures by Detecting and Correcting Bias Components in Some Response Styles. *Journal of Marketing Research,* 29(2): 176–188.

Greenleaf, Eric A. 1992b. Measuring Extreme Response Style. *The Public Opinion Quarterly,* 56(3): 328-351.

Hibbing, J. R & E. Theiss-Morse. 2001. What is it About Government that Americans Dislike? Cambridge Univ Pr.

Hillygus, D. Sunshine. 2005. The Missing Link: Exploring the Relationship between Higher Education and Political Engagement. *Political Behavior*, 27(1): 25-47.

Höfler M, Pfister H, Lieb R, Wittchen H. 2005. The Use of Weights to Account for Non-Response and Drop-out. *Social Psychiatry and Psychiatric Epidemiology*, 40:291–299.

Holbrooke, Allyson L., Jon A. Krosnick, Penny S. Visser, Wendy L. Gardner, John T. Cacioppo. 2001. Attitudes Toward Presidential Candidates and Political Parties: Initial Optimism, Inertial First Impressions, and a Focus on Flaws. *American Journal of Political Science*, 45(4):930-950.

Hui, C. H & H. C Triandis. 1989. Effects of Culture and Response Format on Extreme Response Style. Journal of Cross-Cultural Psychology 20(3): 296.

Hunter, J. D. 1992. Culture Wars: The Struggle to Define America. Basic Books.

Jackman, S., and P. M Sniderman. 2002. The Institutional Organization of Choice Spaces: A Political Conception of Political Psychology. *Political Psychology*, 209–224.

Jackson, D.N. & S. Messick. 1958. Content and Style in Personality Assessment. *Psychological Bulletin*, 55:243-252.

Jacoby, W. G. 2002. Liberal-Conservative Thinking in the American Electorate. In Research in Micropolitics: Political Decision Making, Participation, and Deliberation, ed. Michael X. Delli-Carpini, Leonie Huddy, and Robert Y. Shapiro. Greenwich, CT: JAI Press, 6: 97–147.

Johnson, T., P. Kulesa, I. Llc, Y. I Cho & S. Shavitt. 2005. The Relation Between Culture and Response Styles: Evidence from 19 Countries. *Journal of Cross-Cultural Psychology*, 36(2):264.

Kaplan, D. 2010. A Bayesian Perspective on Methodologies for Drawing Causal Inferences in Experimental and Non-Experimental Settings, Society for Research on Educational Effectiveness, Washington, DC.

Kieruj, N. D, and G. Moors. 2010. Variations in Response Style Behavior by Response Scale Format in Attitude Research. *International Journal of Public Opinion Research*, 22(3): 320.

Kinder, D. R. 1982. Enough Already about Ideology: The Many Bases of American Public Opinion. In A paper presented at the Annual Meeting of the American Political Science Association. Denver, Colorado.

Krosnick, Jon A. & Howard Schuman. 1988. Attitude Intensity, Importance, and Certainty and Susceptibility to Response Effects. *Journal of Personality and Social Psychology*, 54(6):940-952.

Krosnick, Jon A. & Matthew K. Berent. 1993. Comparisons of Party Identi_cation and Policy Preferences: The Impact of Survey Question Format. *American Journal of Political Science,* 37(3):941-964.

Krosnick, Jon A. 1991. The Stability of Political Preferences: Comparisons of Symbolic and Nonsymbolic Attitudes. *American Journal of Political Science,* 35(3):547-576.

Krosnick, Jon A. 1999. Survey Research. *Annual Review of Psychology,* 50:537-67.

Krugman, Paul. The Ignorance Caucus. The New York Times. February 10, 2013. http://www.nytimes.com/2013/02/11/opinion/krugman-the-ignorance-caucus.html?_r=0

Langton, K.P. & M. K. Jennings. 1968. Political Socialization and the High School Civics Curriculum in the United States. *American Political Science Review*, 62(3): 852-867.

Lau, R. R. & D. P. Redlawsk. 2001. "Advantages and Disadvantages of Cognitive Heuristics in Political Decision Making." *American Journal of Political Science,* 951-971.

Layman, Geoffrey C. & Thomas M. Carsey. 2002. Party Polarization and 'Conflict Extension' in the American Electorate. *American Journal of Political Science*, 46(4): 786–802.

Levine, M. & Rubin, D. 1979. Measuring the Appropriateness of Multiple-Choice test Scores. *Journal of Educational Statistics*, 4:269–290.

Lewis, N. A., & J. A. Taylor. 1955. Anxiety and Extreme Response Preferences. *Educational and Psychological Measurement*, 15(2): 111.

Linn, R. L. 2006. The Standards for Educational and Psychological Testing: Guidance in Test Development. *Handbook of Test Development*, 27–38.

Little, RJA & Rubin D, editors. 2002. *Statistical Analysis with Missing Data*. New Jersey, NJ: Wiley

Lord, F. 1952. A Theory of Test Scores. Psychometric Monographs.

Luskin, Robert C. 1987. Measuring Political Sophistication. *American Journal of Political Science,* 31:856-99.

Marin, G., R. J Gamba & B. V Marin. 1992. Extreme Response Style and Acquiescence among Hispanics: The Role of Acculturation and Education. *Journal of Cross-Cultural Psychology,* 23(4):498.

Martin, Andrew D. & Kevin M. Quinn. 2002. Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953-1999. *Political Analysis*, 10(2):134-153.

McGee, R. K. 1962. The Relationship between Response Style and Personality Variables: I. The Measurements of Response Acquiescence. *Journal of Abnormal and Social Psychology,* 64(3):229-233.

Meijer, R. & Sijtsma, K. 2001. Methodology Review: Evaluating Person Fit. *Applied Psychological Measurement*, 25:107–135.

Meisenberg, G. & A. Williams. 2008. Are Acquiescent and Extreme Response Styles Related to Low Intelligence and Education? *Personality and Individual Differences*, 44(7):1539-1550.

Messick, S. 1991. Psychology and Methodology of Response Styles. In Improving Inquiry in Social Science. A volume in honor of Lee J. Cronbach: 161-200.

Miller, Joanne & David A. M. Peterson. 2004. "Theoretical and Empirical Implications of Attitude Strength" *Journal of Politics*, 66 (3): pp. 847-67.

Mondak, J. J & M. R Anderson. 2008. The Knowledge Gap: A Re-examination of Gender-Based Differences in Political Knowledge. *The Journal of Politics,* 66(2):492-512.

Moors, G. 2003. Diagnosing Response Style Behavior by Means of a Latent-Class Factor Approach. Socio-demographic Correlates of Gender Role Attitudes and Perceptions of Ethnic Discrimination Re-examined. *Quality and Quantity,* 37(3): 277-302.

Moors, G. 2004. Facts and Artifacts in the Comparison of Attitudes Among Ethnic Minorities. A Multigroup Latent Class Structure Model with Adjustment for Response Style Behavior. *European Sociological Review*, 20(4):202-320.

Muthen, B. O. 2002. Beyond SEM: General Latent Variable Modeling. *Behaviormetrika*, 29(1):81-117.

Nevitte, N. & M. Kanji. 2003. Authority Orientations and Political Support: A Crossnational Analysis of Satisfaction with Governments and Democracy. *International Studies in Sociology and Social Anthropology*, 157-182.

Niemi, R.G., and Jane Junn. 1998. Civics Education: What Makes Students Learn. New Haven, CT: Yale University Press.

Paul, Richard. 1990. Critical Thinking: What Every Person Needs to Survive in a Rapidly Changing World. Rohnert Park, CA: Center for Critical Thinking and Moral Critique.

Paulhus, Delroy L. 1991. Measurement and the Control of Response Bias. Chapter in Measures of Personality and Social Psychological Attitudes. San Diego, CA: Academic Press.

Poole, Keith T. and Howard Rosenthal. 1985. A Spatial Model For Legislative Roll Call Analysis. *American Journal of Political Science,* 357-384.

Poole, K. T. 2005. The Evolving Influence of Psychometrics in Political Science. *SSRN 1154066*.

Popkin, S. L. 1994. The Reasoning Voter. University of Chicago Press, Chicago.

Ray, J. J & L. V Still. 1987. Maximizing the Response Rate in Surveys may be a Mistake. *Personality and Individual Differences*, 8(4):571-573.

Ray, J. J. 1979. Is the Acquiescent Response Style Problem not so Mythical After All? Some Results from a Successful Balanced F Scale. *Journal of Personality Assessment*, 43(6):638-643.

Ray, J. J. 1983. Reviving the Problem of Acquiescent Response Bias. *Journal of Social Psychology*, 121(1):81-96.

Reise, S. 1990. A Comparison of Item- and Person-fit Methods of Assessing Model-data Fit in IRT. *Applied Psychological Measurement*, 14, 127–137.

Robins JM, Rotnitzky A, Zhao LP. 1994. Estimation of Regression Coefficients when Some Regressors are not Always Observed. *Journal of the American Statistical Association*, 89:846-866.

Rorer, L. G. 1965. The Great Response-Style Myth. *Psychological Bulletin*, 63(3): 129–156.

Samejima, F. 2000. Logistic Positive Exponent Family of modls: Virtue of Asymmetric Item Characteristic Curves. *Psychometrika*, 65:319-335.

Schaefer, N. C & S. Presser. 2003. The Science of Asking Questions. *Annual Review of Sociology*, 29:65-88.

Shadish, W. R, T. D Cook, and D. T Campbell. 2002. "Experimental and quasi-experimental designs for generalized causal inference.".

Shafer, B. E & W. J.M Claggett. 1995. The Two Majorities: the Issue Context of Modern American Politics. Johns Hopkins Univ Press.

Siegel, Harvey. 1988. Educating Reason: Rationality, Critical Thinking, and Education. New York: Routledge.

Smith, P. B. 2004. Acquiescent Response Bias as an Aspect of Cultural Communication Style. *Journal of Cross-Cultural Psychology*, 35(1):50.

Sniderman, P. M., R. A. Brody, & P. Tetlock. 1993. Reasoning and Choice: Explorations in Political Psychology. Cambridge Univ Press.

Sniderman, Paul M., and Edward H. Stiglitz. 2012. The Reputational Premium: A Theory of Party Identification and Policy Reasoning. Princeton University Press.

Stone, D. A. 1989. Causal Stories and the Formation of Policy Agendas. *Political Science Quarterly*, 281-300.

Suh, Y., and D. M. Bolt. Nested Logit Models for Multiple-Choice Item Response Data. *Psychometrika*, 1-20.

Tourangeau, R. & K. Rasinski. 1988. Cognitive Processes Underlying Context Affects in Attitude Measurement. *Psychological Bulletin*, 103:299-314.

Treier, S. & S. Hillygus. 2009. The Nature of Political Ideology in the Contemporary Electorate. *Public Opinion Quarterly*,73 (4): 679-703.

Treier, S. & S. Jackman. 2008. Democracy as a Latent Variable. *American Journal of Political Science*, 52(1):201.

Vanbeselaere, C. 2005. The Shirking Model [microform]: a Theory of how People Answer Survey Questions. University Micro_lms International.

Verba, S., K. L. Schlozman & H. E. Brady. 1995. Voice and Equality: Civic Voluntarism in American Politics. Harvard University Press.

Verba, Sidney & Norman H. Nie. 1987. Participation in America. The University of Chicago Press.

Vermunt, J. K, and J. Magidson. 2004. LATENT GOLD® 4.0. User's Guide. Belmont, MA: Statistical Innovations.

Vermunt, J. K, and J. Magidson. 2005. "Technical Guide for Latent GOLD Choice 4.0: Basic and Advanced." Belmont MA: Statistical Innovations Inc.

Weijters, B. 2006. Response Styles in Consumer Research. Unpublished doctoral dissertation, University of Ghent, Belgium.

Wells, W.D. 1961, The Influence of Yea-Saying Response Style. *Journal of Advertising Research*, 1(4):1-12.

Zaller, J. 1996. The Myth of Massive Media Impact Revived: New Support for a Discredited Idea. *Political Persuasion and Attitude Change*, (19): 17-78.

Zaller, John A. 1992. The Nature and Origins of Mass Opinion. Cambridge University Press.

Zaller, Jon & Stanley Feldman. 1992. A Simple Theory of the Survey Response: Answering Questions versus Revealing Preferences. *American Journal of Political Science,* 36(3):579-616.