Ambivalence is Everywhere: Quantifying Opinion Stability Across Topic Domains[*]

Kevin Kiley, Duke University[†]

1/13/2020

Abstract

A major debate in the social sciences centers on whether people hold consistent attitudes over time or whether attitudes are temporary constructs. A middle ground suggests that stable opinions are a function of social structure and attention, and on any particular issue, some people hold stable attitudes and others do not. This paper uses a finite mixture model approach to quantify the proportion of people who hold stable attitudes, the proportion of people who make durable changes, and the proportion of people who demonstrate vacillating or ambivalent responses for more than 500 survey questions across 10 panel data sets. The results suggest wide variation across questions in the proportion of respondents who hold stable attitudes, with most subject areas demonstrating high levels of inconsistency. Stability is also socially patterned, with people demonstrating stability on related issues, suggesting that over-time instability in responses is not measurement error, but that the general public is divided into "issue publics" that have stable opinions on different issues. Rather than argue that people in general hold or lack opinions, the results show that stable opinions are socially contingent.

1 Introduction

What proportion of the population holds stable attitudes? This question has been a key debate in the social sciences since Philip Converse claimed that "large portions of an electorate … simply do not have meaningful beliefs, even on issues that have formed the basis for intense political controversy among elites for substantial periods of time" (Converse 1964). After decades of research and debate, there is little consensus.

On one hand, researchers point to low wave-to-wave correlations in people's responses to the same question over time (Converse 1964), question-order and -wording effects (Perrin and McFarland 2011), the effects of psychological primes, the complexity of culture, and the general limitations of human cognition (Martin 2010) to argue that people do not carry around fixed beliefs. Instead, they argue, people are ambivalent on issues of public interest and construct opinions in the interview or survey setting, drawing on recent considerations from their social environment (J. R. Zaller 1992). In contrast, other researchers point to the stability of attitudes in the aggregate, both aggregating across questions within people (Ansolabehere, Rodden, and Snyder 2008) and

---

aggregating the same question across a population (Inglehart 1985; Page and Shapiro 1992); the infrequency of durable changes in attitudes over time (Kiley and Vaisey 2020; Vaisey and Lizardo 2016); and the ability of specific attitudes to predict a range of behaviors over time (Vaisey 2009; Vaisey and Lizardo 2010) as evidence that people carry at least some stable latent attitudes or dispositions, even if these positions are obscured by measurement error.

Recent work argues that dispositions are mostly stable by the time people reach adulthood (Kiley and Vaisey 2020; Vaisey and Lizardo 2016; Vaisey and Kiley n.d.). But these findings elide the "non-attitudes" debate by grouping together people who hold stable, unchanging opinions with people who change randomly from wave to wave. In other words, while finding that people tend not to make durable changes in their beliefs over time, they cannot say whether this is because people hold real, stable attitudes or whether it is because they hold no attitudes at all. This is, to put it mildly, an important distinction.

A middle ground – going back to Philip Converse's development of the "non-attitudes" thesis – suggests that on any particular issue (and in any particular window of time) the population comprises three groups: people who hold stable opinions they can report consistently, people who hold weak opinions subject to temporary influences (Freeder, Lenz, and Turney 2019; Hill and Kriesi 2001a; J. R. Zaller 1992), and a small group of people making durable, real change. In other words, rather than assuming that all members of the population either have or lack opinions, this approach says "some do and some do not." But for any particular question, these groups will vary in size and composition. There has been little systematic exploration of which issues demonstrate more or less stability or durable change, especially outside of political attitudes. This has hindered the development of systematic perspectives about the degree to which people form opinions and the social conditions under which they do.

This debate about whether people hold stable attitudes has significant implications for the role of attitudes and beliefs in social behavior. If people do not carry around fixed beliefs, if their dispositions and behavior are swayed by temporary influences, there is substantial room for contemporary social structures, opinion leaders, and momentary situations to shape attitudes, and explanations for attitudes should be rooted in local social structures, as pragmatist theories suggest (Gross 2009; Joas 1996). If people hold stable attitudes that are relatively impervious to momentary changes in social influence, these attitudes would be more likely to shape patterns of behavior, affiliation, and belief over time, and the explanations for attitudes should be located in the past, as recent work suggest. A world where, on any particular issue, some people hold a stable attitude and others do not directs attention to "institutions and contexts and other forms of objectified cultural structure" that facilitate attitude stability (Lizardo and Strand 2010: p. 206; Martin 2010) and suggests that different attitudes might matter in explaining different peoples' behavior.

In this paper, I use a finite mixture model approach to estimate for more than 500 attitude questions across 10 panel data sets the proportion of respondents who hold stable opinions, the proportion of respondents who hold vacillating attitudes or no opinion at all, and the proportion of respondents who make durable changes of opinion. These questions include topics addressed by political scientists in the past, as well sociologically relevant questions about religion, gender roles, race relations, morality, institutional and social trust, and more.

Across these questions, I find wide variation in the proportion of people who hold stable opinions, with some attitude questions demonstrating widespread response stability across the population (70 percent of respondents or more) while other questions demonstrate rampant inconsistency, with fewer than 20 percent of respondents able to give a consistent opinion over time. Overall, rates of vacillating change tend to exceed rates of stable attitudes, and ambivalence can be found in attitudes about political issues, religious and moral beliefs, self-assessments, sentiment toward groups, and more. People disproportionately demonstrate stability on related issues (e.g., stability on one political issue predicts stability on other political issues, but not on religious issues), suggesting different forces give rise to stability in different domains. Rather than argue that people in general "have" or "lack" opinions, these results suggest both are generally true for any particular question, and they reinforce the argument that stable opinions are principally a function of attention and social structure.

2 Stable Opinions and Non-Attitudes

Broad streams of social science work generally assume that people have relatively stable attitudes or dispositions, which is why Converse's original "non-attitudes" findings generated such a robust debate.

The most sustained challenge to the "non-attitudes" model in political science are what are called "measurement-error" models, which posit that all people hold (latent) attitudes but report them with some error (Achen 1975; Ansolabehere, Rodden, and Snyder 2008; Inglehart 1985). Error arises for a variety of reasons, but when responses to the same question are aggregated at the population level (Page and Shapiro 1992), when responses to related questions are aggregated within people (Ansolabehere, Rodden, and Snyder 2008), or certain statistical methods are employed (Inglehart 1985; Judd and Milburn 1980), respondents display much higher levels of stability than when we look at their individual responses. Researchers interpret this as evidence of stable "latent" attitudes.

In cultural sociology, this line of thinking takes the form of suggesting that people have implicit dispositions that shape their behavior over time, even if they cannot articulate these commitments in interview contexts (Hitlin and Piliavin 2004; Vaisey 2009). Vaisey argues that interviews and surveys tap distinct cognitive processes. Interviews tap discursive consciousness, which tends to bring to the surface the breadth of contradictory considerations that people have internalized. Fixed-choice survey questions tend to invoke practical consciousness, gut feelings about which answer is likely correct, which also tends to control social behavior across situations. In support of his argument, he and others show that responses to survey questions about worldviews and values predict a range of behaviors across contexts (Miles 2015; Vaisey 2009; Vaisey and Lizardo 2010) and that quickly assessed questions about the relationship between cultural concepts is predictive of other beliefs (Hunzaker and Valentino 2019).

These works give us an expectation that people should be relatively consistent in their cultural commitments over time, especially in survey responses. When asked to give an opinion about an issue, such as whether they agree or disagree that "Morality is a personal matter and society should not force everyone to follow one standard," people might vacillate within a narrow band, some days agreeing that morals are a personal matter and some days strongly agreeing, but when they look at the "lineup" of options, a person with a disposition

that morality is a personal matter is not going to say that society should enforce one standard, and they are not going to say they lack a position. These results will "feel" wrong, even if people are not conscious of why.

At the same time, a key finding of cultural sociology in the last half-century is that rather than internalize a consistent cultural worldview through a single socialization process, people are exposed to and internalize a diverse, contradictory cultural repertoire of beliefs, practices, and assessments (Swidler 1986; DiMaggio 1997). A long line of research documents this cultural contradiction across domains. In America, love is both a choice entered into freely and a unique and irreplaceable commitment that people cannot leave (Swidler 2001). Morality is fixed and relative (Baker 2004), orthodox and progressive (Hunter 2000). American culture is individualist and collectivist; managerial and therapeutic; biblical and republican (Bellah et al. 1985).

It is not just that culture is diverse and contradictory at the public level – that concepts compete in the public domain but people maintain distinct worldviews – but this contradiction is also apparent in personal culture, the declarative and non-declarative attitudes, worldviews, values, and dispositions that manifest at the individual level (Lizardo 2017). Because of their cognitive limitations, people consume a broad array of cultural information and can not or tend not to engage in the effort to reconcile these cultural contradictions (Martin 2010; J. R. Zaller 1992). As a result, "our heads are full of images, opinions, and information, untagged as to truth value, to which we are inclined to attribute accuracy and plausibility" (DiMaggio 1997: p. 267).

Because they have contradictory commitments in their brains, people seem to struggle to maintain their a single line of cultural reasoning over time (Swidler 1986). In successive interviews (and often in the same interview) people can demonstrate vary different opinions on the same issue without recognizing these contradictions (Swidler 2001). People draw on different cultural resources to justify institutionally constrained behavior (Mills 1940; Scott and Lyman 1968). And when we look at people's responses to the same attitude question over time, they often demonstrate high levels of inconsistency (Alwin 2007; Converse 1964; Hout and Hastings 2016).

This work in cultural sociology echoes with a line of public opinion scholarship on how people form opinions in survey settings. In his Receive-Accept-Sample model of opinion formation, Zaller (1992; J. Zaller and Feldman 1992) argues that the population is characterized by ambivalence toward political issues and tend to be uncritical toward the messages they receive. As a result, they store arguments for "more government spending is good" and "higher taxes are bad" when we hear these messages, and, as Zaller notes, "most of the time, there is no need to reconcile or even recognize their contradictory reactions to events and issues. Each can represent a genuine feeling, capable of coexisting with opposing feelings and, depending on momentary salience in the person's mind, controlling responses to survey questions" (J. R. Zaller 1992: p. 93). When people are called to account for these beliefs, they conjure up some or all of these considerations and construct an answer on the fly. Because these considerations are called up in a haphazard way, which manifest at any particular interview can vary, and people's assessments can be influenced by question-wording, question-order, psychological primes, and changes in their informational environment, such as what was on the news recently.

These research threads suggest that we should expect people to be inconsistent in their cultural or political beliefs when measured over time. On the morality question raised earlier, this framework expects that people will sometimes they might say they lack any opinion at all. Other times, perhaps because they recently heard

a compelling argument in favor of moral absolutism, they might say they agree. Still other times, maybe after hearing a counter argument, they say might say they disagree. Regardless of what they say in the survey, they are unlikely to carry this disposition to a new setting or use it in shaping their behavior.

2.1 Some Do, Some Do Not

The debate about the stability of attitudes persists because at different times and in different contexts, both seem to be true. People sometimes behave as if they have clear, consistent political or cultural beliefs, and at other times appear to behave as if they do not. And this suggests a path forward.

Philip Converse's seminal work (1964; Converse and Markus 1979) on attitude inconsistency is frequently invoked by other researchers to suggest that the American public lacks opinions on political matters, but his "black and white" model got its name by suggesting that on any particular issue, the population could be divided into "a 'hard core' of opinion on a given issue, which is well crystallized and perfectly stable over time" and a group of people whose responses are "statistically random" (Converse 1964: p. 242). On any particular issue, some people hold opinions and others did not. The population did not consistent of ideologically distinct camps, but "issue publics" that care about different issues. Converse's model has received support over time and across data sets (Converse and Markus 1979; Taylor 1983; Hill and Kriesi 2001a), but it faces detractors who favor the measurement error approach.

The measurement error model relies on an assumption that errors are a function of the question being asked, not the people answering questions, meaning no personal features should predict instability. While initial assessments argued that nothing predicted attitude stability (Achen 1975; Erikson 1979), research since finds that individual characteristics are strongly predictive of stability in political views. Zaller (1992) shows that people who have higher political awareness – as measured by factual information about the political space – tend to hold more stable political attitudes. Freeder and colleagues (2019) find that knowledge of where political parties fall on a different issue is associated with stability over time. These findings suggest, in general, some people are better at articulating consistent political opinions than others, undermining support for the measurement-error model.

These debates have tended to be confined to political attitudes, but the theoretical underpinnings of Converse and Zaller's models are in no way confined to political beliefs. There's no reason to assume that a model in which some people hold stable opinions while others express ambivalence in the form of vacillating attitudes does not describe the behavior of a broad range of beliefs.

A "some do, some do not" of attitude consistency can address theoretical problems with "latent beliefs," specifically an unclear sense of what latent beliefs are. If latent beliefs are people's theoretical preferences – preferences for tradition over modernity; government restraint over government intervention; conservative or liberal – that people have a hard time connecting to concrete positions or debates, then people should be more consistent in answering these theoretical positions than concrete ones. But Ansolabehere and colleagues (Ansolabehere, Rodden, and Snyder 2008) find that even when aggregated, these kinds of general attitudes are less consistent than more specific kinds of beliefs such as policy positions. A world where people have a difficult time con-

necting their general dispositions to specific policies, but have a harder time answering questions that explicitly inquire about those general dispositions, is hard to square.

Other researchers posit that latent beliefs are "gut feelings" that take "a good deal of effort" to "put into words" (Inglehart 1985: p. 101). But, given what we know about different methods call forth different forms of culture, we should expect people to be better able to articulate their gut feeling in a forced-choice survey than in any other context, assuming a good enough answer is present in the lineup (Vaisey 2009). If latent beliefs are complicated preferences that are difficult to articulate in any setting, then they are unlikely to control behavior across domains, and we should question whether they are meaningful at all.

A "some do, some do not" model can explain the fact that attitudes are weakly constrained (Converse 1964; Boutyline and Vaisey 2017), but in the aggregate, people's beliefs are relatively stable (Ansolabehere, Rodden, and Snyder 2008) without recourse to latent constructs. If people tend not to hold specific attitudes, but if the attitudes they do have are socially patterned (not everybody who has an opinion on abortion and federal spending, but the people who believe abortion should be prohibited and that federal spending should be decreased), then their aggregated attitudes will exhibit low correlations at the pairwise level, but as more attitudes get aggregated, the aggregation is more likely include more stable attitudes, given the "scale" greater consistency.

A "some do, some do not" model of attitude stability, applied to attitudes more broadly, can also account for other contradictory findings in the sociology of culture. For example, attitudes can be predictive of behavior, but often weakly so. Researchers have invoked this weak predictive power to argue both for and against a model of attitude holding. A "some do, some do not" model suggests that attitudes are influential in shaping behavior when people hold them consistently, but inconsistent beliefs are not meaningful. We should not expect every person to have a position on abortion or a cultural worldview that motivates their behavior, but the people who have consistent beliefs are more likely to be influenced by them.

Finally, a "some do, some do not" model more closely aligns with work in cultural sociology suggesting that social structures – institution, contexts, and externalized public culture – are principally responsible for shaping attitude constraint (Lizardo and Strand 2010; Martin 2010; Rawlings 2020). If social reinforcement and scaffolding through public culture is a necessary condition for holding consistent beliefs, then we would not expect everybody to be equally exposed to these processes. Some people are going to be in environments where they hear contradictory messages and considerations, while other people are in environments where they hear a single line of reasoning that makes holding beliefs relatively easy (J. R. Zaller 1992).

For too long the question has been "do people have stable beliefs or do they lack stable beliefs?" That is a false dichotomy. A better question to ask is on any particular question, what proportion of people demonstrate a stable opinion? And what is the structure of "issue publics" – the co-occurrence of stable attitudes within people? For example, some people might care deeply about national political issues while others care deeply about issues of race, or local socioeconomic issues, or religious issues, or education, or something else. What each person is able to hold a stable opinion on might differ widely, or these issue publics might be socially patterned, forming a branching structure of related attitudes.

3 Varieties of Opinion Behavior

The preceding discussion suggests the question is not whether all people display stable attitudes or whether they demonstrate ambivalence, but whether we can quantify the proportion of the population falling into three groups: opinion holders, who consistently give the same opinion over time; vacillating changers, who lack stable opinions and construct them anew each survey; and durable changers, who hold opinions and change them during the course of the survey or form durable opinions during the survey. It also requires a way of adjudicating whether latent-attitude or measurement-error model could have generated the same data.

What would these three opinion behaviors look like in practice? What response patterns would be indicative of each model? To clarify, consider the question from the General Social Survey's three-wave panels: "Do you agree or disagree with the following statement: Homosexual couples should have the right to marry one another." Respondents are given five options to select from: "strongly agree", "agree", "neither agree nor disagree", "disagree", "strongly disagree."

Opinion holders would be people who report being on the same side of an issue over time, either in agreement or disagreement. A person might "strongly agree" in all three waves, or vacillate between "agree" and "strongly agree" based on recent considerations. But barring actual error in coding the response, it would be unreasonable to assume that a person who supports gay marriage would declare that they oppose it in a survey because of measurement error.[1] Similarly, we would not expect a person who actually holds an opinion on this issue to say they "neither agree nor disagree" with the question.[2]

Durable changers are people who truly hold opinions but go from one side of the issue to the other over the course of the survey – saying "disagree" or "strongly disagree" in wave 1 or wave 1 and 2 but "agree" or "strongly agree" for the remainder. Similarly, some people might truly lack an opinion on the issue and say "neither agree not disagree" in the first wave, and, upon being asked to consider the question, develop an opinion for the next wave (what is commonly called "panel conditioning" (Oh, Yeatman, and Trinitapoli 2019, @halpern-manners2012)).

If these were the only two kinds of people in the population, then we could easily count the number of each. But the presence of vacillating changers complicates the counting process. Technically, any response pattern would be compatible with a model in which individuals select responses on the basis of short-term considerations. While some patterns, such as going from "agree" to "disagree" back to "agree" over three waves would be obvious

---

[1]Measurement error theories are often unclear as to what exactly they mean by "error." Ansolabehere and colleagues attribute measurement error "vague question wordings, vague response categories or categories that do not reflect the individual's actual attitude, inattentiveness on the part of the respondent, and even typographical errors" (Ansolabehere, Rodden, and Snyder 2008: 216). While it is easy to accept that a person who supports gay marriage might vacillate between "agree" and "strongly agree" in response to this question, it is hard to imagine that a person who holds a true position in support gay marriage will give a response that indicates the opposite. There will likely be cases of "measurement error" where the response is coded incorrectly, but it is hard to believe they will be so prevalent as to produce the level of vacillation that we see in some questions.

[2]It is debatable whether the response "neither agree nor disagree" represents ambivalence and the lack of clear opinion, as I assume here, or whether it represents a nuanced position, such as one where a person thinks that homosexuals should have the right to marry in some circumstances and not others. If it is the latter, then we would expect people to maintain this position over time. However, of the 301 people who report this position in wave 1 and respond in each wave, only 31 maintain this position across all three waves. Results below further argue against this position.

indicators of vacillation, some people picking at random could end up picking on the same side of an issue three times, making them look like opinion holders, or say "agree" in wave 1 and "disagree" in waves 2 and 3, making them look like durable changers. Separating this group from the others requires a set of assumptions about how these groups behave.

Most methods tend to conflate two of these opinion behavior groups or avoid quantifying their prevalence at all. Kiley and Vaisey's (Kiley and Vaisey 2020; Vaisey and Kiley n.d.) methods compare the relative prevalence of vacillating changers and durable changers without quantifying the proportion that falls into these groups or the proportion giving stable responses. Focusing on how many people change their responses, the average level of change in responses (Freeder, Lenz, and Turney 2019), or how many people remain stable over time, especially across only two waves, conflates people who make durable change with people who vacillate.

Other methods, especially those focused on measuring correlations over time (Alwin 2007; Hout and Hastings 2016), treat questions as reflecting a continuous scale, rather than respecting the nominal structure of response options as they are presented to the survey participant. For example, many methods treat a change on a five point-point scale from "no opinion" to "disagree" the same as a change from "strongly agree" to "agree," even though the former represents a qualitative shift in belief while the latter represents a shift in degree. These methods also assume that a real position exists underneath measurement error (Achen 1975; Ansolabehere, Rodden, and Snyder 2008), which may not be a valid assumption.

4 The Basic Probability Model

Hill and Kreisi outline a finite mixture model congruent with the above example that they use to estimate the prevalence of these three kinds of response behavior to questions about pollution-abatement policies in a sample of Swiss residents (Hill and Kriesi 2001a, 2001b; Hill 2001). Here I provide an overview of the model and the general assumptions that underlie it. A more detailed explanation of the model can be found in Appendix A and in Hill and Kreisi (2001b).

Finite mixture models allow researchers to specify separate groups with different data-generating processes governed by different parameters, and then estimate the probability a particular observation came from that group (Hill and Kriesi 2001b; Imai and Tingley 2012). The fundamental challenge is that membership in each group is unknown, and many responses could reflect two of these groups. As a result, group membership has to be inferred from the respondent's behavior in the panel study and the overall distribution of response patterns. Applied to the question at hand, I can specify different data-generating processes for each behavioral group, with different parameters and quantify the probability that each process generated the observed data. A person who says "agree"-"disagree"-"disagree" has some probability of coming from the "durable change" group and some probability of coming from the "vacillating change" group, probabilities that can be determined by the behavior of the overall sample.

The starting assumption of the model is that each person is a member of one of the three groups described above for a question during the study window. While a person might be stable during the survey and make durable change at another point in time, in the survey window he can only exhibit one pattern. Let $\pi_j, j = 1, 2, 3$

represent the marginal probability of belonging to each group, with 1 indicating opinion holders, 2 indicating vacillating changers, and 3 indicating durable changers. Below I outline the separate behavioral groups and the assumptions and parameters that govern data generation for these groups.

Opinion Holders: Opinion holders maintain a consistent opinion either in agreement or disagreement with the question for the duration of the survey. For this reason, people who say "neither agree nor disagree" with the statement cannot be classified in this group. Similarly, people who cross the "opinion boundary" from agree to disagree (or disagree to agree) during the survey, cannot be included in this group. Changes between mild and strong agreement (or disagreement) are allowed.

Given these restrictions, two parameters are sufficient to describe opinion holders' responses at any wave:

$$\alpha_1 = Pr(\text{mildly agree or strongly agree}), \text{ and}$$
$$\delta_1 = Pr(\text{strongly agree} \mid \text{mildly agree or strongly agree})$$
$$= Pr(\text{strongly disagree} \mid \text{mildly disagree or strongly disagree})$$

The first parameter (α_1) describes differing probabilities of being on either side of an issue. An assumption of this model is that the probability of giving a "strong" opinion (δ_1) is equal regardless of whether a person agrees or disagrees. This assumption is not strictly necessary, and can be loosened. I include this constraint to minimize the number of parameters estimated for each question.

Vacillating Changers: The core assumption about vacillating changers is that they are not systematic in their response patterns. Unlike opinion holders, who must be on the same side in all waves, vacillating changers' responses are considered independent across waves. Unlike the other two groups where certain patterns are incompatible with category membership, any response pattern is, in theory, consistent with being a vacillating changer.

Vacillating changers' behavior is characterized by three parameters:

$$\phi_2 = Pr(\text{no opinion}),$$
$$\alpha_2 = Pr(\text{mildly agree or strongly agree}), \text{ and}$$
$$\delta_2 = Pr(\text{strongly agree} \mid \text{mildly agree or strongly agree})$$
$$= Pr(\text{strongly disagree} \mid \text{mildly disagree or strongly disagree})$$

Rather than assuming that vacillating changers select options at random, I assume the group has some minimal structure in their responses that reflects the idea that they are constructing their opinions from considerations in the social environment, and that these considerations might not be evenly distributed. They are allowed to have different probabilities of choosing the "no opinion" response (ϕ_2) and responses that indicate agreement

(α_2) – "agree" and "strongly agree" – and disagreement – "disagree" and "strongly disagree." Similar to the opinion holders, I assume that, contingent on giving an opinion, a single parameter explains the likelihood of a "strong" position (δ_2).

The parameters in this model give us an expectation for how frequently response patterns that *look* like stable opinions or durable change are actually reflective of vacillating change, given the distribution of patterns by assuming that the probability that a vacillating changer reports particular responses is consistent over time. That means for every vacillating changer who says "agree-disagree-agree" there should be one who says "disagree-agree-agree" and one who says "agree-agree-disagree." While we could not be certain which of these latter two patterns were vacillating changers and which were durable changers, the overall prevalence of these behaviors in the population, relative to the first pattern, can give us an estimate of the proportion of each group in the population.

Durable Changers: Durable changers update their beliefs in a systematic way, presumably in response to new information from the environment. From a practical perspective, the key assumption is that these respondents change their opinion one time in the study window – moving from one side of the issue to the other or moving from "no opinion" to a either agree or disagree. Respondents who do not change their stated opinion, respondents who change their stated opinion more than once in the study, and respondents who change to the "no opinion" option[3] cannot be considered durable changers.

Because these individuals have more structure associated with their responses, their behavior is described by more parameters. First, I distinguish between people who change from having "no opinion" ($\varphi_3^{(Pre)}$) and people who change between opinions ($\alpha_3^{(Pre)}$). Second, I distinguish between change that occurs after the first wave (τ_3) and change that occurs after the second wave.[4]

Third, the direction of change is allowed to change over time ($\alpha_3^{(Pre1)}, \alpha_3^{(Pre2)}$) as the balance of considerations changes. Finally, like the other groups, I include a parameter for whether people who hold opinions (agree or disagree) express strong opinions (δ_3). This results in the following set of seven parameters:

---

[3]One possibility not accounted for in this model is that someone changes from one side to "neither agree nor disagree" in wave 2 to the other side of the issue in wave 3. For this question, 47 of the 285 people who say "3" in wave 2 follow this pattern (16 percent), suggesting this pattern does not happen more than chance would expect. If these people are all counted as opinion changers, they would increase the probability of durable opinion changers slightly, but not enough to alter any major conclusions.

[4]There is some reason to expect that people with be more likely to make a durable change after wave 1 than wave 2. This is often called "panel conditioning," where participation in the survey drives people to clarify or change their opinion. If much of the durable change in attitudes is due to participation in the survey itself, then it would be hard to generalize rates of the three groups to the population.

$$\varphi_3^{(Pre1)} = Pr(\text{preswitch response is no opinion}|\text{switch occurs after 1st wave});$$

$$\varphi_3^{(Pre2)} = Pr(\text{preswitch response is no opinion}|\text{switch occurs after 2nd wave});$$

$$\alpha_3^{(Pre1)} = Pr(\text{preswitch response is mildly agree or strongly agree}|\text{switch occurs after 1st wave});$$

$$\alpha_3^{(Pre2)} = Pr(\text{preswitch response is mildly agree or strongly agree}|\text{switch occurs after 2nd wave});$$

$$\alpha_3^{(post)} = Pr(\text{postswitch response is mildly agree or strongly agree}|\text{switched from no opinion});$$

$$\delta_3 = Pr(\text{strongly agree} \mid \text{mildly agree or strongly agree})$$

$$= Pr(\text{strongly disagree} \mid \text{mildly disagree or strongly disagree}); \text{ and}$$

$$\tau_3 = Pr(\text{switch occurs after 1st wave})$$

The relative prevalence of these three behavioral groups in the respondent sample (π_1, pi_2, and pi_3) and the 12 parameters describing the behavior of individuals in these groups, are the goal of our estimation.

4.1 Estimation

Rather than estimate the model directly on people's response patterns, which is computationally impossible with a finite sample, I re-express the data as a series of general variables that describe respondents' behavior, such as the number of times they cross the "opinion boundary" and the number of "strong" responses they give. Full details on this recoding and how the model is estimated are outlined in Appendix A.

To estimate the parameters I use a Bayesian estimation procedure known as the Data Augmentation (DA) algorithm. The algorithm has two basic steps. First, it samples the "missing data" – group membership – given the estimated parameters.[5] In other words, it asks "given this person's responses, how common each group is, and how often members of the group give certain response patterns, how likely is it that this pattern came from group 1, group 2, and group 3?" It then samples from those three probabilities and assigns a group membership to that person. Then it estimates the parameters, given the group membership. It asks, for example, "given that these people are in group 1, what's the probability of someone in group 1 saying they strongly agree or strongly disagree (δ_1)?"

The process iterates until it converges on a stationary distribution that summarizes the estimated parameters given the observed data, called the posterior distribution. I use this posterior distribution to generate summary statistics, including the mean and 95 percent confidence intervals for parameter estimates. The posterior distribution reflects reasonable distributions of parameters given reasonable allocation of people to behavioral groups (and reasonable allocation of people to behavioral groups given reasonable distributions of parameter). For example, in some iterations, a response pattern of "agree-agree-disagree" might be considered a vacillating change, while in others it might be considered durable change, affecting parameter estimates.

---

[5]At the first stage, there are no group assignments with which to generate parameter estimates, so random values are drawn from the parameter space. These starting values have minimal, logical constraints (for example, the proportion of people in each behavioral group must sum to 1). To ensure that these starting draws do not unduly influence results, I wait for multiple chains to converge to the same posterior distribution before using draws to summarize the distribution.

Posterior distributions combine the distribution of the data given the unknown parameter values with a prior distribution of the parameters, which quantifies belief about the parameters before seeing any data. I use minimally informative priors outlined in Hill and Kreisi (2001b), the equivalent of adding two additional people to each opinion behavior group and splitting them up among behaviors within these groups.[6] The non-informative priors have the effect of generating slightly conservative parameter estimates. This effect is most clearly illustrated with the durable changer group. Even in iterations where nobody in the sample is assigned to the "durable change" group, which happens quite often, the model still assumes two people display this behavior, inflating the proportion estimates away from 0.

A challenge in using panel data is that people drop out of the sample or refuse to answer questions, leaving missing data. In the analyses presented here, I assume that data is Missing at Random, meaning that once I condition on a person's observed responses, response rates are random. Hill (2001) finds that the distribution of parameter estimates is similar under a variety of assumptions, with the proportion of vacillating changers slightly higher under less restrictive assumptions than the one made here. Accommodating this assumption adds an additional step to each iteration: filling in missing data on the basis of group assignment and the parameter estimates. For example, if a person is assigned to the "stable opinion" group and expresses either agree or strongly agree in their non-missing waves, their missing response is filled in from a draw of "strongly agree" and "agree" with probabilities δ_1 and $1 - \delta_1$.

The model iterates through these three steps – assign group membership on the basis of responses and estimated parameters; estimate parameters on the basis of group membership and responses; and fill in missing data on the basis of parameters, group membership, and observed data – until iterations reach a stable posterior distribution.

4.2 Gay Marriage Example

To demonstrate the model and the interpretation of its parameters, I apply it to the gay marriage question described above. This question has several features that make it a good illustration. Its structure is similar to the questions that Hill and Kreisi developed the model with: a five-point Likert scale with a midpoint that can be thought of as "no opinion" or ambivalence. Prior to and during this period, gay marriage attained a high profile in national political debates, which should ensure that some proportion of the sample had clear opinions. There were also changes in elite opinion on the issue during the window studied. As a result, I expect that some members of the sample will demonstrate durable change.

Figure 1 summarizes the posterior distribution of the gay marriage example, with the mean of the posterior distribution for each parameter indicated by a circle and error bars indicating a 95 percent confidence interval.

As a top-line result, the model estimates that between 49 and 53 percent of the sample gave stable responses to this question over three years, either agreeing ($\alpha_1 = .40$) or disagreeing with the question in all waves, while between 42 and 47 percent of the sample gave vacillating responses. Between 3.28 and 6.45 percent of

---

[6]Hill and Kreisi test a range of non-informative priors – between one and six additional people per behavioral group – and find little sensitivity to different non-informative priors.
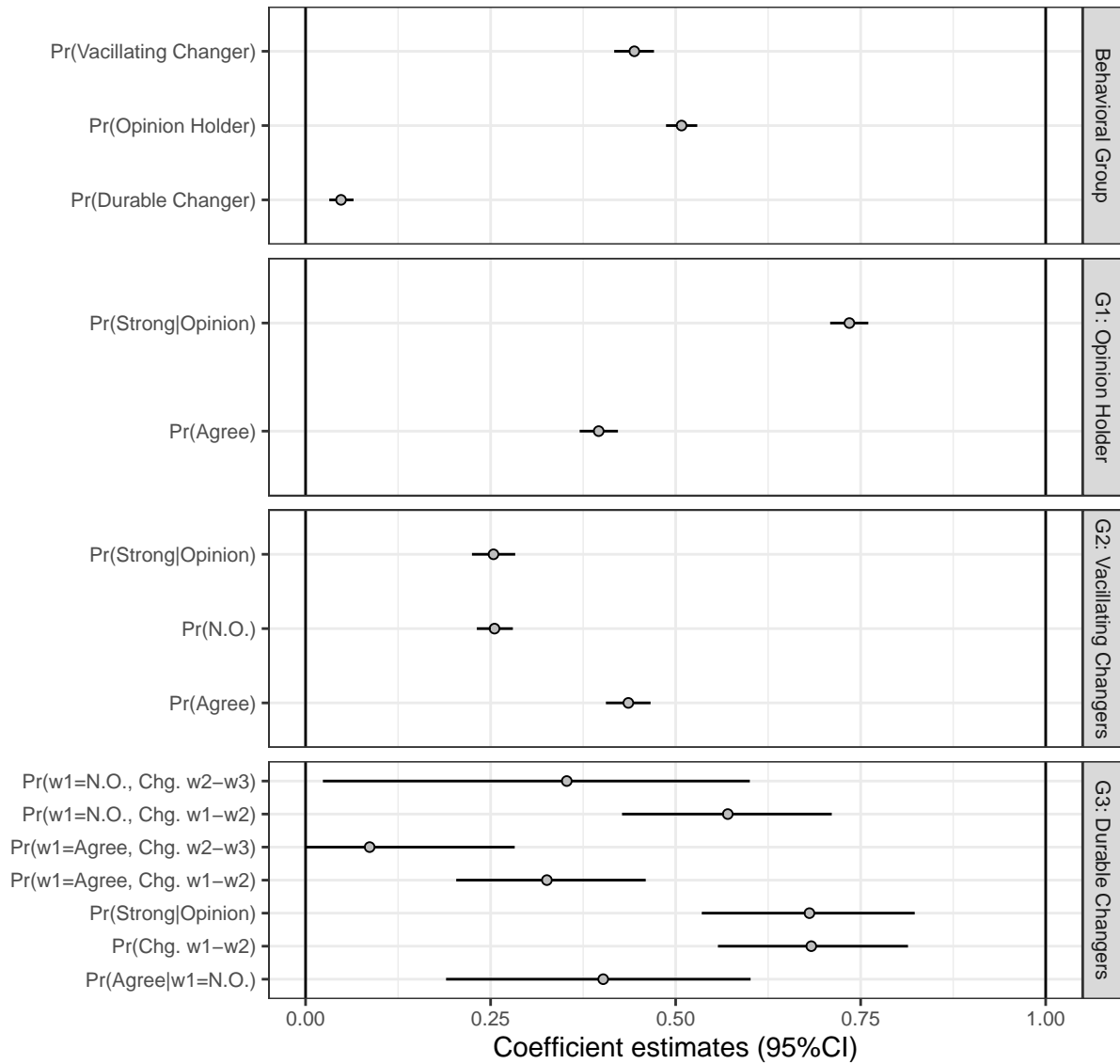
Figure 1: Parameter estimates of Fininte Mixture Model for Gay Marriage question in 2006-14 GSS panels. Posterior distribution estimated with 5 chains, 2500 iterations each, first 500 iterations of each chain discarded. "N.O." inficates "No opinion/Neither agree not disagree" response. "Chg." indicates "change".

the sample made durable changes in attitudes during the survey window (2006-14).

Is about half the population holding a stable view on the gay marriage issue a surprising number? It is significantly higher than the 20 percent of people Converse suggested had stable opinions on the issue of government control of utilities (Converse 1964), and it is roughly comparable to what Hill and Kreisi found when they explored opinions on pollution-abatement policies (Hill and Kriesi 2001a). That being said, it was a dominant political issue of the time, and the fact that more than four in ten people lacked durable opinions on the issue is notable.

Other estimates in Figure 1 are consistent with our intuitions about attitude change and stability. People who hold stable attitudes are more likely to report "strong" attitudes than people who hold vacillating attitudes (Howe and Krosnick 2017). Nothing in the model requires this; it is simply a function of response patterns. Vacillating changers said "neither agree nor disagree" about a quarter of the time, and when they gave an opinion they were about evenly split between agree and disagree ($\alpha_2 = .44$).

Consistent with the historical record, the group of people who made durable change tended to change from opposing gay marriage to supporting it ($\alpha_3^{(pre1)} = 0.318$ and $\alpha_3^{(pre2)} = 0.081$). They were also more likely to change between the first and second wave ($\tau_3 = .68$) than between the second and third wave, consistent with theories of panel updating.[7] Durable changers more closely resemble opinion holders than vacillating changers in their extremity, though there is more uncertainty in parameter estimates for the durable change group, since the group is smaller.

To this point, the model has done nothing to rule out the measurement error model, which suggests that the population comprises one group with different latent positions, responding with errors that are normally distributed. The model presented above could just be partitioning people based on their position in the scale. If this were the case, however, we would expect people in each group to demonstrate similar levels of within-person change over time. People in the stable opinion group would have a latent position near the ends of the spectrum, while vacillating changers and durable changers would have latent positions in the middle of the distribution, and all groups would respond with similar error ranges.

To compare the three-group model to the measurement error model, Hill and Kreisi (2001a) estimate the average within-person standard deviation for each behavioral groups at each iteration. If the measurement-error model is correct, then these group-based standard errors should be comparable across behavioral groups. If the vacillating changer group was composed primarily of people who shifted between "agree" and "neither agree nor disagree" and "disagree" and "neither agree nor disagree," they would have similar standard errors to people who bounce between "agree" and "strongly agree." Table 1 shows the average within-person standard deviation for the three groups.

Behavioral Group	σ (95% C.I.)
Opinion Holder	0.303 (0.295, 0.312)

Behavioral Group	σ (95% C.I.)
Vacillating Changer	0.909 (0.885, 0.932)
Durable Changer	1.469 (1.308, 1.631)

Table 1 shows that the average within-person standard deviations are different across groups. Opinion holders have a small average standard deviation ($sd_{G1} = .303$). Vacillating changers have average within-person standard deviations about three times as large as opinion holders ($sd_{G1} = .909$), and durable changers have an average standard deviation larger still ($sd_{G1} = 1.469$). In other words, vacillating changers and opinion holders do not simply reflect similar behavior at different points in the scale. Vacillating changers demonstrate much more wave-to-wave change than opinion holders. As Hill and Kreisi note in their examination of the Swiss pollution-abatement data, "Given the implausibility of assuming vastly different measurement errors for different people, it seems rather more likely that this variation is composed of both measurement error and true opinion instability" (Hill and Kriesi 2001a: p. 408).

5   Full Data

The above example demonstrates the model's ability to separate out three distinct sets of opinion behavior, and it provides a baseline for what we might expect for rates of stability, vacillation, and durable change. I now turn toward examining a wider range of attitudes across several panel data sets.

A key challenge in the study of attitudes and opinions is that while it is possible to take a random sample of respondents to generalize to a population of people, it is unclear how to randomly sample to draw inferences to a larger population of attitudes. In the absence of a random sample, many studies explore one or a few attitudes and assume that other attitudes behave similarly. I cast a wide net to capture as broad a range of attitudes and beliefs as possible. It is important to keep in mind that these attitudes do not represent a random sample, so statistical tools to draw inferences from this sample will likely be misleading.

The method outlined above requires at least three waves of responses by the same person to the same question. I focus on questions that ask respondents to give a statement of opinion or judgment on an issue. These questions are often structured as five-point Likert scales, which is the structure Hill and Kreisi designed the model to address. However, to apply this approach to a wider range of questions I make a handful of modifications to the model and the questions being examined.

Many questions lack a clear "no opinion" midpoint. However, interviewers tend to record a volunteered "no response" option. Assuming in the absence of a "no opinion" option people to select a second-best option or volunteer "no opinion," the effect would be a decreased value of the φ parameter, but it should not affect the distribution of the behavioral group parameters. Some questions with midpoints are not labeled "no opinion", but rather some other phrase. In cases where the language or substantive interpretation of this midpoint option is equivalent to "no opinion" or ambivalence, I employ the model as it is outlined above. In cases where the midpoint has a substantive interpretation other than "no opinion," I collapse the response into a binary response

structure, which I discuss below.

To accommodate two- or three-point scales – "agree"/"disagree" or "yes/no," often with a "no opinion" or "neither agree nor disagree" option – I remove the δ parameters from the model. The lack of "strong" response options has the consequence of greater uncertainty about whether a particular response pattern falls into one behavioral group or another, and the within-group standard deviation comparison is not possible with these questions as opinion holders will always have no within-person standard deviation.

Similarly, I adapt the model to handle seven-point scales by adding an additional parameter (γ_j) to account for "weakly" agree and "weakly" disagree or comparable points (response 3 and response 5 on a seven-point scale). For most questions, I maintain the assumption that the midpoint of these scales reflects a lack of opinion rather than a substantive "moderate" position. This might be a better assumption for some questions than others. However, if people hold substantive "moderate" positions, it will show up as lower average within-person variance for these questions.[8] Some questions have much larger scale ranges, including 10-, 12-, or 100-point feeling thermometers. When these questions have symmetrical structures, I collapse them questions into five-point scales.[9]

Finally, I modify some questions to conform to the model by changing non-symmetrical scales to a binary structures. For example, the General Social Survey includes a question about people's interpretation of the Bible: "Which of these statements comes closest to describing your feelings about the Bible? a. The Bible is the actual word of God and is to be taken literally, word for word. b. The Bible is the inspired word of God but not everything in it should be taken literally, word for word. c. The Bible is an ancient book of fables, legends, history, and moral precepts recorded by men." I collapse this into whether a person chooses Biblical literalism (choice a) or not. In this way, it conforms to the binary model (without a delta parameter in any group). This approach sacrifices nuance (people who believe the Bible is inspired by God and those who believe it is a book of fables are considered the same, and changes between these beliefs do not count for anything) for direct comparability with other questions. In these cases, I try to focus on the most substantive division in the structure.

A search of multiple panel data sets produced 544 questions for analysis. For this analysis I use data from the General Social Survey's rotating three-wave panel (166 questions)[10]; the American National Election Studies panels from 1956-60 (10 questions), 1972-1976 (71 questions), 1992-1997 (66 questions), and 2000-2004 (57 questions); and the Cooperative Congressional Election Study (36 questions). I also use data from the National Study of Youth and Religion (45 questions), two panels – one of mothers, one of their children – from the Intergenerational Study of Parents and Children (17 questions in the mother panel, 18 in the child panel),

---

[8]Kinder and Kalmoe (2017) suggest that on seven-point political ideology scales, moderates are "indistinguishable from those who remove themselves from ideological analysis" (p. 160) and that we should regard these people as people who lack opinions, not people who hold sophisticated positions in the middle of the distribution. This strengthens my confidence on using the midpoint as a non-attitude point, rather than a moderation position.

[9]On 100-point scales, respondents less than or equal to 20 and 80 as giving considered to give "strong" responses, and individuals at 50 are considered to give ambivalent responses.

[10]The GSS panels were collected over an eight year period as a series of three rotating three-wave panels. Each panel extended over a four-year period (2006-2008-2010; 2008-2010-2012; 2010-2012-2014). Because of the significant overlap in time periods, I treat the GSS panels as a single panel.

and the Longitudinal Study of Generations (58 questions). These data sources have distinct advantages when compared with each other. The GSS, ANES, and CCES panels are representative samples of U.S. adults over the age of 18. The NSYR is a nationally representative sample of young people, who are not included in the GSS panels, and includes questions about religious and moral issues that the adult panels lack. The LSG and ISPC are not a probability samples, so it is hard to make inferences about the larger population from these samples, but they include distinct sets of attitude items and allow for the analysis of some questions in different temporal or generational settings.

The panels also cover different time intervals. The GSS, ANES, and CCES panels cover about four years. The ISPC panel covers 13 years. Other panels fall between these. In general, I suspect that stable opinions will be easier to hold over shorter time frames, so panels covering a longer time frame will see more vacillation and durable change.

It is impossible to say whether the attitudes captured in the 544 questions analyzed here are "representative" of attitudes in general. Because of the nature of the surveys and the kinds of topics get covered in panel studies that extend at least three waves, political views make up a large proportion of questions in this analysis. However, the data sets – particularly the GSS – contain a broad array of questions on other topics including beliefs about race relations, gender roles, family structures, moral worldviews, intergroup relations, and social and institutional trust.

Appendix B lists the full range of questions analyzed here, as well as any modifications to the question structure.

6   Results

I apply the finite mixture model outlined above to these 544 questions. My discussion of the results proceeds in three sections. I first examine the overall distribution of the three response patterns across questions, drawing general conclusions about attitude stability and change across data sets and questions. Next, I use a multiple regression approach to evaluate the effects of question content and structure on rates of stability and change. Finally, using results from the 166 questions from the GSS, I examine the pairwise correlations of stability estimates to look for evidence of "issue publics," or sets of issues where stability tends to co-occur. I focus on the GSS because it includes the most questions touching on the widest variety of issues.

6.1   Overall Pattern Proportions

As a first step in this analysis, Figure 2 presents a series of boxplots, three for each data set, plotting the distribution of mean estimates for the proportion of durable changers, vacillating changers, and opinion holders for each question. The figure contains a lot of information, but it highlights some clear patterns across the data sources.

Starting at the left, the figure shows that durable change is rare across all data sets and almost all questions. For about half of all questions, the mean estimated proportion of respondents demonstrating durable change
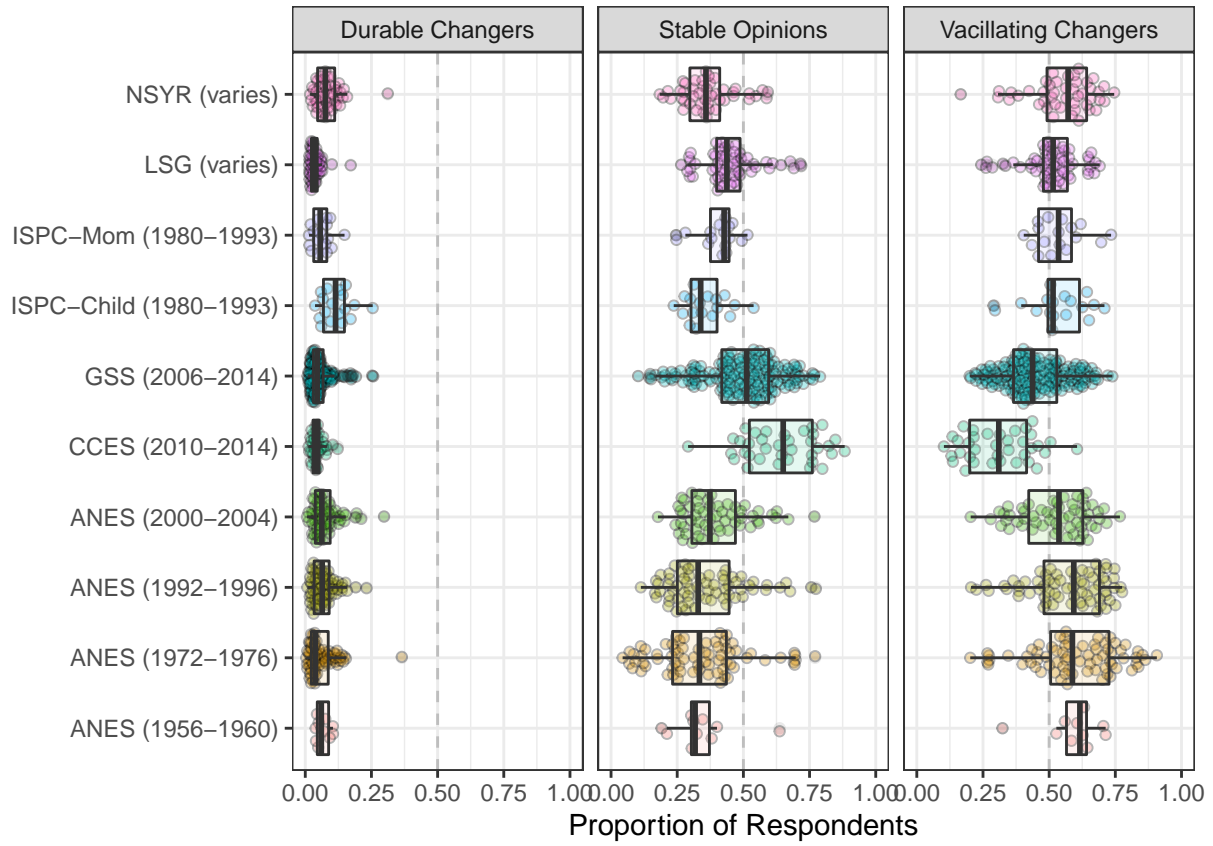
Figure 2: Mean proportion of respondents in each behavioral group, by question and data set. Posterior distribution estimated with 5 chains, 2500 iterations each, first 500 iterations of each chain discarded.

is less than 5 percent, with many of these questions showing rates statistically indistinguishable from 0.[11] Only 17 percent of questions have mean rates of durable change above 10 percent. Attitudes that exhibit very high levels of attitude change typically pertain to major public events, and are exceptions that prove the rule that durable change is, in general, rare.

Second, there is a range of ambivalence and stability across the data sets. Most of the data sets analyzed here include questions that demonstrate both high (greater than 70 percent) and low (less than 30 percent) levels of stability and vacillation. Since within-survey comparisons reflect the same sample, vacillation is not a feature of the population (the kinds of people being surveyed) but an interaction between people taking the survey and the kind of question being asked. In other words, there are some questions that most of the population can answer consistently over time, and there are questions where very few people in the population can answer consistently over time. However, most questions fall somewhere in the middle, with some people answering consistently and others not.

Third, vacillating change is on average more common than stability for the questions analyzed here. For 56 percent of the questions, more than half the sample demonstrated vacillating change. The 1956-60 ANES panel, which Converse analyzed to generate his original insights about non-attitudes, displays the highest average level of vacillation, but it is not an outlier. Almost all other surveys include questions that surpass that average in ambivalence. The GSS and the CCES are the only data sets where more questions show higher levels of stability than vacillation.

There are other features of the results that should give us confidence that the finite mixture approach is separates distinct sets of opinion behavior and is not just partitioning respondents based on their positions on the scale. Figure 3 plots the proportion of opinion holders and vacillating changes who give strong responses, conditional on them expressing an opinion (not saying "no opinion" or "neither agree nor disagree"), for questions with four, five, or seven response options. Because they lack "strong" responses, questions with two and three response options are excluded here. Points above the line suggest that people in the opinion holder group give "strong" responses at a higher rate than vacillating changers, conditional on giving an opinion.

For almost all questions with a "strongly agree" option, people with stable attitudes are more likely than people with vacillating attitudes to respond with that position, often much more likely. This suggests vacillating changers are generally expressing weaker opinions than opinion holders. At the same time, respondents who vacillate still select the "strong" option, often quite frequently suggests that they tend to have a larger response range than opinion holders.

If the measurement error model is correct, within-person standard deviations should be comparable across behavior groups. If standard deviations are not comparable, we have more confidence that these groups exhibit different opinion behaviors. Figure 4 presents two sets of plots. The top row compares vacillating changers to durable changers, and the bottom row compared vacillating changers to opinion holders. Because questions with 2 or 3 response options will always have a within-person standard deviation of 0 for the opinion holders

---

[11] Because of the conservative nature of the estimation process, the estimate for the proportion of durable changers can never be 0. But most of these questions have confidence intervals that extend below 1 percent.
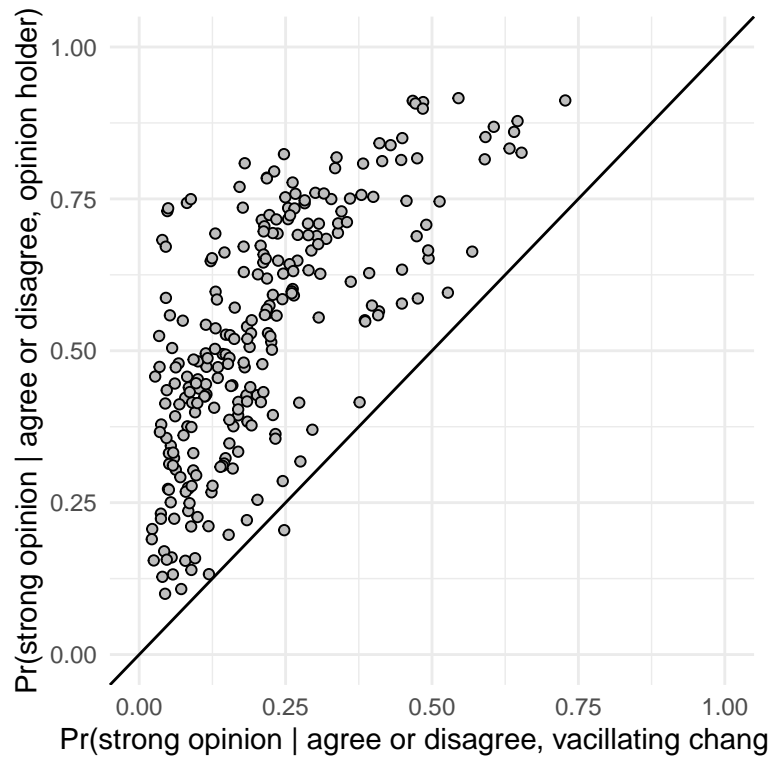
Figure 3: Mean estimated proportion of "strong" responses in opinion holder and vacillating changer behavioral group. Posterior distribution estimated with 5 chains, 2500 iterations each, first 500 iterations of each chain discarded.

group, they are omitted from this comparison. Points above the line indicate questions where the vacillating changers have greater within-person variance than opinion holders.
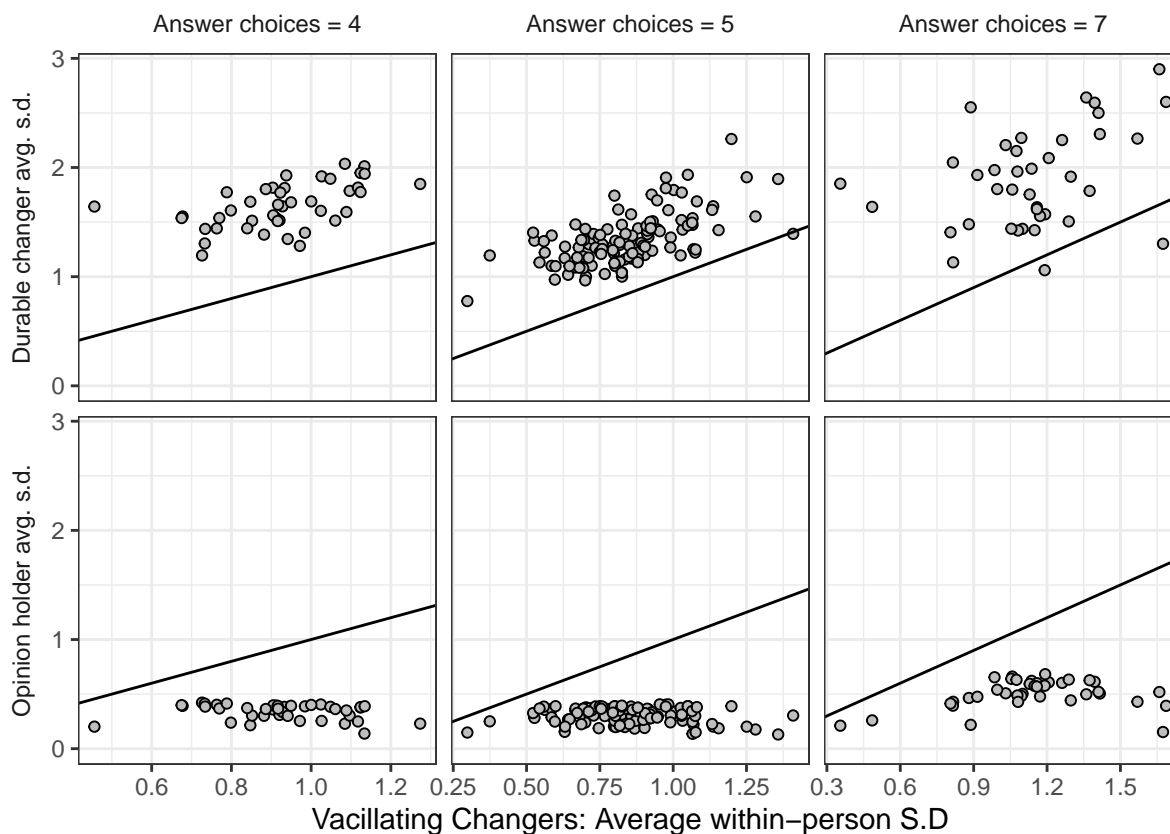

Figure 4: Average within-person standard deviation, by behavioral group and question response options. Black line indicates equal within-group standard deviations. Posterior distribution estimated with 5 chains, 2500 iterations each, first 500 iterations of each chain discarded.

Consistent with the gay marriage example presented above, Figure 4 shows that vacillating changers uniformly have higher within-person standard deviations for almost all questions with four or more response options, regardless of question structure, and almost uniformly lower within-person standard deviations than durable changers.[12] This suggests that durable changers are more likely to move from extreme ends of the scale, while vacillating changers are more likely to cluster in the middle of the scale (though with more variation than opinion holders). In other words, we can be fairly confident that the model is separating three different groups with distinct sets of behavior. Notable here is the fact that the pattern is consistent with most seven-point questions. On five-point questions, vacillating changers might have more room to move around and express their nuanced opinions (three positions: agree, neither agree nor disagree, and disagree) than opinion holders (strongly agree and agree). But even when opinion holders have room to give a wider range of responses, they still demonstrate uniformly less within-person variation than vacillating changers.

---

[12] The three questions where durable changes have lower average within-person standard deviation than vacillating changers ask respondents if if the government should bus students to ensure racial equality (72-76 ANES), whether the government should do more to improve the conditions of blacks (92-96 ANES), and whether they are happy with their bodies (NSYR).

As an overall conclusion, we can say that the model is detecting three distinct sets of behavior, consistent with Converse's original black and white model, with the addition of a durable change group. Opinion holders tend to exhibit strong opinions at the ends of scales. Vacillating changers, rather than reflecting relatively settled views at scale midpoint, tend have a much wider range of responses that vacillate on either side of the issue. Durable changers tend to move between ends of the scale, often from one extreme to another. While we can say that overall, durable change is rare, there is a large variance in stability and vacillating change that remains to be explained.

6.2 Question Content and Structure

Give the large number of questions examined and the covariation in content and structure, it is difficult just looking at the overall patterns to say whether it is question content or structure that drives overall rates of stability or vacillation. The GSS has many more questions with fewer response options (questions with two, three, or four choices make up most of the GSS questions), but also more questions about non-political content. The ANES panels tends to favor questions with more response options, typically five or seven, but focuses almost exclusively on questions of politics. An additional feature of question structure that might matter is the explicit option of a midpoint to signal ambivalence.

To better assess these influences, I tag questions with broad indicators for content areas: government and politics; civil liberties; medicine and science; race; sex; family and gender[13]; socioeconomic position; work; morality; foreign relations; self-evaluations; religion; crime; social trust; and other. These categories are not mutually exclusive, and many questions are asked in a survey because feature the intersection of multiple domains. Because of this, I focus on an explicit reading of the question and code items parsimoniously. Because more than half of all questions analyzed here touch on government and politics in some capacity, I instead use an indicator to signal that a question does *not* explicitly deal with politics. I also include indicators for whether a question asks people to assess another person's position, and indicators for whether the question refers to a clear group of people or a particular person.

I also create indicator variables for question structure. First, I include five indicators about question format. The first format asks whether people agree with a statement. These include traditional five- and four-point Likert scales and two- and three-point agree/disagree or yes/no statements. The second format asks people to select a position select between two poles or options. The third format, common the ANES panels, asks respondents to place a respondent on a "feeling thermometer." A fourth format asks people whether groups have "too much" influence in politics. The final category includes other question structures. Most questions in this last category evaluate the intensity of something, such as whether an amount should increase, stay the same, or decrease. While many questions with "other" structures were collapsed to reflect a "yes/no" structure, I code based on the original wording. I also include indicator variables for the number of response options (2, 3, 4, 5, 7) and indicator variables for data set.

I then regress the proportion of respondents falling into each behavioral group on these indicators. As noted

---

[13]Ideally I would include separate indicators for questions relating to gender and questions relating to family structures. However, almost all questions about gender touch on some other subject material, with family structures being the largest overlap.

above, the questions explored here do not represent a random sample of questions, so standard confidence intervals and statistical inferences are not meaningful. However, I incorporate uncertainty in two ways. First, I bootstrap coefficient estimates, randomly sampling questions with replacement 10,000 times. Second, to incorporate the uncertainty in each question's proportion estimate, each time a parameter is sampled in the bootstrap, I draw a random value from that parameter's posterior distribution. By bootstrapping the estimates we can be relatively confident that whatever patterns we observe are not driven by a single question or small group of questions with unusual behavior, even if I cannot generalize to all attitudes.

In all regressions, the reference group consists of general questions about government (particularly government's role in the domestic economy) on a five-point agree/disagree scale, with the General Social Survey serving as the reference data set.

6.3 Vacillation

Figure 5 presents coefficient estimates for a regression of the proportion of respondents giving vacillating response patterns on question content and structure. An unpictured regression of the proportion of respondents giving durable change response patterns on the same parameters is almost a mirror-image of the figure presented here, with the variables predicting higher vacillation predicting lower stability. As a result, I will discuss both stability and vacillation in reference to the Figure 5.

A general five-point agree/disagree question about domestic government and politics from the GSS has an expected proportion of people vacillating of about .58 with a standard error of .033. Two content areas demonstrate notably less vacillation than other kinds of questions: questions about religious beliefs and questions about sex and homosexuality. On the other end of the scale, questions about general morality, questions about civil liberties, and questions about medicine and science show more vacillation than other kinds of questions. When questions fall outside of the political space ("Not politics" in Figure 5), they tend to show higher levels of vacillation. Notably, questions about race, foreign relations, work, crime, and family and gender do not appear substantially different from questions about politics once I control for question structure.

Questions that ask people to assess others – where do politicians or parties fall on particular issue scales – demonstrate much higher vacillation than questions about people's own opinions. This is consistent with previous work (Alwin 2007) and suggests a lack of knowledge about actors' in the political sphere.

In terms of question structure, stability decreases and vacillation increases as the number of response options increases. This helps explain the higher levels of stability for GSS and CCES questions relative to the other data sets. Questions about groups' political influence show more stability than other kinds of questions, but beyond that, question structure does not seem to matter much.

Finally, even controlling for question structure, data sets demonstrate different rates of vacillation. Notably, the 1972-1976 ANES and the NSYR demonstrate more vacillation than other data sets. The CCES demonstrates lower rates of vacillation.

Structure, content, and data set explain about 57 percent of the variation in the proportion of vacillating re-
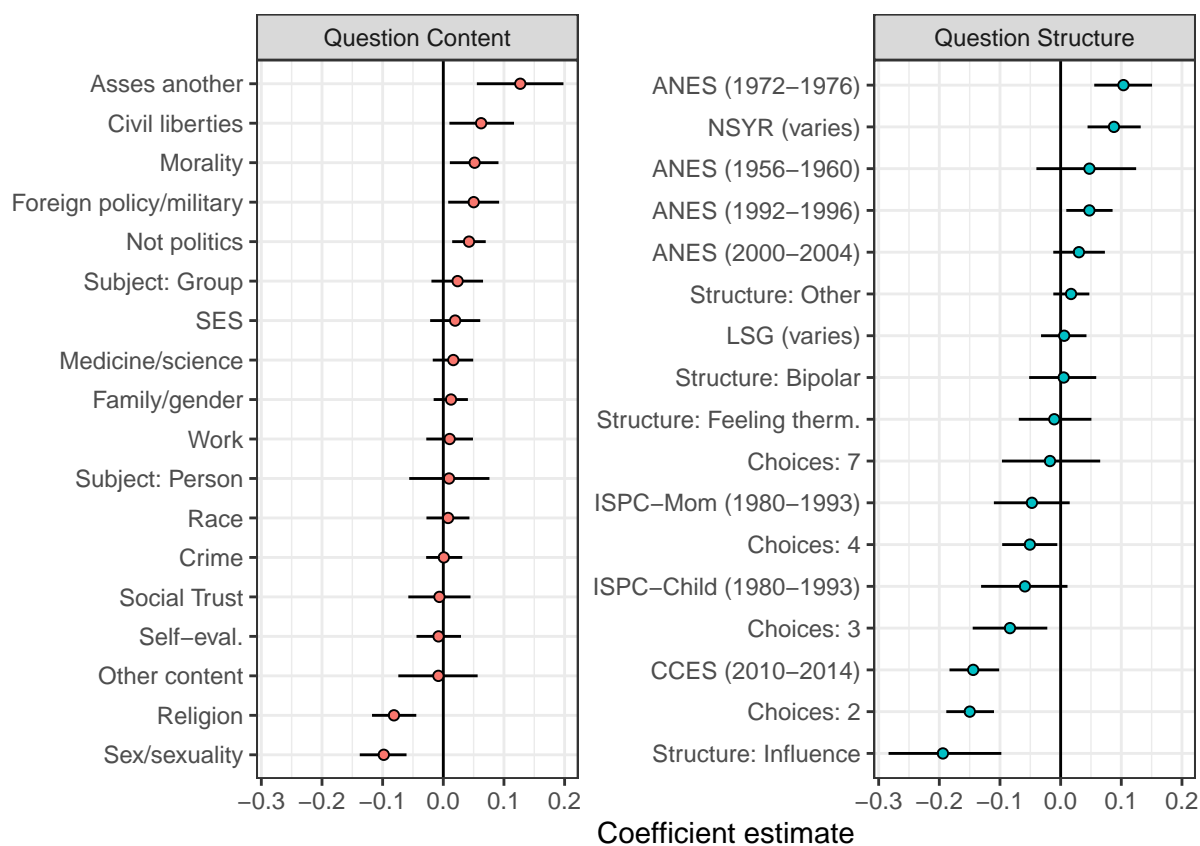
Figure 5: Coefficient estimates of regression of proportion of respondents of giving a vacillating response on question content and structure indictors. Confidence intervals generated from 10,000-iteration bootstrap.

sponses for the questions analyzed here, but there are some questions that are poorly explained by these categories. Across data sets, partisan identification is much more stable than other bipolar measures of general political ideology (position on liberal-conservative scale; role of government in economy; role of government in equal opportunity, etc.) with which it is often classified. People also tend to be more stable on issues that deal with schools, such as whether prayer should be allowed in schools and whether busing should be used to integrate schools.

Another pattern in the residuals is that questions about specifics – specific policies and specific behaviors – tend to be more stable than predicted while questions about general concepts (ideology, broad morality) tend to vacillate more than predicted. This helps explain why the CCES, which includes many questions about specific policies, has such a high rate of stability.

6.3.1 Durable Change

Figure 6 plots coefficient estimates of the proportion of respondents demonstrating durable change. The coefficients associated with question content and structure are much smaller on durable change than on vacillating change, reflecting its truncation in low end of the proportion rage, and the parameters here do not explain nearly as much variance as the model for vacillating change. There is almost no appreciable effect of question content on the rate of durable change.

Two data sets demonstrate substantially higher levels of durable change: the NSYR and the ISPC-Child, the two panels with the lowest average age.

There is a less clear pattern on question structure for durable change than for vacillating change. Questions with five response options (the reference category) demonstrate the most durable change, followed by questions with three responses. Questions with two, four, and seven responses demonstrate comparable, low levels of durable change.

The largest residual in the model pertains to people's evaluation of Richard Nixon in the 1972-76 ANES panel, in which a large proportion of people shifted from a positive opinion or no opinion to a negative opinion. Given that this window covers from Nixon's victory in the 1972 election through the Watergate scandal and his eventual resignation, this shift is not surprising. Other questions with high rates of durable change include a shift in the NSYR in which many adolescents go from saying that people should wait to have sex until they are married to saying "not necessarily," with very few shifting the other way, and a large change of opinion around whether the presidential in 2000 was fair.

6.4 Issue Publics

The preceding results suggest that, on any particular issue, some people seem to hold opinions while other people seem lack them. While this is strongly suggestive of an issue publics model, it does not rule out the probability that what drives stability resides at the individual level, and that questions simply vary in the degree to which people can stably answer them. To test whether these results are consistent with the "issue publics" theory, I explore the correlation of the pairwise probabilities of giving stable responses to questions, generated
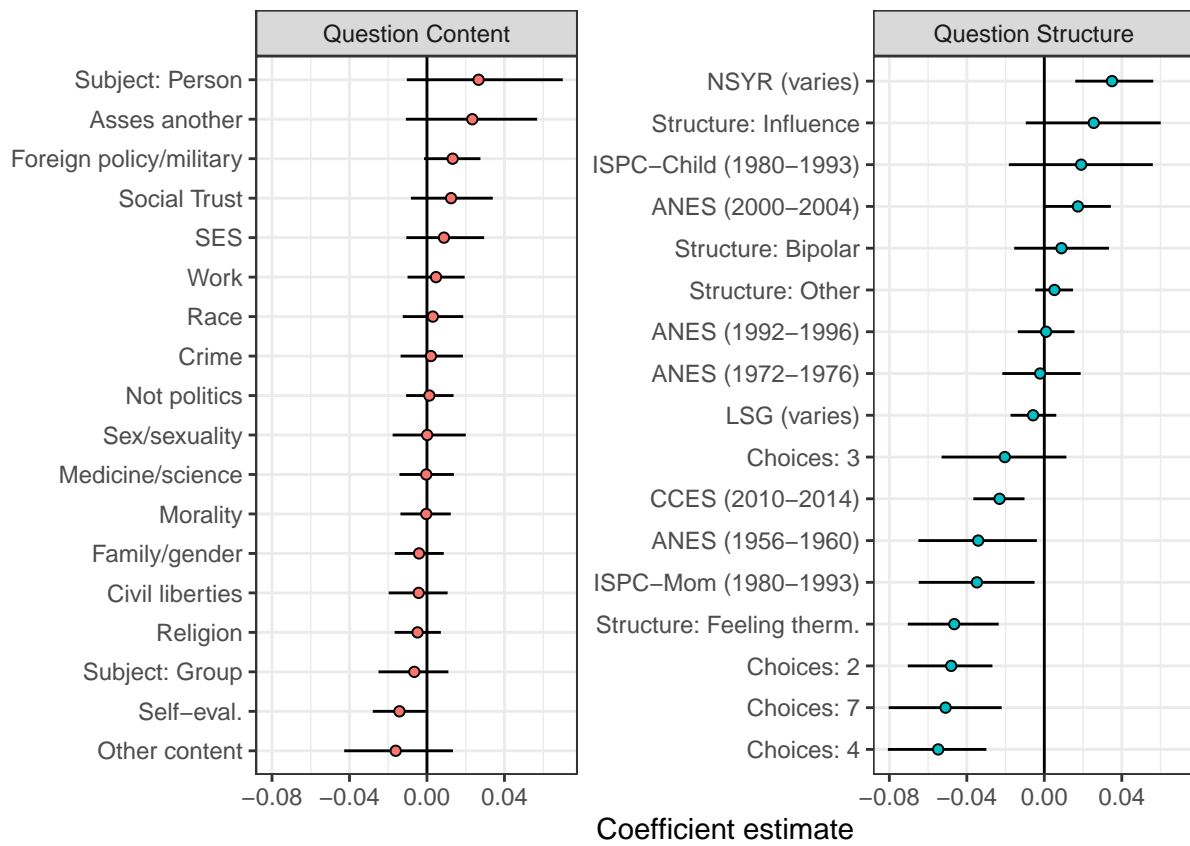
Figure 6: Coefficient estimates of regression of proportion of respondents of giving a durable change response on question content and structure indictors. Confidence intervals generated from 10,000-iteration bootstrap.

by the proportion of times that a respondent is classified in each group over the 10,000 iterations of the model. If stability is socially patterned, then people should be stable on related questions but not necessarily stable on unrelated questions. If a measurement error model is correct, then the probability of being stable on one question should not be related to their stability on any other questions. If stability is principally a function of individual cognitive ability or other individual feature, then pairwise correlations should be strong, as people who hold stable opinions do so across the board and other people fail to.

The 166 questions in the General Social Survey panels examined here produce 13,695 pairwise correlations.[14] In general, pairwise correlations are very low, with an average correlation of .04, suggesting that stability in one attitude is not a strong predictor of stability in another attitude. This provides additional support for the argument that attitude stability is not principally a function of the person answering the question. At the same time, some pairwise correlations are quite strong. To better understand the distribution of these strong correlations, I plot them as a network diagram in Figure 7. For parsimony, I focus on the 541 correlations greater than $\rho > .2$, about 2 percent of all correlations. Ties between issues indicate that people who are stable on one issue are stable on the other, and that people who vacillate on one issue vacillate on the other.

Figure 7 reinforces the "issue public" nature of attitude stability, showing a series of spanning trees, with the occasional sense cluster of attitudes. It is important to keep in mind that the figure does not reflect the correlation of the underlying issues themselves, but rather the correlations of the probability of stability. First, 58 questions in the GSS do not demonstrate any substantial correlations with any other questions. This group includes a heterogeneous mix of questions about gun laws, whether there is an afterlife, whether premarital sex is acceptable, and more. However, closely related questions tend to demonstrate simultaneous stability. The largest component of the network is centered on a dense cluster of questions about civil liberties, but it spans out to questions of police use of force, abortion, and the values that should be demonstrated by children. At the same time, this component is totally disconnected from questions about political views.

Also notable in the figure are the lack of connections between certain issues. Despite having a similar structure and appearing next to each other in the survey, sets of questions about abortion are completely disconnected in the figure – holding a stable belief about abortion in some circumstances (child defect, women's health, and rape) does not necessarily beget a stable position on other circumstances (single mother, as a method of birth control, if the mother is poor, and any circumstance the woman wants). The same is true of beliefs about suicide. Similarly, stability in general political beliefs (liberal-conservative ideology, partisan identification) is completely disconnected from a broad set of questions about government spending. Questions about general social trust are disconnected from specific questions about confidence in institutions, and questions about specific morality are disconnected from questions about general morality.

6.5 General Discussion of Results

The results presented above suggest several general conclusions about the behavior of attitudes, at least the attitudes that tend to get measured in social science surveys. First, it is wrong to say that people, in general, do

---

[14]There are seven federal spending questions where different versions are tested for different people, examined separately above. The two versions show minimal differences, so I consolidate each pair into a single item for this analysis.

Figure 7: Network graph of pairwise correlations of co-occuring stability greater than .2. Probabilities generated from group assignments over 10,000 draws from posterior distribution.

not hold durable opinions. On most issues, there is some proportion of the population that consistently locates itself on one side of an issue or another. Consistent with psychological research on opinions, when people hold opinions stably, they are disproportionately likely to say they hold this opinion strongly (Howe and Krosnick 2017). While simpler questions – with fewer answer choices or less abstract wording – seem easier for people to answer consistently over time, some proportion maintain stable attitudes even on complicated, abstract, and conceptual questions.

At the same time, it is also wrong to say that people in general "have" opinions, as in common in latent attitude and measurement-error models. On any particular question there is typically a large group of people who do not maintain consistent opinions, vacillating between ends of the scale in ways that suggest they are subject to short-term considerations. Examination of the average within-person standard errors for different groups suggest that people who vacillate are not just people at the middle of the scale who report with error; they vacillate more widely than people who hold stable opinions. These patterns are strongly consistent with Zaller's notion of ambivalence, that on any particular issue, "people are likely to internalize many contradictory arguments" and "form considerations that induce them both to favor and oppose the same issues" (1992: p. 59). This is true not just of politics, but general morality, religious beliefs, and more.

A striking pattern in the results presented here is similarity of attitudes across topic domains. It is not the case that different kinds of attitudes behave differently. While there was a large range in the amount of stable opinions and vacillating change across questions, there were no questions analyzed here where all people demonstrated stability and no questions where all people demonstrated vacillating attitudes.[15] This suggests that the general mechanisms that produce attitude stability or vacillation are present across domains, but the specifics might vary widely. Similarly, most questions demonstrate low levels of durable change, with the exceptions reflecting high-profile events that are, by necessity, rare. It is not the case that people are forming and reforming attitudes that they carry with them over medium time frames.

In contrast to expectations, people are not inherently better at answering non-political questions than political questions, with the exception of questions about religious beliefs and sex. While questions about political issues do demonstrate high levels of ambivalence, they are not unique in that regard. People are similarly ambivalent about morality, civil liberties, race, gender, family structures, medicine, and more. In fact, when issues in other domains – race, social trust, medicine, etc. – interact with the political space, people seem more likely to report the same opinion consistently. This suggests, consistent with Zaller's model, that people get their cues about what to believe from opinion leaders.

Across domains, specific attitudes tend to be more stable than general attitudes. People are much more likely to be consistent on specific moral prohibitions than on general questions about morality (whether morality is a personal matter, whether issues are black and white or contain shades of gray, whether morality should change with the times). They are more stable on specific questions about government spending on various priorities than on general questions about the role of government and political ideology. This suggests that people work inductively from specifics to general principles, and that they seem to have a hard time reconciling disparate

---

[15]It is plausible that such questions exist but make for bad survey questions. Survey questions tend to focus on areas where there are disagreement in the public.

specifics. They do not seem to apply general principles to specific outcomes. In the political domain, this is consistent with the notion that people are "ideologically innocent" (Kinder and Kalmoe 2017), and further argues against general or latent beliefs. This, combined with the fact that issues demonstrate greater stability when they intersect with the political realm, could also suggest that stability on issues is a function of what people hear, with specific issues more frequently discussed than general principles.

6.6   Conclusions

This paper sought to adjudicate a longstanding divide present in both cultural sociology and public opinion scholarship between theories suggesting that people lacked clear opinions and those suggesting that people held stable, real opinions. Going back to Converse's original formulation of the "black and white" model, I suggest that for any particular question, both were likely true: some people hold real, stable attitudes while other people lack stable attitudes altogether. Using a finite mixture model approach, I found that all questions could be divided into people who hold stable opinions, people who vacillate, and a small group of people who form an opinion or change their opinion during the course of the survey, with the vacillating group tending to outnumber opinion holders. This set of results is strongly consistent with Zaller's model of ambivalence, but also suggest that some people in the population do form stable, real opinions.

For cultural sociology, the results presented here reinforce notions that opinion stability is a function of social scaffolding – institutions, context, and social networks. The large variation in stability across questions and the clustering of stability on related topics, suggest that features of the social environment facilitate attitude stability. Exactly what kinds of social scaffolding, and where in the life course their influence matters most, is unclear. It could be the case, as Kiley and Vaisey (Kiley and Vaisey 2020) suggest that people with stable dispositions acquire these early in life, while people who do not acquire these early never do. It could also be the case that people who hold stable attitudes do so because they are embedded in social structures that facilitate them. Adjudicating the mechanisms that facilitate stability requires further work.

The results have broad implications for theories of social change. Kiley and Vaisey (2020; Vaisey and Kiley n.d.; Vaisey and Lizardo 2016) posited that people have relatively stable views by the time they reach adulthood, especially for the GSS questions they analyzed. This is reinforced here by the low proportions of people demonstrating durable change. The NSYR demonstrates the highest rate of durable change. Similarly, the child panel of the ISPC demonstrates much higher rates of durable change than the mother panel, and both panels contain the same questions, extend for the same duration, and are asked in the same years.

At the same time, their conclusion that people have settled dispositions is challenged by the high amount of vacillation in many questions. While we can say that large swaths of the population have settled dispositions about questions about sexual morality, religious beliefs, and some specific governmental policies, they lack consistent opinions on more general conceptual questions – political ideology, general beliefs about the role of government, general morality, and abstractions about civil liberties. If people tend to be ambivalent, then it is possible for their attitudes to be influenced by a consolidation of elite opinion around a topic. However, so long as culture remains heterogeneous, it seems likely that large groups of the population will vacillate.

These results have implications for methods involving attitudes. The high degree of inconsistency in opinions – inconsistency that seemingly cannot be attributed to measurement error – should caution researchers away from assuming that a cross-sectional measure of attitudes is a good proxy for a person's belief. For many questions, more than half a sample might be responding with a temporary attitude construct, while others report real attitudes that matter to them. Given this heterogeneity, we should not expect strong predictive ability for that attitude in general, but that tells us nothing about how it influences the people who truly hold that attitude. At the same time, it is not clear how many respondents need to be making stable responses for an attitude to be a good predictor for behavior (a coarsened version of the question Vaisey (2009) used to predict behavior was only stable for about 27 percent of respondents).

Ultimately, these results suggest that researchers should devote attention to the social conditions that facilitate stable attitudes and dispositions, whether they are located in people's past or in their contemporary environment. While we know that political awareness tends to facilitate stability in political beliefs, we have no expectation that political awareness should facilitate stability beliefs about religion, morality, or more, but it does suggest that domain-specific awareness might predict domain-specific stability. Attention to the mechanisms that create attitudes can help researchers develop a better understanding of the role of culture in behavior.

7 Appendix A: Model Estimation

The following is a summary of the model outlined in Hill and Kreisi (Hill and Kriesi 2001b) with modifications for three waves, rather than four. The goal of the Data Augmentation algorithm is to get draws from the posterior distribution $p(\theta|Z)$. It has two main steps: Drawing group membership indicators, given the parameter, and drawing parameters given the group membership indicators.

7.1 Re-Expressing the Data

Rather than estimate parameters on the basis of full patterns, such as "agree"-"disagree"-"strongly disagree," I characterize a respondent by a limited number of general variables. These variable retain the features of the data without having to generate a probability for each pattern separately. For example, by re-expressing the data in this manner I collapse the distinction between someone who says "agree"-"strongly agree"-"agree" and someone who says "agree"-"agree"-"strongly agree." Both patterns become someone who remains on the same side of an issue and gives one strong response.

$$A_i = \begin{cases} 1 & \text{is the } i\text{th person's initial response is a 4 or 5} \\ 0 & \text{otherwise} \end{cases}$$

B_i = number of the ith individual's responses that are either 1 or 5 across all *t*

C_i = number of the ith individual's responses that are 3

D_i = number of times the ith individual crosses an opinion boundary

$$E_i = \begin{cases} 0 & \text{if } D_i \neq 1 \\ t_i^* & \text{otherwise} \end{cases}$$

$$F_i = \begin{cases} 0 & \text{if the } i\text{th individual's preswitch response is a 1, 2, or 3, or } D_i \neq 1 \\ 1 & \text{if the } i\text{th individual's preswitch response is a 4 or 5} \end{cases}$$

$$H_i = \begin{cases} 0 & \text{if the } i\text{th individual's preswitch response is a 1, 2, 4, or 5, or } D_i \neq 1 \\ 1 & \text{if the } i\text{th individual's preswitch response is a 3} \end{cases}$$

$$M_i = \begin{cases} 0 & \text{if the } i\text{th individual's postswitch response is a 1, 2, or 3, or } D_i \neq 1 \\ 1 & \text{if the } i\text{th individual's postswitch response is a 4 or 5} \end{cases}$$

$$Q_i = \begin{cases} 0 & \text{if the } i\text{th individual's postswitch response is a 1, 2, 4, or 5, or } D_i \neq 1 \\ 1 & \text{if the } i\text{th individual's postswitch response is a 3} \end{cases}$$

R_i = number of the ith individual's responses that are either 1 or 2 across all *t*

I denote the vector of these random variables for individual i as Z_i.

7.2 Drawing Group Indicators Given Parameters

I use the observed data Z_i to determine the probability that a person falls into each behavioral group. The conditional probability of belonging to each behavior group given the observed responses is expressed by the following functions:

$$p_{1,i}^z = (\alpha_i^{(1-A_i)}(1-\alpha_1)^{A_i}(1-\delta_1)^{(3-B_i)}\delta_1^{B_i})I(C_i=0)I(D_i=0)$$

$$p_{2,i}^z = \varphi_2^{C_i}\alpha_2^{R_i}(1-\varphi_2-\alpha_2)^{(3-C_i-R_i)}\delta_2^{B_i}(1-\delta_2)^{(3-C_i-B_i)}$$

$$p_{3,i}^z = [\tau_3^{I_{(E_i=1)}}(1-\tau_3)^{I_{(E_i=2)}}(\varphi_3^{(pre1)^{H_i}}\alpha_3^{(pre1)^{(1-F_i)(1-H_i)}}(1-\alpha_3^{(pre1)}-\varphi_3^{(pre1)})^{F_i})^{I_{(E_i=1)}}$$

$$\times\,(\varphi_3^{(pre2)^{H_i}}\alpha_3^{(pre2)^{(1-F_i)(1-H_i)}}(1-\alpha_3^{(pre2)}-\varphi_3^{(pre2)})^{F_i})^{I_{(E_i=2)}}$$

$$\times\,((1-\alpha_3^{(post)})^{M_i}(\alpha_3^{(post)})^{(1-M_i)})^{H_i}\delta_3^{B_i}(1-\delta_3)^{(3-C_i-B_i)}]I(D_i=1)I(Q_i=0)$$

where $I(.)$ equals 1 if the condition in the parentheses is met and 0 otherwise. These probabilities represent the expected frequency of a particular pattern given the parameters that describe behavior in that opinion behavior group. For example, if a person responds "no opinion" or "neither agree nor disagree" in one wave, then $D_i \neq 0$, and $p_{1,i}^z$ will equal 0 for that person, since people who express "no opinion" can not be opinion holders. If $\alpha_1 = .5$ and $\delta_1 = .75$, we expect patterns where a person give a "strongly disagree" response in all three waves to make up about 21 percent of responses in group 1.

Generating these probabilities require parameter estimates. In the first iteration of the model, these parameters are drawn from a random place in the parameter space.

With these probabilities and draws for π_1, π_2, and π_3 (sampled randomly from the parameter space in the first iteration), I sample from the following trinomial distribution for each person to generate group membership indicators for each iteration.

$$p(G_i|\theta, Z_i) = \text{Mult}\left(\frac{\pi_1 p_{1,i}^z}{\sum_{j=1}^3 \pi_j p_{j,i}^z}, \frac{\pi_2 p_{2,i}^z}{\sum_{j=1}^3 \pi_j p_{j,i}^z}, \frac{\pi_3 p_{3,i}^z}{\sum_{j=1}^3 \pi_j p_{j,i}^z}\right) \tag{1}$$

These draws generate group membership indicators for a single iteration of the model. The probability that a person's responses are generated from a particular pattern are a function of the likelihood that a pattern was generated by a particular behavioral group, $p_{i,j}$, and the overall prevalence of that behavioral group in the population, π_j. While a pattern "strongly agree"-"agree"-"strongly agree" might be most consistent with a stable opinion, as the proportion of the population vacillating increases, the the probability that the vacillating group produced this pattern increases.

7.3 Drawing Parameters Given Group Indicators

I then draw parameter estimates from their distribution conditioning on the data (Z_i) and the group indicators drawn in the prior step. In other words, I fit separate models for each (temporarily assigned) behavioral group to generate parameter estimates for that iteration.

I use Beta and Dirichlet distributions for prior distributions for each parameter. Assuming a priori indepen-

dence of the appropriate parameters, I factor $p(\theta)$ into six independent Beta distributions and four independent Dirichlet distributions.[16] Parameters are then drawn from the appropriate posterior distributions, using only people in that particular behavioral group, listed below:

[16]In the seven-point-scale model, δ and γ estimates are drawn from a Dirichlet distribution (one for each behavioral group), rather than the Beta distribution for the δ estimate in the five-point-scale model. In the three-point-scale model, no δ parameter is estimated.

$$p(\alpha_1, 1 - \alpha_1) = \text{Beta}[(N - \sum_{i=1}^{N} A_i + 1), (\sum_{i=1}^{N} A_i + 1)]$$

$$p(\delta_1, 1 - \delta_1) = \text{Beta}[(\frac{\sum_{i=1}^{N} B_i}{3} + 1), (N - \frac{\sum_{i=1}^{N} B_i}{3} + 1)]$$

$$p(\varphi_2, \alpha_2, 1 - \varphi_2 - \alpha_2) = \text{Dirichlet}[(\frac{\sum_{i=1}^{N} C_i}{3} + \frac{2}{3}),$$

$$(\frac{\sum_{i=1}^{N} R_i}{3} + \frac{2}{3}),$$

$$(N - \frac{\sum_{i=1}^{N} C_i}{3} + \frac{2}{3} - \frac{\sum_{i=1}^{N} R_i}{3} + \frac{2}{3})]$$

$$p(\delta_2, 1 - \delta_2) = \text{Beta}[(\frac{\sum_{i=1}^{N} B_i}{3} + \frac{2}{3}), (N - \frac{\sum_{i=1}^{N} B_i}{3} + \frac{2}{3})]$$

$$p(\varphi_3^{(pre1)}, \alpha_3^{(pre1)}, 1 - \varphi_3^{(pre1)} - \alpha_3^{(pre1)}) = \text{Dirichlet}[(\sum_{i=1}^{N} H_i I(E_i = 1) + \frac{2}{6}),$$

$$(N^{(E_i=1)} - \sum_{i=1}^{N} F_i I(E_i = 1) - \sum_{i=1}^{N} H_i I(E_i = 1) + \frac{2}{6}),$$

$$(\sum_{i=1}^{N} F_i I(E_i = 1) + \frac{2}{6})]$$

$$p(\varphi_3^{(pre2)}, \alpha_3^{(pre2)}, 1 - \varphi_3^{(pre2)} - \alpha_3^{(pre2)}) = \text{Dirichlet}[(\sum_{i=1}^{N} H_i I(E_i = 2) + \frac{2}{6}),$$

$$(N^{(E_i=1)} - \sum_{i=1}^{N} F_i I(E_i = 2) - \sum_{i=1}^{N} H_i I(E_i = 2) + \frac{2}{6}),$$

$$(\sum_{i=1}^{N} F_i I(E_i = 2) + \frac{2}{6})]$$

$$p(\alpha_3^{(post)}, 1 - \alpha_3^{(post)}) = \text{Beta}[(\sum_{i=1}^{N} H_i - \sum_{i=1}^{N} M_i I(H_i = 1) + \frac{2}{6}), (\sum_{i=1}^{N} M_i I(H_i = 1) + \frac{2}{6})]$$

$$p(\delta_3, 1 - \delta_3) = \text{Beta}[(\frac{\sum_{i=1}^{N} B_i}{3} + \frac{2}{3}), (N - \frac{\sum_{i=1}^{N} B_i}{3} + \frac{2}{3})]$$

$$p(\tau_3, 1 - \tau_3) = \text{Beta}[(\sum_{i=1}^{N} I(E_i = 1) + 1), (\sum_{i=1}^{N} I(E_i = 2) + 1)]$$

$$p(\pi_1, \pi_2, \pi_3) = \text{Dirichlet}[(\sum_{i=1}^{N} G_i = 1 + 2), (\sum_{i=1}^{N} G_i = 2 + 2), (\sum_{i=1}^{N} G_i = 3 + 2)]$$

The formulas look complicated, but they amount to counting the number of response patterns that follow the particular rule the parameter pertains to. This is most obvious in the case of the pi_j estimates, which are a

Dirichlet draw from the count of each person assigned to that group for the current iteration. Similarly, the estimate for $alpha_1$, $(N - \sum_{i=1}^{N} A_i + 1)$, is simply the total number of people in Group 1 (opinion holders), N, minus the number of people who disagree, $\sum_{i=1}^{N} A_i$, plus 1 (the prior). Since opinion holders can only agree or disagree, this gives the count of people who agree, which is what α_1 indicates.

As noted above the priors are functionally equivalent to adding two people to each behavior group and splitting up their behavior within those groups. Hill and Kreisi find that changes in these priors have minimal effects on the actual estimates. The greatest instability is observed in the durable changers group, which is often small and therefore more susceptible to the influence priors.

7.4 Convergence

The model iterates until it reaches a stable distribution, from which draws were taken to capture the posterior distribution. Functionally, I used five chains, each with 2500 iterations. I discarded the first 500 iterations to eliminate estimates generated prior to convergence. Diagnostics suggest that, for most questions, parameter estimates converged well before the 500th iteration.

8 Appendix B: Questions

The table below lists the 544 questions analyzed here, actual question wording, how a question was recoded (if at all), the number of response options (2, 3, 4, 5, or 7) included in each model, and the structural and content tags associated with the question.

References

Achen, Christopher H. 1975. "Mass Political Attitudes and the Survey Response." *The American Political Science Review* 69 (4). [American Political Science Association, Cambridge University Press]: 1218–31. https://doi.org/10.2307/1955282.

Alwin, Alwin, Duane F. 2007. *Margins of Error: A Study of Reliability in Survey Measurement.* Hoboken, N.J.: John Wiley & Sons.

Ansolabehere, Stephen, Jonathan Rodden, and James M. Snyder. 2008. "The Strength of Issues: Using Multiple Measures to Gauge Preference Stability, Ideological Constraint, and Issue Voting." *American Political Science Review* 102 (2): 215–32. https://doi.org/10.1017/S0003055408080210.

Baker, Wayne. 2004. *America's Crisis of Values: Reality and Perception.* Princeton, New Jersey: Princeton University Press.

Bellah, Robert N., Richard Madsen, William M. Sullivan, Ann Swidler, and Steven M. Tipton. 1985. *Habits of the Heart, with a New Preface.* Berkeley: University of California Press. https://www.ucpress.edu/book/9780520254190/habits-of-the-heart-with-a-new-preface.

Boutyline, Andrei, and Stephen Vaisey. 2017. "Belief Network Analysis: A Relational Approach to Understanding the Structure of Attitudes." *American Journal of Sociology* 122 (5). The University of Chicago Press: 1371–1447. https://doi.org/10.1086/691274.

Converse, Philip E. 1964. "The Nature of Belief Systems in Mass Publics (1964)." In *Ideology and Discontent*, edited by D. E. Apter, 18:206–61. New York: Free Press. http://www.tandfonline.com/doi/abs/10.1080/08913810608443650.

Converse, Philip E., and Gregory B. Markus. 1979. "Plus ca Change...: the New CPS Election Study Panel." *The American Political Science Review* 73 (1). [American Political Science Association, Cambridge University Press]: 32–49. https://doi.org/10.2307/1954729.

DiMaggio, Paul. 1997. "Culture and Cognition." *Annual Review of Sociology* 23: 263–87.

Erikson, Robert S. 1979. "The SRC Panel Data and Mass Political Attitudes." *British Journal of Political Science* 9 (1). Cambridge University Press: 89–114.

Freeder, Sean, Gabriel S. Lenz, and Shad Turney. 2019. "The Importance of Knowing 'What Goes with What': Reinterpreting the Evidence on Policy Attitude Stability." *The Journal of Politics* 81 (1): 274–90. https://doi.org/10.1086/700005.

Gross, Neil. 2009. "A Pragmatist Theory of Social Mechanisms." *American Sociological Review* 74 (3). SAGE Publications Inc: 358–79. https://doi.org/10.1177/000312240907400302.

Halpern-Manners, Andrew, and John Robert Warren. 2012. "Panel Conditioning in Longitudinal Studies: Evidence from Labor Force Items in the Current Population Survey." *Demography* 49 (4): 1499–1519. https://doi.org/10.1007/s13524-012-0124-x.

Hill, Jennifer L. 2001. "Accommodating Missing Data in Mixture Models for Classification by Opinion-Changing Behavior." *Journal of Educational and Behavioral Statistics* 26 (2). American Educational Research Association: 233–68. https://doi.org/10.3102/10769986026002233.

Hill, Jennifer L., and Hanspeter Kriesi. 2001a. "An Extension and Test of Converse's "Black-and-White" Model of Response Stability." *The American Political Science Review* 95 (2). [American Political Science Association, Cambridge University Press]: 397–413.

———. 2001b. "Classification by Opinion-Changing Behavior: A Mixture Model Approach." *Political Analysis* 9 (4). [Oxford University Press, Society for Political Methodology]: 301–24.

Hitlin, Steven, and Jane Allyn Piliavin. 2004. "Values: Reviving a Dormant Concept." *Annual Review of Sociology* 30 (1): 359–93. https://doi.org/10.1146/annurev.soc.30.012703.110640.

Hout, Michael, and Orestes P. Hastings. 2016. "Reliability of the Core Items in the General Social Survey: Estimates from the Three-Wave Panels, 2006–2014." *Sociological Science* 3 (November): 971–1002. https://doi.org/10.15195/v3.a43.

Howe, Lauren C., and Jon A. Krosnick. 2017. "Attitude Strength." *Annual Review of Psychology* 68 (1): 327–51. https://doi.org/10.1146/annurev-psych-122414-033600.

Hunter, James Davison. 2000. *The Death of Character : Moral Education in an Age Without Good or Evil*. 1st ed. New York : Basic Books, c2000. https://find.library.duke.edu/catalog/DUKE002791044.

Hunzaker, M.B. Fallin, and Lauren Valentino. 2019. "Mapping Cultural Schemas: From Theory to Method." *American Sociological Review* 84 (5). SAGE Publications Inc: 950–81. https://doi.org/10.1177/0003122419875638.

Imai, Kosuke, and Dustin Tingley. 2012. "A Statistical Method for Empirical Testing of Competing Theories: EMPIRICAL TESTING OF COMPETING THEORIES." *American Journal of Political Science* 56 (1): 218–36. https://doi.org/10.1111/j.1540-5907.2011.00555.x.

Inglehart, Ronald. 1985. "Aggregate Stability and Individual-Level Flux in Mass Belief Systems: The Level of Analysis Paradox." *American Political Science Review* 79 (1): 97–116. https://doi.org/10.2307/1956121.

Joas, Hans. 1996. *The Creativity of Action*. Chicago: University of Chicago Press.

Judd, Charles M., and Michael A. Milburn. 1980. "The Structure of Attitude Systems in the General Public: Comparisons of a Structural Equation Model." *American Sociological Review* 45 (4): 627. https://doi.org/10.2307/2095012.

Kiley, Kevin, and Stephen Vaisey. 2020. "Measuring Stability and Change in Personal Culture Using Panel Data." *American Sociological Review* 85 (3). SAGE Publications Inc: 477–506. https://doi.org/10.1177/0003122420921538.

Kinder, Donald R., and Nathan P. Kalmoe. 2017. *Neither Liberal nor Conservative: Ideological Innocence in the American Public*. Chicago: University of Chicago Press.

Lizardo, Omar. 2017. "Improving Cultural Analysis: Considering Personal Culture in Its Declarative and Nondeclarative Modes." *American Sociological Review* 82 (1). SAGE Publications Inc: 88–115. https://doi.org/10.1177/0003122416675175.

Lizardo, Omar, and Michael Strand. 2010. "Skills, Toolkits, Contexts and Institutions: Clarifying the Relationship Between Different Approaches to Cognition in Cultural Sociology." *Poetics* 38 (2): 205–28. https://doi.org/10.1016/j.poetic.2009.11.003.

Martin, John Levi. 2010. "Life's a Beach but You're an Ant, and Other Unwelcome News for the Sociology of Culture." *Poetics* 38 (2): 229–44. https://doi.org/10.1016/j.poetic.2009.11.004.

Miles, Andrew. 2015. "The (Re)Genesis of Values: Examining the Importance of Values for Action." *American Sociological Review* 80 (4): 680–704. https://doi.org/10.1177/0003122415591800.

Mills, C. Wright. 1940. "Situated Actions and Vocabularies of Motive." *American Sociological Review* 5 (6): 904. https://doi.org/10.2307/2084524.

Oh, Jeong Hyun, Sara Yeatman, and Jenny Trinitapoli. 2019. "Data Collection as Disruption: Insights from a Longitudinal Study of Young Adulthood." *American Sociological Review* 84 (4): 634–63. https://doi.org/10.1177/0003122419859574.

Page, Benjamin I., and Robert Y. Shapiro. 1992. *The Rational Public*. Chicago: University of Chicago Press. https://press.uchicago.edu/ucp/books/book/chicago/R/bo3762628.html.

Perrin, Andrew J., and Katherine McFarland. 2011. "Social Theory and Public Opinion." *Annual Review of Sociology* 37 (1): 87–107. https://doi.org/10.1146/annurev.soc.012809.102659.

Rawlings, Craig M. 2020. "Cognitive Authority and the Constraint of Attitude Change in Groups." *American Sociological Review* 85 (6). SAGE Publications Inc: 992–1021. https://doi.org/10.1177/0003122420967305.

Scott, Marvin B., and Stanford M. Lyman. 1968. "Accounts." *American Sociological Review* 33 (1): 46. https://doi.org/10.2307/2092239.

Swidler, Ann. 1986. "Culture in Action: Symbols and Strategies." *American Sociological Review* 51 (2): 273. https://doi.org/10.2307/2095521.

———. 2001. *Talk of Love: How Culture Matters*. Chicago: University of Chicago Press.

Taylor, Marylee C. 1983. "The Black-and-White Model of Attitude Stability: A Latent Class Examination of Opinion and Nonopinion in the American Public." *American Journal of Sociology* 89 (2): 373–401. https://doi.org/10.1086/227870.

Vaisey, S., and O. Lizardo. 2010. "Can Cultural Worldviews Influence Network Composition?" *Social Forces* 88 (4): 1595–1618. https://doi.org/10.1353/sof.2010.0009.

Vaisey, Stephen. 2009. "Motivation and Justification: A Dual⊠Process Model of Culture in Action." *American Journal of Sociology* 114 (6): 1675–1715. https://doi.org/10.1086/597179.

Vaisey, Stephen, and Kevin Kiley. n.d. "A Model-Based Method for Detecting Persistent Cultural Change Using Panel Data." SocArXiv. Accessed January 14, 2021. https://doi.org/10.31235/osf.io/ec46t.

Vaisey, Stephen, and Omar Lizardo. 2016. "Cultural Fragmentation or Acquired Dispositions? A New Approach to Accounting for Patterns of Cultural Change." *Socius* 2 (January). SAGE Publications: 2378023116669726. https://doi.org/10.1177/2378023116669726.

Zaller, John, and Stanley Feldman. 1992. "A Simple Theory of the Survey Response: Answering Questions Versus Revealing Preferences." *American Journal of Political Science* 36 (3). [Midwest Political Science Association, Wiley]: 579–616. https://doi.org/10.2307/2111583.

Zaller, John R. 1992. *The Nature and Origins of Mass Opinion*. Cambridge Studies in Public Opinion and Political Psychology. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511818691.