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

Kevin Kiley, Duke University[†]

2/8/2021

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 inconsistent 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 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, question-order and -wording effects, the effects of psychological primes, the complexity of culture, and the general limitations of human cognition to argue that most people do not carry around fixed beliefs (Converse 1964; Martin 2002, 2010; Perrin and McFarland 2011). Instead, they argue, people carry around diverse considerations on issues of public interest and construct opinions in the interview or survey setting, drawing on recent considerations from their social environment (Swidler 2001; Zaller 1992). In contrast, other researchers point to the stability of attitudes in the aggregate, both aggregating across questions within people and aggregating the same question across a population; the infrequency of durable changes in attitudes over time; and the ability of specific attitudes to predict a range of behaviors over time as evidence that people carry at least some stable latent attitudes or dispositions, even if these positions are obscured by measurement error (Ansolabehere, Rodden, and Snyder 2008; Inglehart 1985; Kiley and Vaisey 2020; Vaisey 2009).

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 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, and a small group of people making durable change (Freeder, Lenz, and Turney 2019; Hill and Kriesi 2001a; Zaller 1992). 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," a position more consistent with advances in psychology distinguishing between strong attitudes, which govern behavior and thought, and weak attitudes, which do not (Howe and Krosnick 2017). 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 implications for the role of attitudes and beliefs in social behavior. If people do not carry around fixed beliefs, and if their dispositions and behavior are swayed by temporary influences, there is substantial room for contemporary social structures, opinion leaders, and situations to shape attitudes, and explanations for attitudes should be rooted in contemporary social structures, as pragmatist theories suggest (Gross 2009; Joas 1996). If people hold attitudes that are relatively impervious to social influence, these attitudes would be more likely to shape patterns of behavior, affiliation, and belief over time, and explanations for attitudes should be located in people's backgrounds, rather than contemporary social structures. 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 formalize three models of attitude behavior and 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 questions demonstrating widespread response stability (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. Overall, rates of vacillating change tend to exceed rates of stable attitudes, and high rates of ambivalence are 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 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 the social sciences 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 over time. 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 reasoning, which tends to bring to the surface the breadth of contradictory considerations that people have internalized. Fixed-choice survey questions tend to invoke practical reasoning, 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 "stable attitude" theories suggest a model in which people are 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. On some days, a moral relativist might "agree" that morals are a personal matter and some days might "strongly agree," 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, and the person will avoid them, even if they 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; 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 very 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 a line of public opinion scholarship. In his Receive-Accept-Sample model, Zaller (1992; 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 they 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" (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. Because considerations are called up in a haphazard way, which spe-

cific considerations manifest at any interview can vary, and people's assessments can be influenced by question wording, question order, psychological primes, and changes in their information environment, such as what was on the news recently.

Neither the RAS model nor the cultural cognition model suggest that people lack considerations. Both models argue that people have internalized heterogeneous considerations that make them express differing, often strongly felt, opinions over time. Similarly, neither model prohibits people from holding strong contradictory beliefs – support for more government spending, lower taxes, and a balanced budget, for example – that they present consistently over time. But these models suggest that on questions that call forth these conflicts, people will give inconsistent responses as some considerations are pushed to the foreground each time.

These "inconsistent attitude" theories 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 be torn between their considerations and 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 disagree. Still other times, maybe after hearing a counter argument, they say might say they agree. 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 in the long term.

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 that inform their behavior, and at other times appear to behave as if they do not. And this suggests a path forward.

Researchers invoke Converse's work (1964; Converse and Markus 1979) 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 held opinions and others did not. The population did not consist 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).

In recent decades, psychological research has distinguished between strong attitudes, which are "resistant to change, stable over time, influential on cognition, and influential on action" and weak attitudes, which are not (Howe and Krosnick 2017: p. 329). This research suggests that attitude strength is a multi-dimensional concept with diverse antecedents, and that people vary widely in the strength they attach to any particular attitude.

These approaches contrast with the measurement error model, which relies on an assumption that errors are a function of the question being asked, not the people answering questions, meaning no individual characteristics should predict instability. While initial assessments argued that nothing predicted stability (Achen 1975; Erikson 1979), subsequent research finds that individual characteristics are predictive of stability in political views. For example, people who have higher political awareness as measured by factual information about the political space tend to hold more stable political attitudes (Freeder, Lenz, and Turney 2019; Zaller 1992).

The underpinnings of Converse and Zaller's models are in no way confined to political beliefs, and there is 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" model of attitude consistency can address theoretical problems with "latent belief" models, specifically an unclear articulation 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 articulating these theoretical positions than concrete ones. But Ansolabehere and colleagues (2008) find that even when aggregated, these kinds of general attitudes are less consistent than more specific beliefs regarding policy positions.

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. While there are undoubtedly nondeclarative forms of culture that shape behavior over time but would be difficult to capture in a survey (Lizardo 2017), there is no reason to expect that aggregating more questions would capture these aspects of culture better, given their absence of symbolic content.

A "some do, some do not" model does not need the concept of latent beliefs to explain the fact that attitudes demonstrate low constraint (Converse 1964; Boutyline and Vaisey 2017) but are relatively stable in the aggregate (Ansolabehere, Rodden, and Snyder 2008). Psychological research demonstrates that when people decide that an issue is important to them, they become more stable on that attitude and seek out other attitudes to reinforce it (Howe and Krosnick 2017). For example, a person who opposes abortion might recognize that they should also oppose government intervention in the economy because they see other people holding those two attitudes simultaneously (Goldberg and Stein 2018). But the extent to which they attach various attitudes, and which attitudes they attach, are going to vary from person to person, shaped by time, attention, and social networks. This will result in what Converse (1964) called "issue publics" – socially patterned groups of people who are stable on different but overlapping sets of issues. When more attitudes get aggregated, more of these "issue publics" get folded into the average, giving the scale greater consistency as inconsistent (weak) beliefs cancel out.

Finally, a "some do, some do not" model more closely aligns with work in cultural sociology suggesting that social structures – institutions, contexts, and public culture – are principally responsible for shaping attitude structure and constraint (Lizardo and Strand 2010; Martin 2002, 2010; Rawlings 2020). Martin (2002), drawing on Durkheim's formulation of "social facts" (1895), argues that consistent patterns of thought require external social supports. 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 (Zaller 1992).

2.2 Expectations

The preceeding discussion suggests that the question "do people have stable attitudes or do they lack stable attitudes?" is a false dichotomy, rooted in a outdated understanding of attitudes. Instead, we should expect that on any particular issue that is measured in a survey, some people will hold strong, stable attitudes, while other people will demonstrate weak, ambivalent attitudes. Importantly, the distinction between these opinion behaviors might only be aparent when we look at attitudes over time, as weak attitudes make themselves apparent through their inconsistency.

There are a handful of other general expectations that emerge from the preceeding discussion. First,

most issues should demonstrate higher proportions of weak, vacillating attitudes than strong, stable attitudes. Questions asked in social science surveys tend to be asked because because there is public disagreement on the issue, meaning most people will tend to consume heterogeneous arguments in favor and against the issue in question.[1] Because screening out competing considerations takes cognitive work, time, and attention, which are limited resources, people are only likely to be stable on some issues while displaying ambivalence toward others.

Second, consistent with the issue publics model that roots stability in institutions and external scaffolding, people should demonstrate stability on related issues. For example, we should expect people who have stable attitudes on one question of general morality to have stable opinions on other questions of general morality, fostered through reflection on this topic, but this person might not have stable opinions on politics, religion, or even questions of specific moral prohibitions. Stable opinions on one political issue (opposition to abortion) might span to general political ideology (conservatism), but not necessarily to other political issues, such as opposition to government spending (Baldassarri and Gelman 2008).

Finally, we should expect very few examples of people durably changing opinions, even as we see high rates of change in the population. Previous work suggests that people tend to form dispositions early in life and carry them with them over time, rather than update attitudes over time (Kiley and Vaisey 2020; Vaisey and Lizardo 2016). As Howe and Krosnick note, "Attitudes that can be easily changed are weak and unlikely to shape behavior. The attitudes that most powerfully shape behavior are the hardest to change" (2017: p. 328).

3   Masuring Opinion Behaviors

The proceeding discussion suggests that the population comprises three groups of people with distinct attitude behaviors: some will behave as if they have settled (strong) attitudes, another will vacillate as if they construct (weak) opinions anew in each survey wave; and a third comprises people who change their attitudes over time or ambivalent people who become strong attitude holders in the survey window.

Most methods for measuring opinion behavior tend to conflate two of these groups or avoid quantifying their prevalence at all. Kiley and Vaisey's (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

---

[1]While there are likely issues that all people are stable on because they are taken-for-granted assumptions about the social world, for that same reason they are unlikely to be asked. Similar, there might be issues so far outside the domain of public discussion such that nobody holds strong stable beliefs, but these are also not likely asked.

groups or the proportion giving stable responses. Focusing on how many people change their responses over time, 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 that measure correlations over time assume that all members of the population follow the same process and treat questions as reflecting a continuous scale, rather than respecting the nominal structure of options as they are presented to the survey participant (Alwin 2007; Hout and Hastings 2016). For example, many methods treat a change on a five-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.

Finally, latent class methods can be used to deductively group cases such that their manifest variables are statistically independent (Bonikowski and DiMaggio 2016). Using such a method, we might be able to quantify how many people consistently give "agree" responses, "disagree" responses, and responses that change from one side to the other (similar to Taylor 1983). However, such models cannot adjudicate between processes that generated these patterns.

An ideal approach would formally specify the three different opinion behavior groups and quantify the probability that a respondent's set of responses came from each of these theoretical models. Such a model should respect the nominal structure of the data – especially the unique nature of "no opinion" and other ambivalent response categories. And it should allow for a comparison with the measurement error 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 Kreisi 2001a, 2001b; Hill 2001). Using such a model, I can formalize the three distinct theoretical processes that might have generated a person's set of responses and estimate the probability that each one did. I outline that model below.

3.1 Behavioral Groups

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."

Survey respondents answered this question three times over four years. How would the theoretical models outlined above expect people to behave?

Opinion Holders: The "stable attitude" theories discussed above suggest that because they have strong dispositions about an issue or because they are embedded in social structures that facilitate a single line of response over time, people should report being on the same side of an issue in successive interviews. With regard to the question above, a person might "strongly agree" in all three waves, or switch 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.[2] 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.[3]

If all people in the population behaved according to this model, we could model their responses over time with only two or three parameters. First, we would need to know what proportion of people agreed (α_1) or disagreed ($1 - \alpha_1$) with a statment, an unchanging disposition for each person, and then, in a particular wave, draw whether they give a strong (δ_1) or weak response.[4] Theoretically, α_1 would represent the balance of forces in society that create stable opinions in favor of an issue, such as socializing institutions.

Vacillating Changers: The second theoretical model of opinion formation – the inconsistent attitudes model – suggests a different attitude-formation process in which opinions are constructed anew in each wave on the basis of stored considerations and local influences. As noted above, several theories posit that people indiscriminately consume information and construct opinions by drawing on these considerations in the survey or interview setting – what Zaller (1992) calls "ambivalence." These people have likely heard some arguments in favor of homosexual marriage and some arguments against it, but have not taken much time to reconcile them. In the survey interview, they are asked to give an opinion and, upon reflection, call forth some of these considerations, which might be influenced by a variety of factors, such as recent discussions. These influences

---

[2]Measurement error arises from "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: p 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 harder to imagine that a person who supports gay marriage to give a response that indicates the opposite. There will likely be cases of "measurement error" where the response is coded incorrectly, but it is unlikely that these cases will be so prevalent as to produce the level of vacillation that we see in some questions.

[3]It is debatable whether "neither agree nor disagree" represents ambivalence, as I assume here, or a nuanced position where a person thinks that homosexuals should have the right to marry in some circumstances and not others. If it is the latter, then people should maintain this position over time. However, of the 301 people in the GSS panel 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.

[4]For simplicity, the model makes the assumption that people expressing opinions are equally likely (within opinion behavior groups) to give them strongly. This assumption can be relaxed.

on responses are local, reflecting the idea that people are disconnected from public debates on the issue and therefore should not be swayed by changes of elite opinion.

In these theories, people's dispositions are mostly irrelevant in shaping their responses in any particular wave. What matters instead is the balance of considerations in the social environment. We would only need to know the probability that a person gives a "no opinion" response (φ_2), the probability that, if they give an opinion it is some version of "agree" (α_2), and the probability that if they give an opinion they give it strongly (δ_2). In this model, these parameters reflect the distribution of considerations in society that people can draw on to give responses, including the social factors that might make admitting to no opinion acceptable, including the existence of an explicit "no opinion" response option. Because their opinions are statistically independent across waves, each person's response in each wave would be a multinomial draw of the five response options ("strongly agree", "agree", "neither agree nor disagree", "disagree", "strongly disagree") defined by those three parameters.

Durable Changers: The final group, which I call "Durable Changers," comprise two sets of behaviors. The first 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 waves 1 and 2 but "agree" or "strongly agree" for the remainder.[5] People in this group might be reacting to the Democratic party's evolution on gay marriage or have a close family member of friend come out as gay, leading this person to truly change their mind on the topic. Second, some people might 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, @halpernmanners2012)).

Because this group is defined by their switching behavior, parameters to summarize their behavior would revolve where they start in their responses and where they end, as well as when during the survey they change their attitude, which might also influence who changes their mind and how. The first parameter that describes their behavior, τ_3, captures the probability of change after the first wave (instead of the second wave). The second and third parameters, $\varphi_3^{(Pre1)}$ and $\varphi_3^{(Pre2)}$, quantify the probability that an opinion changer starts by saying they lack an opinion and changes after the first and second waves, respectively. A fourth parameter, $\alpha_3^{(post)}$ defines the probability that these people change to agreeing with the statement.

Two more parameters, $\alpha_3^{(Pre1)}$ and $\alpha_3^{(Pre2)}$, define the probability that someone who changes from

---

[5]It is plausible that in a long enough survey interval, a person might truly change their opinion twice. The surveys explored here tend to look at intervals of about four years, reducing my concern for this possibility. A change that reverts in two years does not seem very "durable," and seems to reflect more local influences than real change.

one side of the scale to the other starts by agreeing with the statement. Finally, a single parameter, δ_3, quantifies the probability that if a person gives an opinion, they do so strongly.

The defining feature of this group is that they change opinions once, so response patterns that switch twice or do not switch at all would be incompatible with this model.

3.2   Mixture Model

The key question for this analysis is what proportion of respondents display each of these behaviors in a given survey. 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.

The fundamental challenge of quantifying response patterns into these groups is that these group labels are unobserved. If we knew which group each person was in, we could count the number of opinion holders who say they agree with a statement to get an estimate of α_1. Conversely, if these parameters were known, we would have a simple time calculating which theoretical process was most likely to generate each response. However, while some patterns could only be generated by one of the models (e.g., "agree"-"disagree"-"agree" could only be generated by a vacillating changer), most could be generated by the vacillating changer group and one of the other groups, especially in a three-wave context.

What we can do is estimate a range of plausible values of these parameters ($\pi_1, \pi_2, \pi_3, \alpha_1, \delta_1, ..., \tau_3$), as well as estimates about the proportion of the sample giving each response, that are compatible with the data we observe and the assumptions made above about these three groups.

To do this, I use a Bayesian estimation technique known as the Data Augmentation algorithm. The algorithm has two basic steps. First, it samples the "missing data" – group membership – given a set of randomly drawn parameters.[6] In other words, it asks "given this person's responses, how common each group is, and how often members of each 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. It uses these group memberships to estimate the parameters, asking, 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 model draws a plausible value for each parameter from a specified posterior distribution.

---

[6] At the first iteration, 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 with different starting values 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.[7] The non-informative priors have the effect of generating slightly conservative parameter estimates in each iteration. 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.

The process – assign groups based on parameters and estimate parameters based on group assignments – iterates until it converges on a stationary distribution that summarizes plausible values of the parameters given the observed data and the model assumptions. Over these iterations, ambiguous patterns will be alternately assigned to different behavioral groups, producing the range of plausible values. I can then draw values from this distribution to summarize it. In practice, this range of plausible parameters is quite small. Full details of the estimation procedure are outlined in Appendix A.

The model Hill and Kreisi (2001b) outline is designed to estimate parameters on four waves of data, but the model is compatible with three waves of data using minor modifications. Comparing models estimated on four waves of data to the modified model estimated on only three of those waves dropping an internal wave (wave 2 or 3) produces almost identical parameter estimates, especially for group size parameters (π_1, π_2, and π_3), affirming the model's ability to generate valid inferences using three waves.

3.3 Missing Data

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

---

[7]Hill and Kreisi test a range of non-informative priors – comparable to between one and six additional people per behavioral group – and find little sensitivity to different non-informative priors.

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.

3.4 Gay Marriage Example

To demonstrate the model and the interpretation of its estimates, 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.

Table 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 an interval preresenting the central 95 percent of values most compatible with the data and the model assumptions.

[Table 1 about here.]

As a top-line result, the model estimates that between 49 and 53 percent of the sample gave stable responses (π_1) 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 (π_2). Between 3.28 and 6.45 percent of the sample made durable changes in attitudes during the survey window of 2006-2014 (π_3).

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 Table 1 are consistent with our intuitions about attitude change and stability. Peo-

ple 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 people who do not give strong responses changing between sides of the scale in successive waves. 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.[8] Durable changers more closely resemble opinion holders than vacillating changers in their opinion strength, 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 (bouncing between 1 and 2 or 4 and 5), while vacillating changers and durable changers would have latent positions in the middle of the distribution (bouncing between 2 and 3 or 3 and 4), and all groups would respond with similar average errors. While people with extremely large variation would have to be in the vacillating changer or durable changer group, they should be relatively rare so that, on average, the groups are more or less comparable.

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. Table @ref{tab:groups} shows the average within-person standard deviation for the three groups.

[Table 2 about here.]

Table @ref{tab:groups} 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

---

[8]The full panel collapses three three-wave panels, so wave 1 actually represents three two-year periods (2006-08, 2008-10, and 2010-12). This makes it difficult to draw conclusions about timing of change, though analyzing the panels separately can provide clarification.

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, choosing strong responses more often than we would anticipate if they had similar error ranges to opinion holders but had an average near the scale midpoint. Importantly, people who make durable change tend to go from one end of the spectrum to the other, a behavior very different than vacillating changers. 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" (2001a: p. 408).

4   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.[9] 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.[10]

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

---

[9]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 moderate position.

[10]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.

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)[11]; 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), 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.

Presenting a full list of all 544 questions and their wording would be cumbersome. A full list of questions, their wording, any modification, and group proportion estimates is available from the author.

---

[11] 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.

5   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.

5.1   Overall Pattern Proportions

As a first step in this analysis, Figure 1 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.

[Figure 1 about here.]

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 is less than 5 percent, with many of these questions showing rates statistically indistinguishable from 0.[12] 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 gives clear opinions over time, and there are questions on which very few people in the population articulate clear opinions over time. However, most questions fall somewhere in the middle, with some people

---

[12]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.

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 model approach separates distinct sets of opinion behavior and is not just partitioning respondents based on their positions on the scale. Figure 2 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.

[Figure 2 about here.]

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. In other words, vacillating changers tend to say that their opinions are weak and, as we would expect of weak opinions, switch sides of the scale. At the same time, respondents who vacillate still select the "strong" option, often quite frequently, suggesting 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 3 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 group, they are omitted from this comparison. Points below the line indicate questions where the vacillating changers have greater within-person variance than the comparison group (durable changers in the top row, opinion holders in the bottom).

[Figure 3 about here.]

Consistent with the gay marriage example presented above, Figure 3 shows that vacillating changers

uniformly have higher within-person standard deviations than durable changers for almost all questions with four or more response options, regardless of question structure, and almost uniformly lower within-person standard deviations.[13] 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 in the figure above is the fact that the pattern is consistent with almost all 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, stably clustering in the "strong" end of the scale.

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.

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 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. Appendix B outlines a regression approach to evaluate the competing influences of structure and content on attitude stability and change.

Overall, structure and content explain about 50 pecent of the variation in rates of vacillation. Ques-

---

[13]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).

tions with fewer response options demonstrate higher stability and less vacillation. In terms of content, questions about sexuality and religion demonstrate slightly higher rates of stability than questions about politics, and questions about general morality and civil liberties show higher rates of vacillation. Questions about race, foreign relations, work, crime, and family and gender are not more or less stable than general political questions once I control for question structure.

There are some questions that are poorly explained by these tags. Across data sets, partisan identification is more stable than other measures of general political ideology (position on liberal-conservative scale; role of government in economy; role of government in equal opportunity, etc.). 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, etc.) tend to vacillate more than predicted. This helps explain why the CCES, which includes many questions about specific policies, has a higher rate of stability than other political surveys.

Content explains less of the variation in rates of durable change, and there is no clear pattern in terms of question structure. However, the NSYR and the child panel of the ISPC, the two panels with with youngest average age, show higher rates of durable change than other panels, even controlling for questions strucutre and content. The largest residual in the durable change 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.

5.2 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 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, people should be stable on related questions but not necessarily stable on unrelated questions. If a measurement error model is correct, 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 4. 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 4 about here.]

Figure 4 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.

Among the remaining questions, closely related issues 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 mainstream political views – government spending, ideology, and political identification.

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, two sets of questions about abortion are 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. Stability in general political beliefs (liberal-conservative ideology, partisan identification) is completely

[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.

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.

5.3 General Discussion of Results

The results presented above suggest several general conclusions about the behavior of attitudes, at least the attitudes that get measured in social science surveys. First, it is wrong to say that people, in general, do 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 and in their extremity in ways that suggest they are subject to short-term considerations. Examination of the average within-person standard devisions for different groups suggest that people who vacillate are not just people at the middle of the scale who report with error; they often span the scale. 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 the similarity of the behavior 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, potentially located in people's heads, but the specific social features that produce consistency on a particular

---

[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.

issue 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, regardless of domain.

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 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.

5.4 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 with three theoretically grounded models of opinion behavior, I quantified the proportion of survey respondents who held stable opinions, who vacillated or gave ambivalent responses, and who formed an opinion or changed their opinion during the course of the

survey. The vacillating group tending to outnumber opinion holders. This set of results is strongly consistent with Zaller's model of ambivalence, but also suggests 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 from these results. It could be the case 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 that can be facilitated by adding group-predicting parameters to the fininte mixture model.

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 or ambivalence in many questions. Many survey respondents agree with a statement in one wave and disagree in another, suggesting they do not have "settled" opinions. 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 also have implications for methods involving attitudes. Researchers have invoked the weak predictive power of cultural attitudes to argue both for and against a link between attitudes and behaviors (Jerolmack and Khan 2014; Vaisey 2014). The high degree of weak attitudes – 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. "Attitude effects" are going to be shaped by what proportion of the population really holds an atittude, the distribution of stable attitude holds in the population, and the actual effect of the attitude on behavior. 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.

6   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 the four they use in their analysis. Comparing a four-wave model to a three-wave model using data from the NSYR and LSG, which have more than three waves, to a model with the same data, dropping an interior wave (wave 2 or 3), produces comparable estimates for most coefficients, especially behavioral group proportions (π_1, π_2, and π_3), suggesting the model behaves just as well with three waves.

The goal of the Data Augmentation algorithm is to draw a set of parameters, $p(\theta|Z)$, compatible with the observed data (Z) and the assumptions made in the model. It has two main steps: Drawing group membership indicators, given the parameter, and drawing parameters given the group membership indicators. I first outline the variables that go into the model and then outline those steps below.

6.1 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 variables 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. For the seven-point scale, I include an additional variable that counts "weak" responses.

6.2 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 and estimated parameters

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 requires 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.

6.3   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 independence 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.

6.4 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.

7 Appendix B: 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[17]; 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 do-

---

[17] 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.

mains. 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 in 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 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.

7.1 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.

[Figure 5 about here.]

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, which have fewer response options on average, 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.

7.1.1 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.

[Figure 6 about here.]

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.

In addition to the Nixon question discussed above, 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 election in 2000 was fair.

8 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.

Baldassarri, Delia, and Andrew Gelman. 2008. "Partisans Without Constraint: Political Polarization and Trends in American Public Opinion." *American Journal of Sociology* 114 (2): 408–46.

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.

Bonikowski, Bart, and Paul DiMaggio. 2016. "Varieties of American Popular Nationalism." *American Sociological Review* 81 (5). SAGE Publications Inc: 949–80. https://doi.org/10.1177/0003122416663683.

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.

Durkheim, Émile. 1895. *The Rules of the Sociological Method.* Translated by Sarah A. Solovay and John H. Mueller. New York: Free Press.

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.

Goldberg, Amir, and Sarah K. Stein. 2018. "Beyond Social Contagion: Associative Diffusion and the Emergence of Cultural Variation." *American Sociological Review* 83 (5). SAGE Publications Inc: 897–932. https:

//doi.org/10.1177/0003122418797576.

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.

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.

Jerolmack, Colin, and Shamus Khan. 2014. "Talk Is Cheap: Ethnography and the Attitudinal Fallacy." *Sociological Methods & Research* 43 (2): 178–209. https://doi.org/10.1177/0049124114523396.

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. 2002. "Power, Authority, and the Constraint of Belief Systems." *American Journal of Sociology* 107 (4): 861–904. https://doi.org/10.1086/343192.

———. 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.

———. 2014. "The 'Attitudinal Fallacy' Is a Fallacy: Why We Need Many Methods to Study Culture." *Sociological Methods & Research* 43 (2): 227–31. https://doi.org/10.1177/0049124114523395.

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. 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.

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.
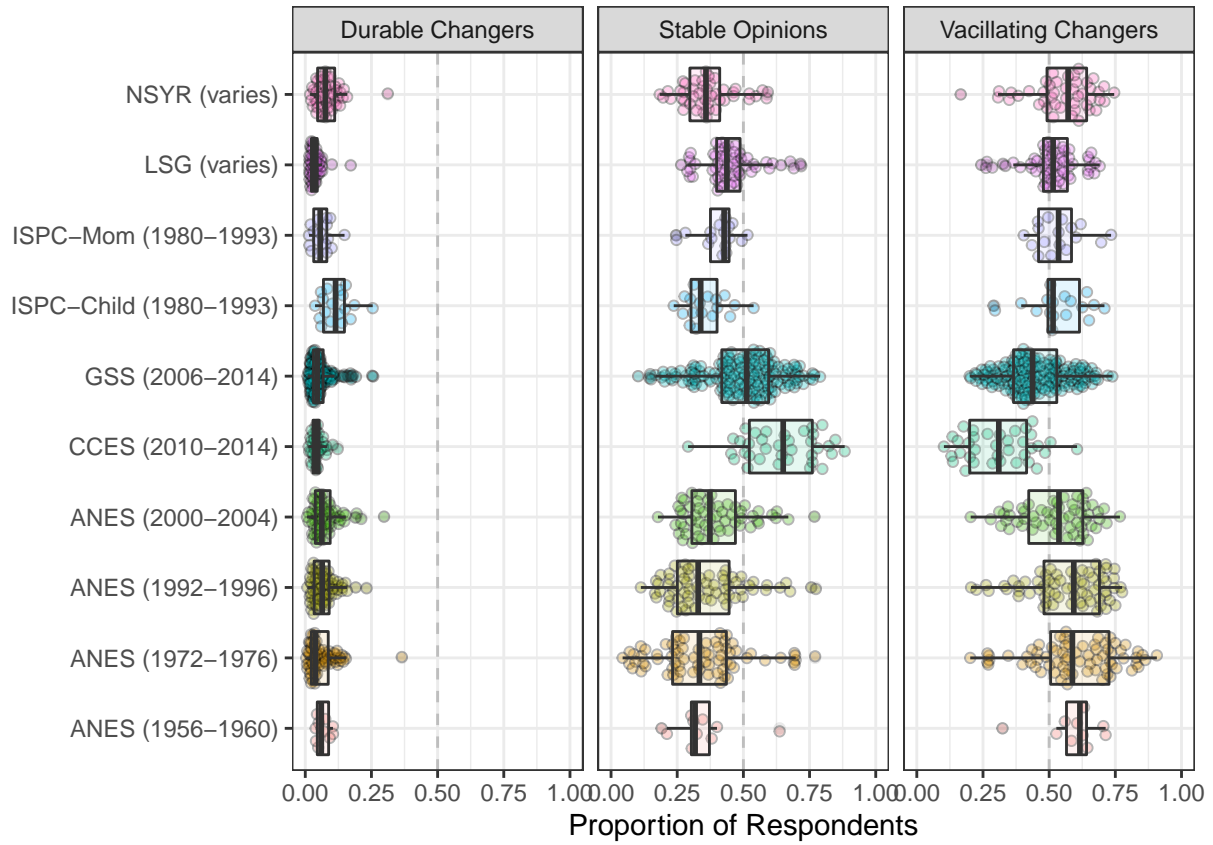
Figure 1: Mean proportion of respondents in each behavioral group, by question and data set. Posterior distribution estimated with 5 chains, 2500 iterations each, first 500 iterations of each chain discarded.
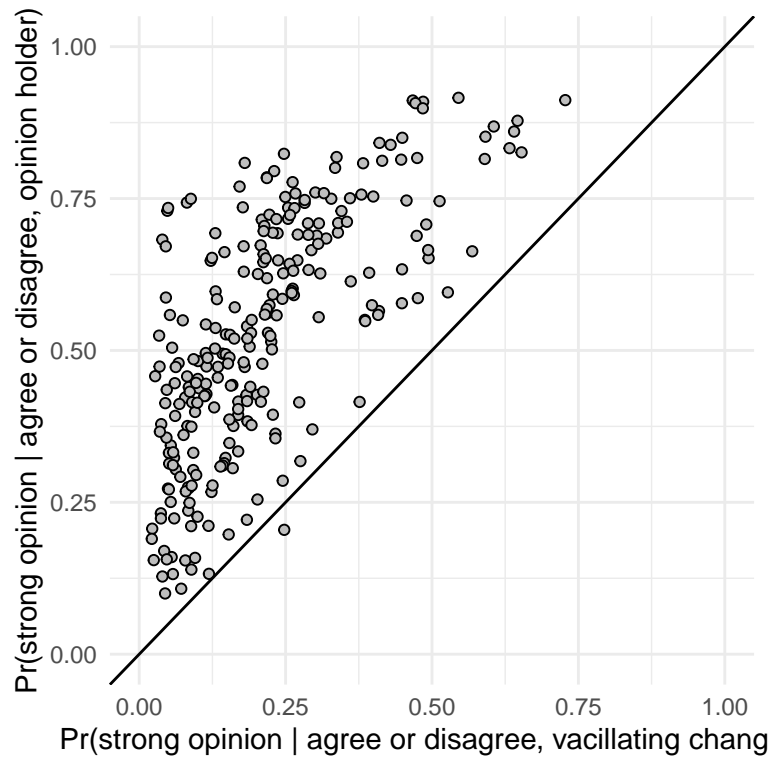
Figure 2: Mean estimated proportion of "strong" responses in opinion holder and vacillating changer behavioral group. Posterior distribution estimated with 5 chains, 2500 iterations each, first 500 iterations of each chain discarded.
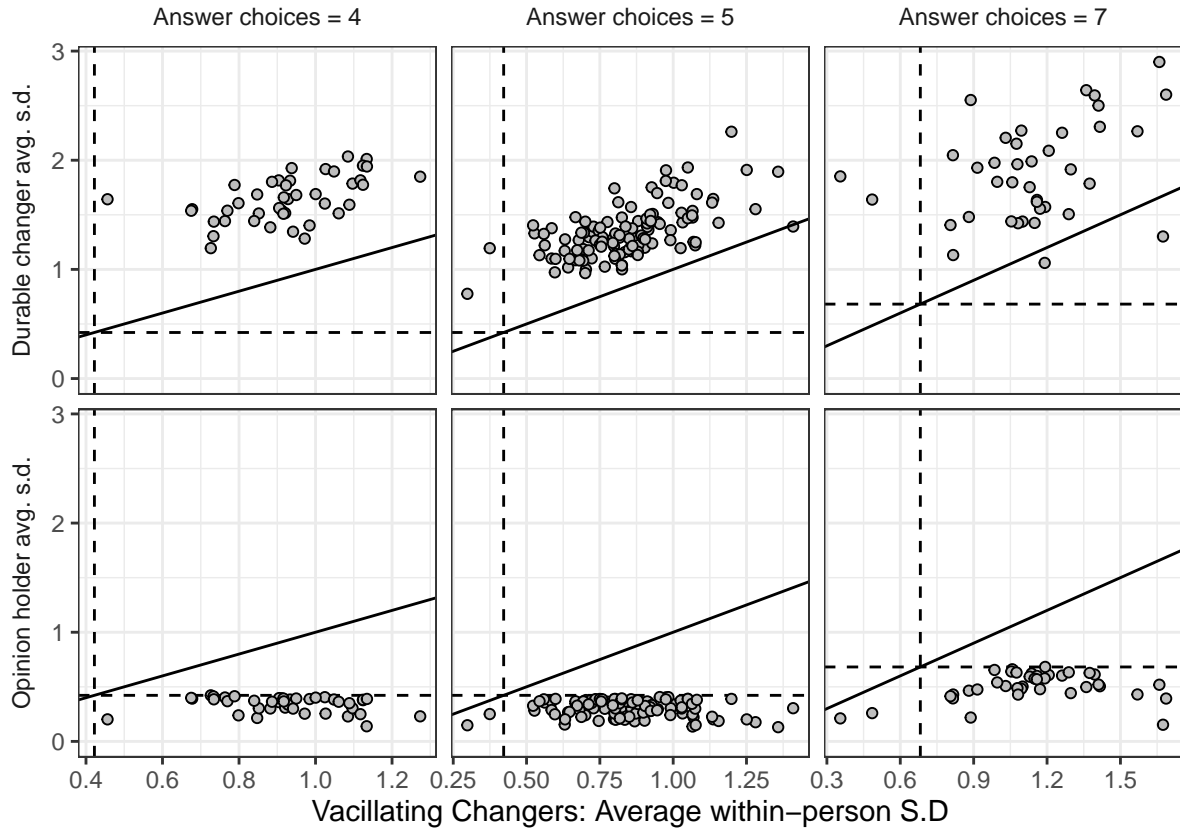
Figure 3: Average within-person standard deviation, by behavioral group and question response options. Solid black line indicates equal within-group standard deviations. Dashed line indicates largest average within-person standard deviation for stable opinion holders. Posterior distribution estimated with 5 chains, 2500 iterations each, first 500 iterations of each chain discarded.
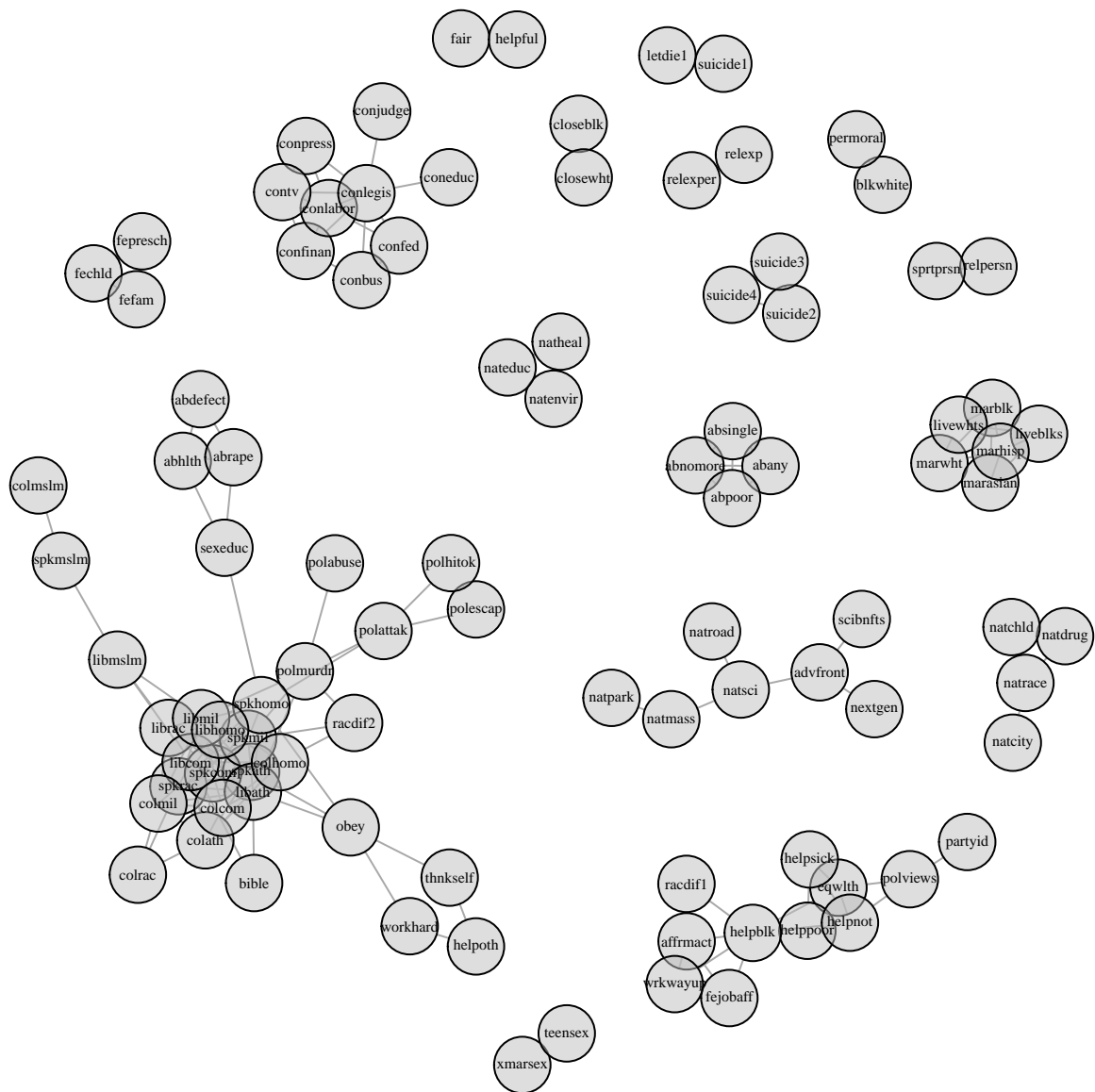
Figure 4: Network graph of pairwise correlations of co-occuring stability greater than .2. Probabilities generated from group assignments over 10,000 draws from posterior distribution.
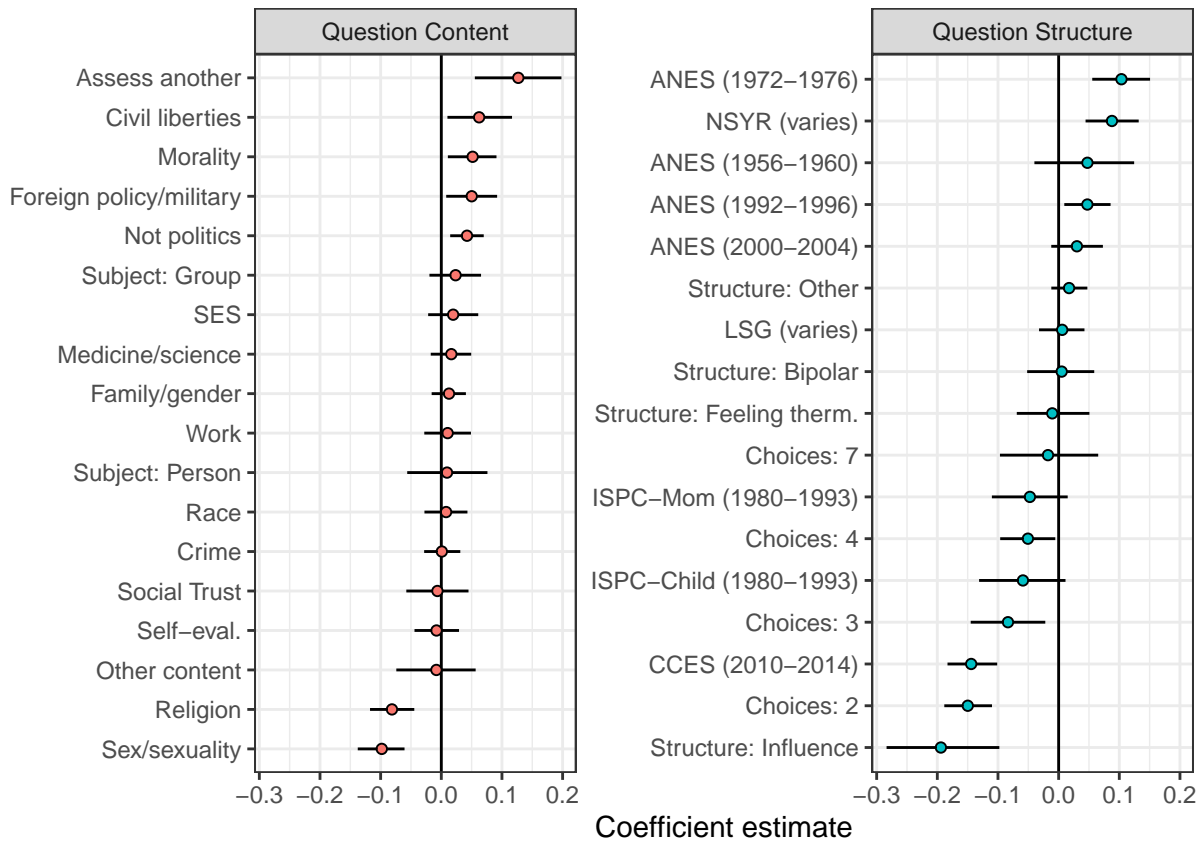
Figure 5: Coefficient estimates of regression of proportion of respondents of giving a vacillating response on question content and structure indictors. Confidence intervals generated from 10,000-iteration bootstrap.
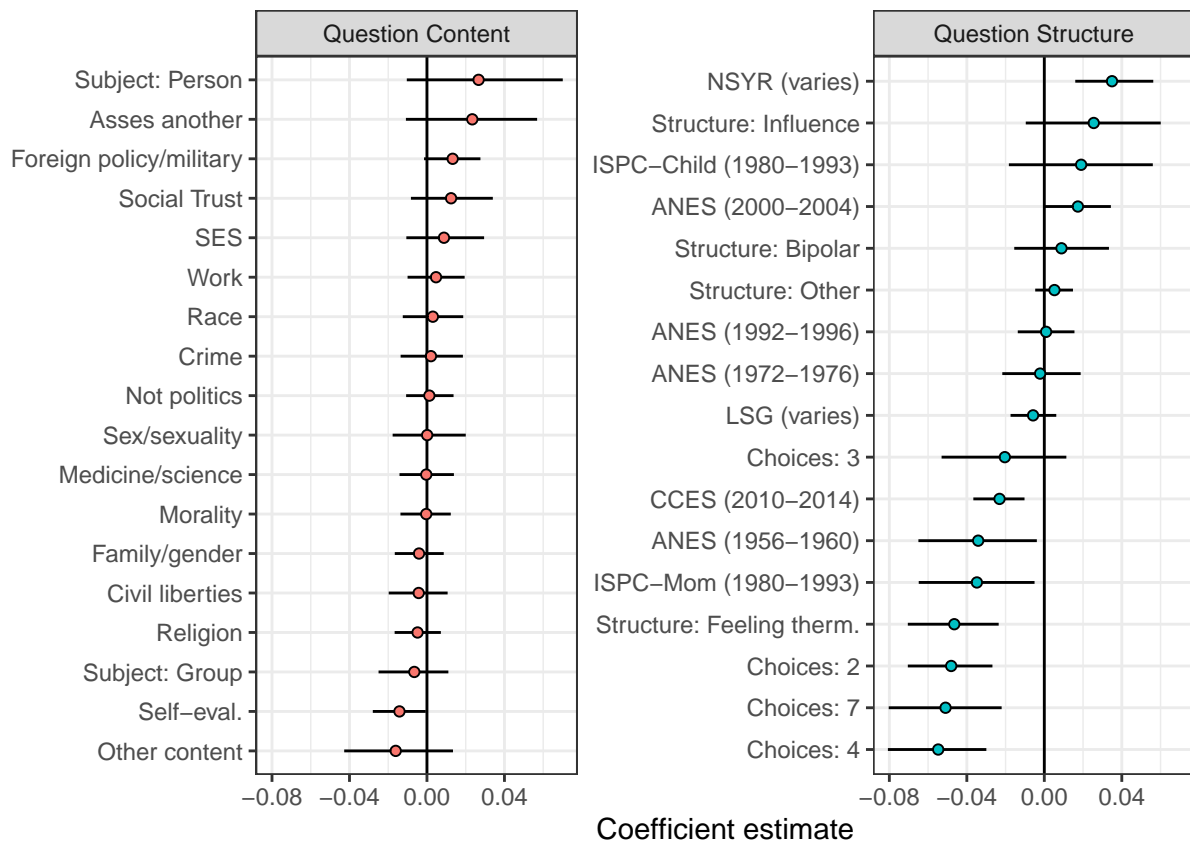
Figure 6: Coefficient estimates of regression of proportion of respondents of giving a durable change response on question content and structure indictors. Confidence intervals generated from 10,000-iteration bootstrap.

Table 1: Fininte Mixture Model parameter estimates for gay marriage question (2006-14 GSS).

Group	Parameter	Mean	Interval
Behavioral Groups	π_1	0.51	(0.49, 0.53)
	π_2	0.44	(0.42, 0.47)
	π_3	0.05	(0.03, 0.06)
Opinion Holders	α_1	0.40	(0.37, 0.42)
	δ_1	0.73	(0.71, 0.76)
Vacillating Changers	α_2	0.44	(0.41, 0.47)
	δ_2	0.25	(0.22, 0.28)
	φ_2	0.26	(0.23, 0.28)
Durable Changers	$\alpha_3^{(post)}$	0.40	(0.19, 0.6)
	$\alpha_3^{(pre1)}$	0.33	(0.2, 0.46)
	$\alpha_3^{(pre2)}$	0.09	(0, 0.28)
	δ_3	0.68	(0.54, 0.82)
	$\varphi_3^{(pre1)}$	0.57	(0.43, 0.71)
	$\varphi_3^{(pre2)}$	0.35	(0.02, 0.6)
	τ_3	0.68	(0.56, 0.81)

Table 2: Average within-group standard deviation, by opinion behavior group.

Behavioral Group	σ	Plausibility Interval (95 percent)
Opinion Holder	0.303	(0.295, 0.312)
Vacillating Changer	0.909	(0.885, 0.932)
Durable Changer	1.469	(1.308, 1.631)