(Almost) Nobody Changes

Comparing Theoretical Models of Cultural Change Using Panel Data

Kevin Kiley
Stephen Vaisey

Duke University

Introduction

Culture is an important part of social life but cultures are continuously evolving. In 1972, for example, over 40% of US adults supported a law outlawing interracial marriage. Three decades later, this opinion had become so uncommon that the question was removed from the US General Social Survey. How does this kind of cultural change happen?

Although rarely made explicit, sociological models of cultural change generally fall into two broad classes. The first is an *active updating* model that emphasizes the role of changing discourses, environments, and interactions. In this model, people continuously revise their views and practices based on the people, discourse, and events around them. The second is a *settled dispositions* model, which emphasizes the continuing influence of durable dispositions acquired early in life. In this model, adults' views and practices are relatively robust to changes in the social environment.

Both accounts have affinities with traditions in sociological theory. The active updating model is most compatible with a broadly pragmatist approach to social action, which claims that contemporary social environments and problems provoke people to adapt their views and make new meaning (e.g., Joas, Gross). The settled dispositions model has affinities with the Bourdieusian tradition that de-emphasizes (though does not ignore) the current environment in favor of the "past conditions of production" (e.g., Bourdieu, Vaisey and Lizardo)

In this paper, we make these two core models explicit, deduce some of their empirical implications, and derive a simple strategy for comparing them using panel data. We then apply this method to 184 items from the 2006-2014 General Social Surveys (GSS). Because of data limitations, we cannot speak to all types of cultural objects (e.g., music styles, baby names). We do, however, investigate a wide variety of opinions (like views on politics, free speech, race, and gender roles) and practices (like socializing at at bars or attending church) that are important in contemporary society.

Our analysis yields several results. First, we find that the majority of what appears to be individual-level change in belief or practice probably reflects short-term (non-persisting) change or measurement error rather than actual persistent change. Simply put, there is little evidence that U.S. adults are changing their beliefs or practices in lasting ways over this period of study. Second, the persistent change that we *do* see in the data is somewhat more concentrated among younger respondents. That is, on several items at least, it appears that younger adults are still in the process of acquiring dispositions and habits they will take into later life. Third, we find that changes in social behavior (e.g., church attendance, party membership, socializing) are more likely to be persistent than changes in private attitudes (e.g., political ideology). This suggests that interactional and institutional mechanisms may provide stronger support for lasting change than pressures for intrapsychic consistency.

These findings are generally more consistent with the settled dispositions model than with the active updating model. However, there is a pattern of exceptions and caveats that can help us understand how institutions and events shape the process of cultural change. We argue that there is a place for both models in our theory of cultural evolution but that we need more evidence on the circumstances under which each is more likely to apply.

In addition to their theoretical significance, our findings also have practical methodological implications. We argue that short duration panel studies of adults will be of limited utility unless they have a clear focus on the sorts of things that are likely to change in persistent ways over the study period. Otherwise, researchers will waste resources measuring "change" that turns out to be noise. In general terms, then, we suggest that social scientists would do well to conduct more panel studies of children and adolescents while using cross-sectional designs to study adults.

Theoretical Background

How does cultural change happen? Let us make the question concrete by imagining a person who answers the same question each year for several years. The question could be anything, but assume it is this GSS question: "Please tell me whether you strongly agree, agree, disagree, or strongly disagree [with this statement]: 'a working mother can establish just as warm and secure a relationship with her children as a mother who does not work.'"

How does the respondent formulate a response to that question each time, year after year? To make things as explicit as possible, we can write the data-generating process formally. Although this presentation may make it seem like we are assuming rationality or conscious deliberation, this way of writing the models makes no particular cognitive assumptions at all. We will explain this in greater detail below. For now, consider the following two models:

$$y_{it} = y_{it-1} + \nu_{it} \tag{1}$$

$$y_{it} = \mu_i + \nu_{it} \tag{2}$$

These models may seem similar at first glance, but they have very different implications. Equation 1 represents an *active updating* model and Equation 2 represents a *settled dispositions* model, as we explain below. Figure 1 (XXX) shows these models in graphical form, which helps highlight their differences. In the next two sections, we consider each of these models and briefly discuss their connections to sociological theory. In the section after that, we discuss the conditions under which each model is more likely to apply.

Active Updating Model

Equation 1 represents the active updating model. The respondent forms her answer by starting with what she said last time (y_{it-1}) and then incorporating any new considerations (ν_{it}). There is no need to remember responses from even earlier time points (e.g., y_{it-2}) because this information gets folded into the updated response each time.

More informally, this AUM posits an agent that is updating her views in the face of social experience. There are formal Bayesian ways of modeling updating but we need not rely on any assumptions of rationality, optimality, or conscious thought for this basic process to apply. Following Gross's (2009: 367) pragmatist account, we could instead regard this model as consistent with an "active and creative relation to the world" that "lead[s] actors to see themselves in new ways, to value different kinds of goods, and to become attached to problem solutions that they could not have imagined previously."

Regardless of whether the exact process is rational, heuristic, discursive, embodied, or any combination of these, the key notion is that the actor continues to be influenced by the (social) environment in ways that might lead to long-term, persistent change in beliefs, practices, and identities (Gross, DeGloma). As the person encounters new considerations throughout her life (e.g., adding working mothers to one's social network, seeing changing media representations of working mothers), she can continue to revise her views.

The AUM makes no assumption about the distribution of ν_{it}. Specifically, it does not assume that ν_{it} has an expected value of 0, either for any time t or for any person i. This leaves open the possibility of a population-wide shift in responses as many people react to the same changes in the environment.

Settled Dispositions Model

Equation 2 represents the settled dispositions model. Here each respondent begins the study period with a set tendency to respond to the question in a particular way (μ_i). Temporary considerations, like current events, can play a role in what response she gives at each time (part of the ν_{it}), but these considerations have no lasting impact beyond time t. As the right panel of Figure 1 makes clear, there is no mechanism by which a particular consideration can "propagate up" into the settled disposition and change the baseline for future responses. Such considerations are temporary influences only.

The SDM is most consistent with the Bourdieusian model of action that emphasizes "the past conditions of production" (XXX:XXX). In other words, updating happened, but it happened in the past, prior to the time of the study. In this sense, μ_i reflects the "habitus." We will say more about this below.

Like the AUM, the SDM does not require ν_{it} to be 0 in expectation at time t, but it does assume that ν_{it} has an expected value of 0 within an individual over time. That is, current considerations can move a person from their baseline temporarily, but there is a tendency to "bounce back" to that baseline. The SDM thus allows for population-wide shifts in beliefs, practices, or identities at a particular time (temporary period effects), but it assumes that within individuals these shifts will be erased over time as individuals return to their baselines.

Metaphorically, one can think of these models as travelers. In the active updating model, the traveler wakes up in the morning, walks in a random direction for a while, goes to sleep, and then wakes up the next day to pick up the journey where she left off. In the settled dispositions model, the traveler also travels in a random direction each day but wakes up to find herself in the same bed every morning.

The models thus differ fundamentally on their emphasis on the character of personal change. In the AUM, changes tend to persist because they propagate into future responses, shifting the baseline. In the SDM, changes tend to revert because each person returns to their existing baseline. We can therefore summarize the models' implications for individual change as follows: the AUM predicts *persisting change* whereas the SDM predicts *non-persisting change*. We consider the implications of these predictions below.

Combining the Models

Although our presentation above highlights differences between these models, they are not necessarily contradictory. There are at least two ways to understand how both models could apply without contradiction.

Age and Cohort Formation

The easiest way to reconcile the AUM and SDM is to posit that they operate sequentially over a person's life. To continue the metaphor above, one can imagine travelers wandering around for a while (active updating) before setting up their permanent home base (settled disposition). This is, in fact, the logic that underlies the notion of a "cohort effect." The idea of a cohort effect is that the contemporary environment influences the baseline response tendency when a person is young and then stabilizes and stays "baked in" for the remainder of one's life (Elder, Bartels and Jackman). Any notion of a settled disposition (including the Bourdieusian one) requires an earlier "formative period" where something like the active updating model was operative (Vaisey and Lizardo). To put it slightly differently, cohort formation requires an early period of persistent change followed by a later period characterized by either absolute stability (where the variance of ν_{it} is low) or by temporary, non-persisting changes that disappear as people revert to baseline.

This discussion of cohorts brings us close to the vast literature on age-period-cohort decomposition models, which is focused on statistical techniques to partition variance among sources of social change in repeated cross-sectional data (e.g., Yang and Land). Although this work is important, it is focused on modeling changes in population aggregates whereas we are concerned with modeling cultural change from the point of view of the same actor moving through time.

The two are related and have implications for each other, of course. Again, the idea of a cohort requires that some people (i.e., younger people) are influenced by the environment to make persistent change in ways that others (i.e., older people) are not. This implies a transition, perhaps a gradual one, from active updating to a settled disposition. A further connection between the two is that the world implied by the SDM can only change in the aggregate due to cohort replacement. That is, if persisting change is extremely rare among adults, the only way for persisting change to occur in the population is for older cohorts to die and be replaced with younger cohorts with (perhaps) different baselines. For our purposes, the most important idea here is that "cohortization" implies that the AUM might describe the data-generating process better among younger respondents whereas the SDM might describe it better among older respondents.[1]

Different Items, Different Processes

Another way to reconcile the AUM and the SDM is to consider that they apply to different sorts of questions a researcher might ask a respondent. We can illustrate this idea with (non-cultural) examples we understand well and then consider how less obvious cultural processes might be similar or different.

First, if we repeatedly ask a sample of women "how many children have you ever given birth to?", we know the data is generated by a process that looks like equation 1. Leaving measurement error aside, the response is a simple combination of how many children the respondent had last time we asked (y_{it-1}) and whether or not she has had any since (ν_{it}).

Second, if we ask repeatedly ask a sample of adults "how many full siblings did you have at age 16?", we know the true answer cannot change and therefore equation 2 is the best model. This is because μ_i refers to a fixed quantity and cannot change over time. Responses could still vary due to measurement error (e.g., misunderstanding the question) meaning that the variance of ν_{it} need not be 0. But the underlying reality is, by definition, a settled matter.

Finally, there are responses that have no stable component but are entirely functions of local circumstances. If we repeatedly ask a sample of people in the same city "did it rain yesterday?", their responses would not be described better by Equation 1 or Equation 2. In this case, everything that matters is in the contemporary circumstances (ν_{it}) and thus the terms that distinguish equation 1 from equation 2 (y_{it-1} or μ_i) are ignorable. (These types of items are not important for adjudicating our research questions but we still need to be aware of this possibility.)

What we want to determine, then, is how much a particular item resembles these different processes. Consider repeated requests for (dis)agreement with the prompt "a working mother can establish just as warm and secure a relationship with her children as a mother who does not work." Does the aggregate pattern of responses look more like an active updating process where a person might be changing her mind? Or does it look more like asking a person to report repeatedly on a settled matter?

When it comes to beliefs and behaviors, which items will demonstrate persistence likely depends on a number of factors. [REWRITE SHORT – 2 paragraphs max]

Toward Theoretical Synthesis

The active updating model and the settled dispositions model are different and they do have affinities with different theoretical traditions in sociology. Indeed, the difference in emphasis between environmental cues, creative meaning-making, and public discourse on the one hand and settled dispositions, "habitus," and cohort replacement on the other has been a basic theoretical tension in cultural sociology for more than a decade (Vaisey, Lizardo, Vaisey and Lizardo, etc.).

---

[1] We do not have space for a full discussion here, but we regard these approaches as complementary. Repeated cross-section approaches must make strong assumptions about individual processes but are able to use many years of data. A panel approach is better able to capture micro-level processes but at the cost of using only a few years of data; this requires making strong assumptions about the generalizability of the process across historical time.

One model may, in fact, be a better starting place than the other. Previous work using cross-sectional data suggests that settled dispositions are hugely important (Vaisey and Lizardo). Despite these findings, the evidence is still mixed. Therefore our goal here is neither to declare victory for one of the theoretical perspective nor to simply say that both "matter." Rather, our objective is to improve sociological models of cultural evolution by more precisely specifying when and where different types processes are at work. We believe achieving a better understanding of these processes will be relevant for many subfields of sociology as well as for other social science fields that study changes in beliefs and behaviors.

Research Questions and Expectations

With these considerations in mind, we ask the following questions:

First, *are patterns of cultural change generally better described by the active updating model or the settled dispositions model?* Previous work on cultural change using cross-sectional data has suggested that cohort effects are generally more important than period effects (Vaisey and Lizardo). This implies that, in the repeated measures data on adults we will use here, we should find that the settled dispositions model performs better on most items because cohort formation should be (mostly) complete.

Second, *is there evidence that younger respondents are doing more active updating than older respondents?* As we just mentioned, in a sample of adults, the possibility exists that cohort formation may be complete for most beliefs and behaviors before the study period (i.e., before age 18). However, if some cohort formation is still occurring among younger respondents, we should see that evidence consistent with the AUM is disproportionately located among younger respondents.

Third, *are there systematic differences in item content between questions that are better described by each model?* Items differ from each other in many ways and it would be challenging to come up with clear predictions for all possible types of items. In general, however, we expect (following Vaisey and Lizardo) that core beliefs about gender, family, race, and trust should be more consistent with settled dispositions because their formation should be already be complete for adults. Items that require public commitments such as changing political parties, socializing, or religious attendance and affiliation should be open to persistent change because there are social mechanisms (not just cognitive ones) to *maintain* the change. [ONE MORE SENTENCE? MAKE SURE THE SENTENCES IN "DIFF ITEMS, DIFF PROCESSES" PRESAGE THIS DISCUSSION.]

Analytic Strategy

To investigate these ideas empirically, we examine 184 survey items from the 2006-2014 General Social Survey panels in search of evidence in favor of an active updating model. This period of the GSS contains three different three-wave panels, each of which surveys a sample of adults three times over a four-year period (e.g., 2008-2008-2010). Three waves of data is the minimum amount needed to compare the predictions of the active updating model and the settled dispositions model. We discuss item selection below.

Statistical Model

Basic Models

Our two different models make different predictions in the three-wave panel context. The AUM makes the following prediction for wave 3:

$$E(y_{i3}) = y_{i2} \tag{3}$$

That is, the AUM predicts that a respondent's most recent response is the best available predictor of her next response. If change is persisting (i.e., if our metaphorical traveler wakes up in the same bed she went to sleep in last night), then our best guess is that she will be close to where we saw her last.

The SDM makes the following prediction:

$$E(y_{i3}) = \mu_i \tag{4}$$

Since the best estimate of μ_i is the mean of the respondent's two previous answers, we can rewrite the SDM prediction like this:

$$E(y_{i3}) = \frac{y_{i2} + y_{i1}}{2} \tag{5}$$

That is, the SDM predicts that the average response is the best predictor of the next response. If change is non-persisting (i.e., if our metaphorical traveler ends up back in her own bed every morning), then taking the average of the last two places we saw her will be our best guess about the location of her home base.

Combined Model

Both of the models above include y_{i2} as a predictor of y_{i3}, but only the settled dispositions model includes y_{i1} as a predictor as well, and its is treated as equally predictive as y_{i2}. Therefore, to test for evidence of active updating we use a model that evaluates whether y_{i1} carries any additional predictive power for y_{i3} once we control for y_{i2}. If the two previous estimates are equally predictive, then we can be relatively confident that the data we observe likely came from a settled dispositions model. However, if y_{i2} is a better predictor of y_{i3} than y_{i1}, we begin to observe evidence that at least some individuals in the population engage in active updating. We use the following non-linear model to estimate the relative influence of y_{i1} and y_{i2}:

$$E(y_{i3}) = \alpha + \phi\beta y_{i2} + (1 - \phi)\beta y_{i1} \tag{6}$$

Rather than estimate separate parameters for y_{i2} and y_{i1}, this model generates two parameter estimates of interest for each item in our analysis: β, which captures how well any combination of previous waves predicts a person's response at wave 3, and ϕ, the relative proportion of wave 3 explained by wave 2 compared to wave 1. If the Settled Dispositions Model is the preferred data-generating process for an item, then both y_{i2} and y_{i1} should be equally predictive of y_{i3}, and ϕ will equal .5, meaning the best estimate of wave 3 is some transformation of the mean of previous waves, consistent with Equation (5). In the active updating model is present in some respondents, then wave 2 will be more predictive than wave 1, and ϕ will increase converge with Equation (3) in certain circumstances.

Limitations of the Method

Three challenges limit our ability to evaluate the relative presence of settled dispositions and active updating models in the data set using our approach. The first is that our model is designed to allocate variance explained to each wave, rather than assign probabilities to each data-generating process. Because of this, few large persisting changes can inflate the *phi* estimate even if most individuals make small non-persisting changes. If we were willing to assume that persisting and non-persisting changes were drawn from the same distribution, we could make inferences about the relative proportion of people coming from each data generating process, but we have no reason to believe this assumption.

The second challenge is measurement error, which is by definition non-persisting change. As a result, estimates of ϕ will be biased toward .5 in the presence of measurement error. There is evidence that many of the items explored in our analysis are measured with significant error (Alwin, Hout and Hastings). On the other hand, previous studies of measurement error often conflate measurement error and non-persisting belief change,

meaning that while we might have good estimates for the combination of these two processes, we cannot separate them. Because of this issue of measurement error, it is unlikely that ϕ and β will be one for any item.

The third challenge of our analysis is that we focus exclusively on predicting wave 3. If for some reason individuals have a high likelihood of changing between waves 2 and 3, then our ability to predict responses at wave 3 with either wave will be limited and β will be low. Our model relies on the assumption that "persisting" change is relatively rare and that individuals who change between waves 1 and 2 do not also make persisting changes between waves 2 and 3. If the rate of active updating is so high that individuals make changes between each wave, then the model becomes indistinguishable from the settled dispositions model.

In addition to these two limitations, there are two forms of change that our model is not well designed to account for. The first is overall mean change. Since our model includes an intercept, changes that shift all responses toward one end of the scale are not accounted for in our ϕ estimates. The second is change in the variance of responses. If all individuals shift outward or inward toward the population mean but maintain their relative position in the overall distribution, this change will not enter into ϕ.

Despite these limitations, the model is capable of detecting the presence of persisting change even in the presence of high levels of measurement error. Because of this, it is best to think of our approach as seeking any evidence in favor of active updating, rather than allocating probabilities to each model. We can only detect whether there is evidence of persisting belief changes, and therefore whether there is any evidence that active updating is present.

Analysis Steps

Our analysis proceeds in three steps. We first evaluate the evidence in favor of the Active Updating model. To do this, we compare for each item the Bayesian Information Criteria (BIC) of a model estimated using (6) with a free estimate of ϕ to a model that constrains $\phi = .5$. If the model with the constraint if preferred, then there is little evidence that respondents engage in an active updating process with respect to that item during the window we study.

Next, for variables that do show at least some evidence of active updating ($\phi > 0.5$), we ask whether the persistent change is disproportionately concentrated among younger respondents. To test this, we re-estimate our original model and include a dummy term indicating whether respondents fall below an age cutoff. Rather than test a single age cutoff, we again use BIC comparisons to evaluate whether including the dummy variable improves the model fit using a cutoff of every age between 20 and 45. We test a range of cutoffs to ensure robustness of the overall pattern to specific ages. If the cutoff is too restrictive, there might not be enough cases to draw valid inferences. If there are strong differential effects of age in a narrow age range, a large cutoff might dilute the effects of age on active updating.

Finally, with these findings in hand, we consider whether there are any meaningful patterns in the relative distribution of evidence for AUM across variables.

Item Selection

We seek to test our model on as broad a range of items tapping attitudes, beliefs, and social behaviors as possible. To do so, we estimate the model in equation 6 on data from the three three-wave panel data sets of the General Social Survey, conducted from 2006 through 2014.

We sought measures of attitudes, beliefs, self-assessments, self-perceptions, and behaviors that were asked in three waves. These questions tended to come from the "core" of the GSS, a set of questions asked in each wave. Rotating topical modules asked during the panels were only asked in select waves or were not asked consistently to the same people over time, leaving too few cases with complete data to analyze.

Since our theoretical framework focuses on attitudes, beliefs, and social behaviors, we excluded from our analysis questions that focused on demographic characteristics (marital status, household size, region,

gender, race, ethnicity), work activity (employment status, income, hours worked, size of workplace), objective socioeconomic status (years of education and highest degree, home ownership), and evaluations of a respondent by the questioner. We also exclude questions that ask about an individuals' childhood. While testing these questions with our method is possible, they are beyond the scope of our theoretical framework. To organize our presentation of such a large number of variables, we follow Hout and Hastings (XXXX) and group questions into 15 categories based on their subject material. Questions in the same category tend to be asked in the same block during the survey and have the same structure, such as questions about confidence in institutional leaders, questions about public spending, and questions about social life.

We also follow Hout and Hastings (XXXX) in recreating some commonly used scales designed to capture attitudes about gender roles, access to abortion, and social trust. This includes Rossi's six-question scale of support for abortion and a seven-question scale which includes the question asking about abortion under any circumstances (abany). We use Smith's (1997) scale of "misanthropy" by combining questions about how helpful, fair, and trustworthy people are. We use four questions to create a scale of gender role attitudes (Cotter, Hermsen, and Vanneman 2011).

Like Hout and Hastings, we combine civil liberties items into six scales about the freedom of atheists, communists, militarists, racists, and, in the 2010-14 panel, Muslim clergymen. We combine four parallel questions about how frequently individuals socialize to create a "social life" scale. We combine four questions about support for suicide under different circumstances. We also created a scale of support for police use of violence against criminal suspects by averaging five binary questions about the conditions under which individuals support police use of violence.

In total, we test the model on 184 GSS items, including the composite scales.[2]

Results

Our model generates two parameter estimates of interest for each item in our analysis: β, which captures how well any combination of previous waves predicts a person's response at wave 3, and ϕ, the relative proportion of wave 3 explained by wave 2 compared to wave 1. If responses are generated through a Settled Dispositions Model, then ϕ will be .5. As the evidence of active updating increases, ϕ will increase toward 1. Both ϕ and β equaling 1 would indicate that all individuals who changed between waves 1 and 2 persisted in their change, that an item was measured with no measurement error, and there was no additional change between waves 2 and 3. This is a highly unlikely scenario.

Evidence for Active Updating

To evaluate the evidence in favor of the Active Updating model, we compare for all 184 items the Bayesian Information Criteria (BIC) of a model with a free estimate of ϕ to a model that constrains $\phi = .5$. If the model with the constraint if preferred, then there is little evidence that respondents engage in an active updating process with respect to that item.

Figure 1 plots the distribution of ϕ estimates for the 184 items evaluated in this analysis and the probabilities that the model without the constraint fits the data better.

The right side of figure 1 shows that while the majority of items prefer the free parameter, 72 items (about 40 percent of those examined) prefer the constraint, meaning these items show no evidence of active updating during the time frame examined here. While individuals might shift their response to these items in any particular wave because of measurement error or real ideological change, they will tend to revert to their previously held position over time. On the left side of the figure, we see that the majority of ϕ estimates fall between .5 and .6, meaning that wave 2 is only a slightly better predictor than wave 1 for most items. This

---

[2]To ensure that our estimates of ϕ are not simply artifacts of response scale construction, we estimate the model on coarsened versions of items, generated by collapsing responses to questions with more than three response options into scales of two or three response options. These results are reported in an appendix.
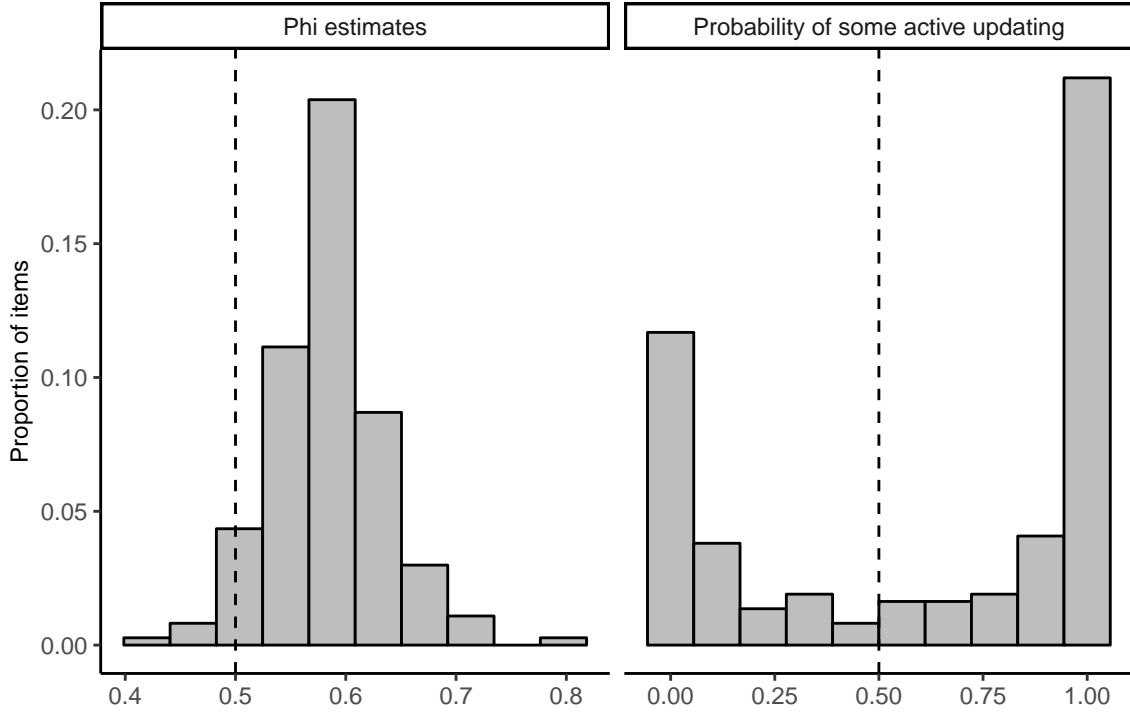
Figure 1: Distribution of probabilities that items show evidence of Active Updating.

suggests that if active updating is happening in these responses, it is relatively infrequent or small compared to temporary change and measurement error.

The group of items preferring the constraint includes a broad swath of items across a number of subject areas. It includes about half of the items about civil liberties, half of the items about confidence in institutional leaders, half of all questions about race and immigration, half the items about gender and family, and half the items about sexuality and abortion. On the other hand, most items about religion, suicide, politics and government, and public spending show evidence of active updating.

Figure 2 plots the ϕ and β estimates for items colored by whether they prefer the constraint or not, with a few items labeled. Items tend to prefer the $\phi = 0.5$ constraint for a combination of two reasons: because wave 1 and wave 2 have equal predictive power (ϕ is close to .5) or because the measure is so unpredictable (β is low) that neither wave 1 nor wave 2 has much predictive power. Included in the first group, where waves 1 and 2 are both predictive, are items asking whether a person hunts (labeled in Figure 2), views about the morality of sexual relations among teenagers and extramarital sex, several questions tapping gender roles and gender discrimination, items capturing stereotypes about race, questions about civil liberties, and most questions and scales about abortion.

The second group of items, those for which neither wave 1 nor wave 2 are particularly predictive and as a result prefer the constraint of $\phi = .5$, comprises a set of questions about racial stereotypes, including whether blacks and whites are intelligent. This is consistent with previous work finding a high level of measurement error for these items (Hout and Hastings). This group also includes items strongly shaped by the Great Recession, such as whether individuals believe they will lose their jobs.

Items showing evidence for active updating tend to have ϕ estimates greater than .55, and most have β estimates greater than .6. A small group of variables, including confidence in the leadership of the executive branch of the federal government have low β estimates, meaning that prediction at wave 3 is difficult, but have large ϕ estimates, meaning that wave 2 is still a better predictor than wave 1.
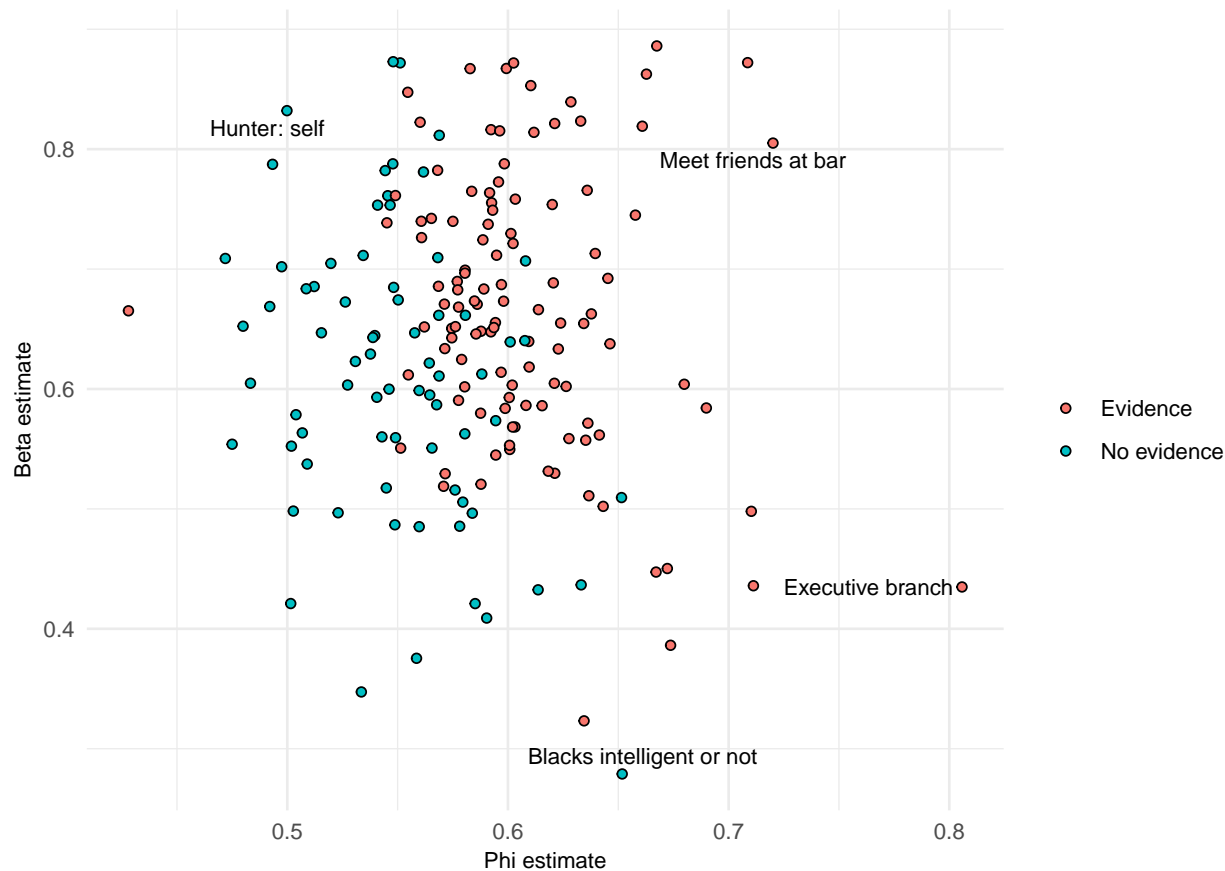
Figure 2: Distribution of beta and phi estimates for GSS items, by whether model prefers phi = .5 constraint.

Age Heterogeneity

While it's impossible to say what proportion of people are following each data-generating process for each item, it is possible within an item to compare whether evidence of active updating is disproportionately concentrated in younger respondents.

Of the 112 items that showed evidence for active updating, 23 of these showed differential effects of age for more than 50 percent of cutoff ages. To get a sense of the magnitude of difference between older and younger individuals, Figure 3 plots the estimates of these 23 items for individuals above or equal to or below 30.

The majority of items that show evidence for age concentration show that active updating is more prevalent among younger respondents. These items include views on affirmative action, women in the workforce, and politics, several civil liberties items, general views of whether people can be trusted, and views on whether doctors should let terminal patients die. These items tend to be in categories where a large proportion showed no evidence of active updating, which suggests an overall trend of these view being formed earlier in life (perhaps prior to inclusion in the GSS) and remaining relatively stable for the rest of life.

For some items, such as whether individuals can be trusted, political views, whether physicians should allow terminal patients to die, and whether companies should make special efforts to hire and promote women to address pass discrimination, all evidence of active updating disappears for individuals over 30. This suggests that early adulthood is a time when these opinions are still malleable, but that over time they harden. For other items, such as how important people believe it is for children to be popular and views on how much the government should spend on health care, there is still evidence of active updating in older individuals even though it is substantially less than for younger individuals.
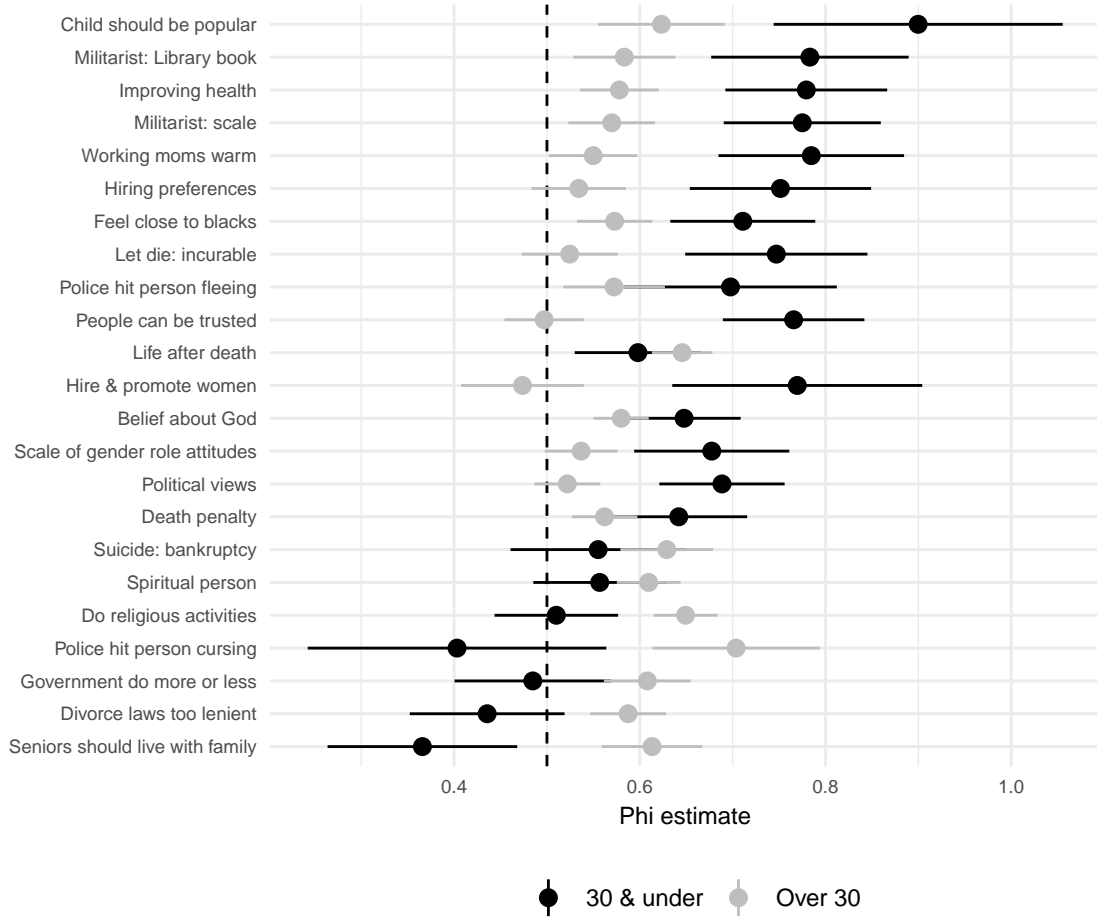
Figure 3: Comparison of phi estimates for individuals over and equal to to less than 30 years old.

Eight of these items show a negative effect of being below the age cutoff on the ϕ value, meaning that younger people showed less evidence of active updating. These items include how often individuals were active in religious activities, views on suicide in the case of bankruptcy, views on whether aging parents should live with their children, and views on whether police should be allowed to use force against suspects who are verbally abusive toward them. Some of these items, such as views on whether aging parents should live with their children might be things young people are not forced to really consider until later in life and as a result do not form clear opinions until that exposure.

The remaining 89 items explored in this analysis (just under half of all items) show evidence for active updating but do not show consistent age heterogeneity. However, this does not mean they show strong evidence of an active updating model. We now turn to exploring the extent of evidence for active updating for items in the data set.

Item heterogeneity

Which items show the strongest evidence for active updating? Figure 4 through 7 plot ϕ estimates for all items included in the analysis, grouped by subject material. We constrain items that showed no evidence of active updating to $phi = .5$ in these plots. For ease of interpretation, we omit from these plots one item that showed a statistically significant negative ϕ estimate: whether respondents can imagine situations in which they would approve of a policeman striking an adult male citizen.
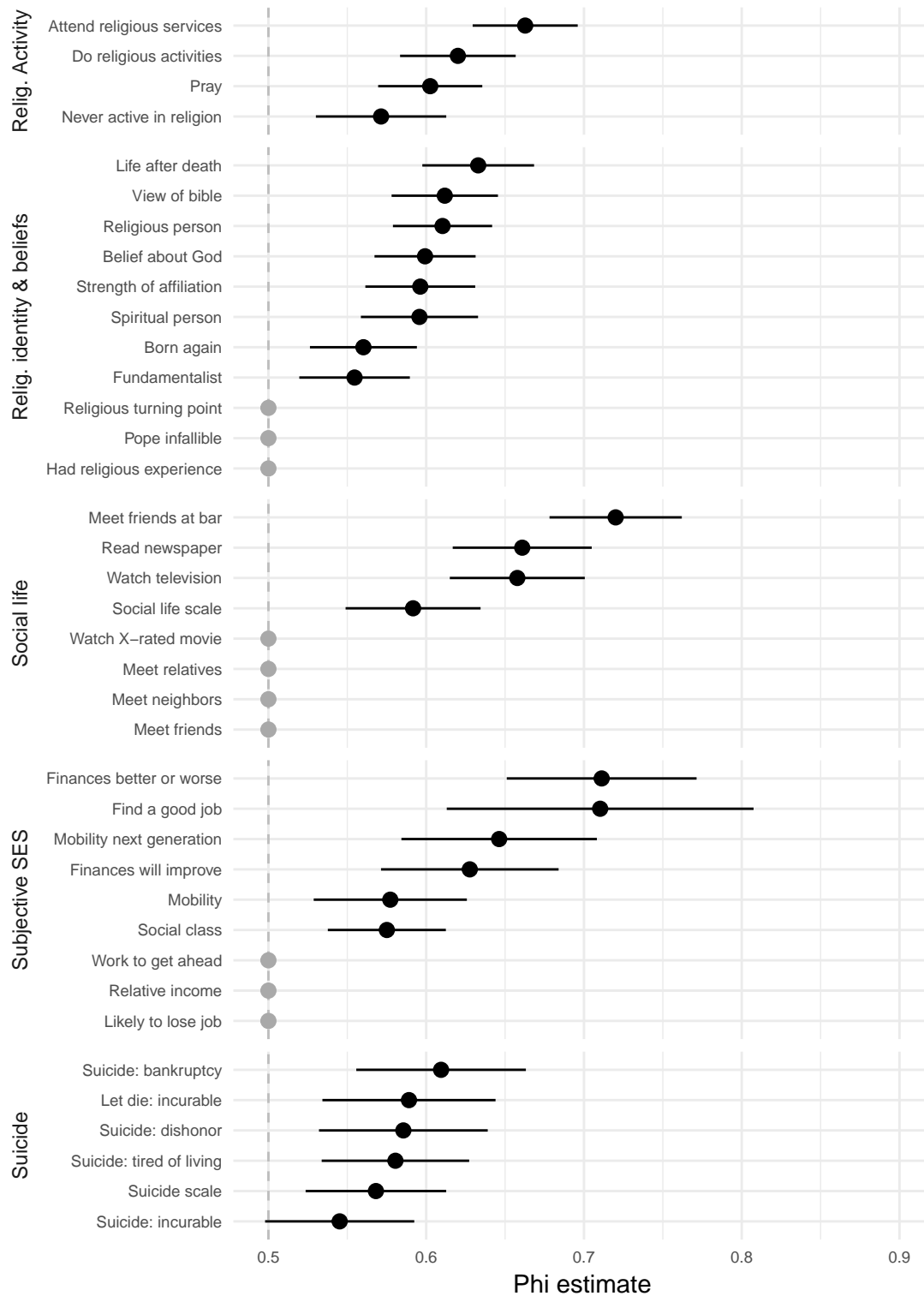
Figure 4: Phi estimates for items about religious activity and beliefs, social life, subjective SES, and suicide.

The overwhelming takeaway from these figures is that even for those items where there is some evidence, the majority of change in these items is non-persisting, either because there is significant measurement error in
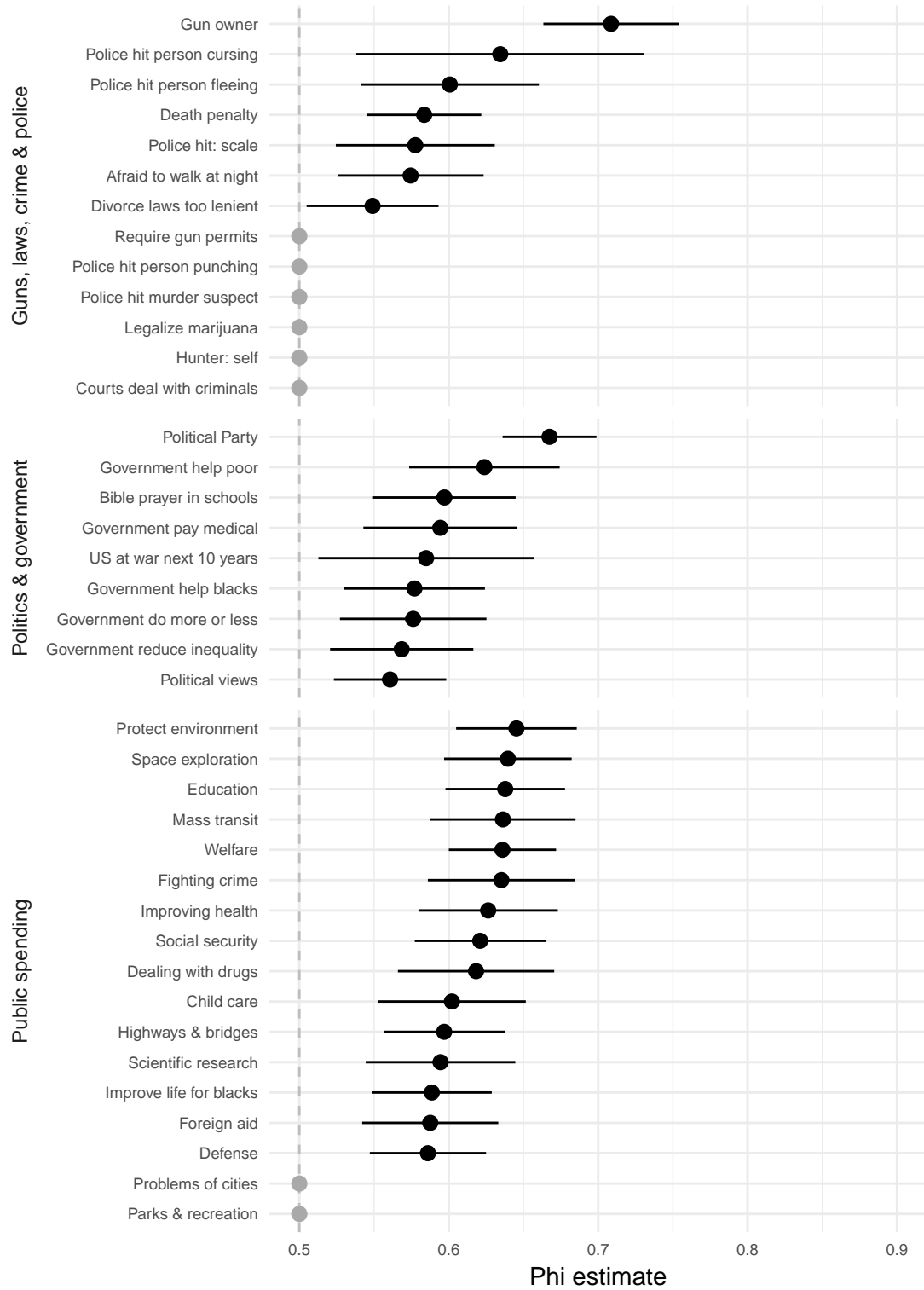
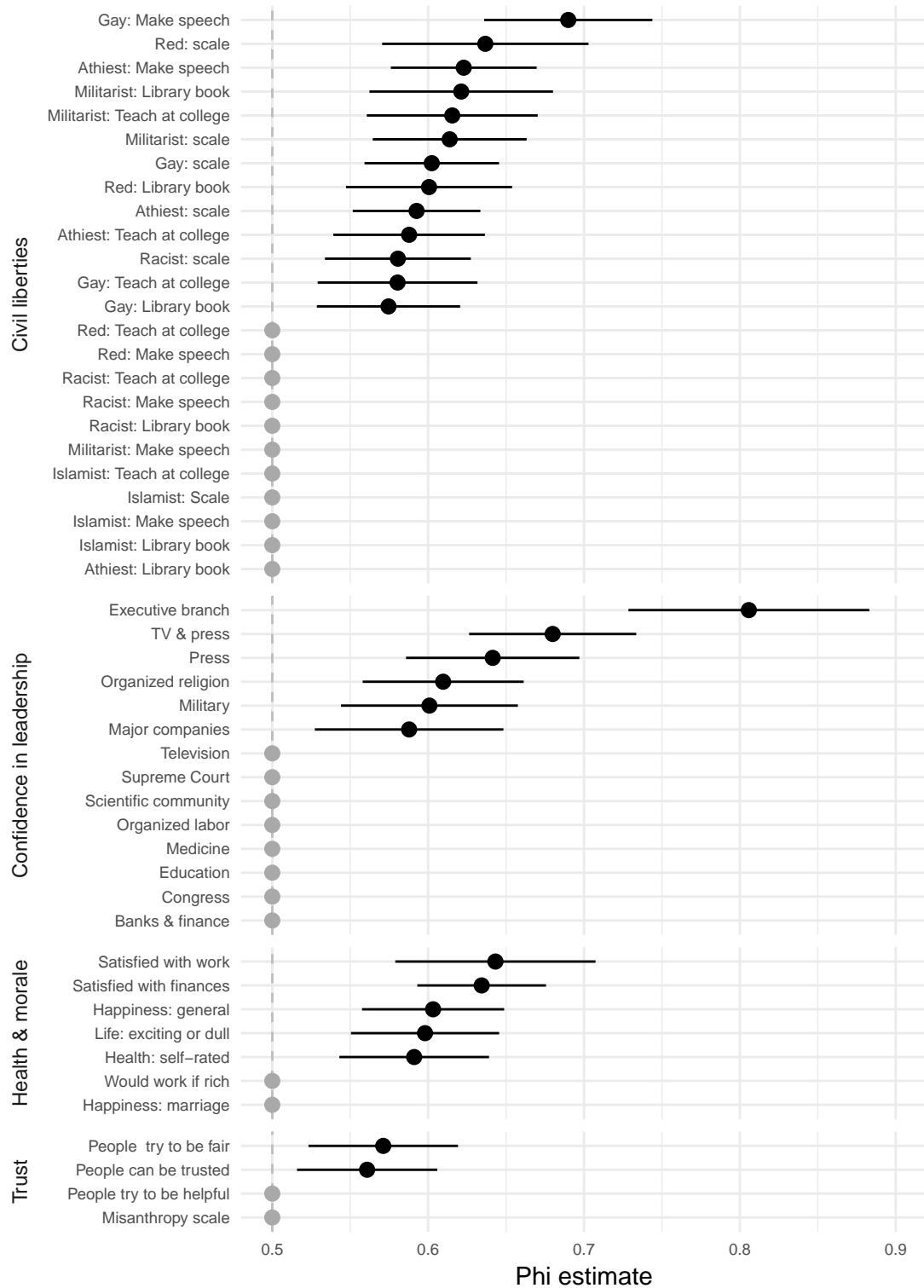Figure 5: Phi estimates for items about guns, law, crime and policing; politics and government; and public spending.

Figure 6: Phi estimates for items about civil liberties, confidence in leadership, health, morale, and social trust.

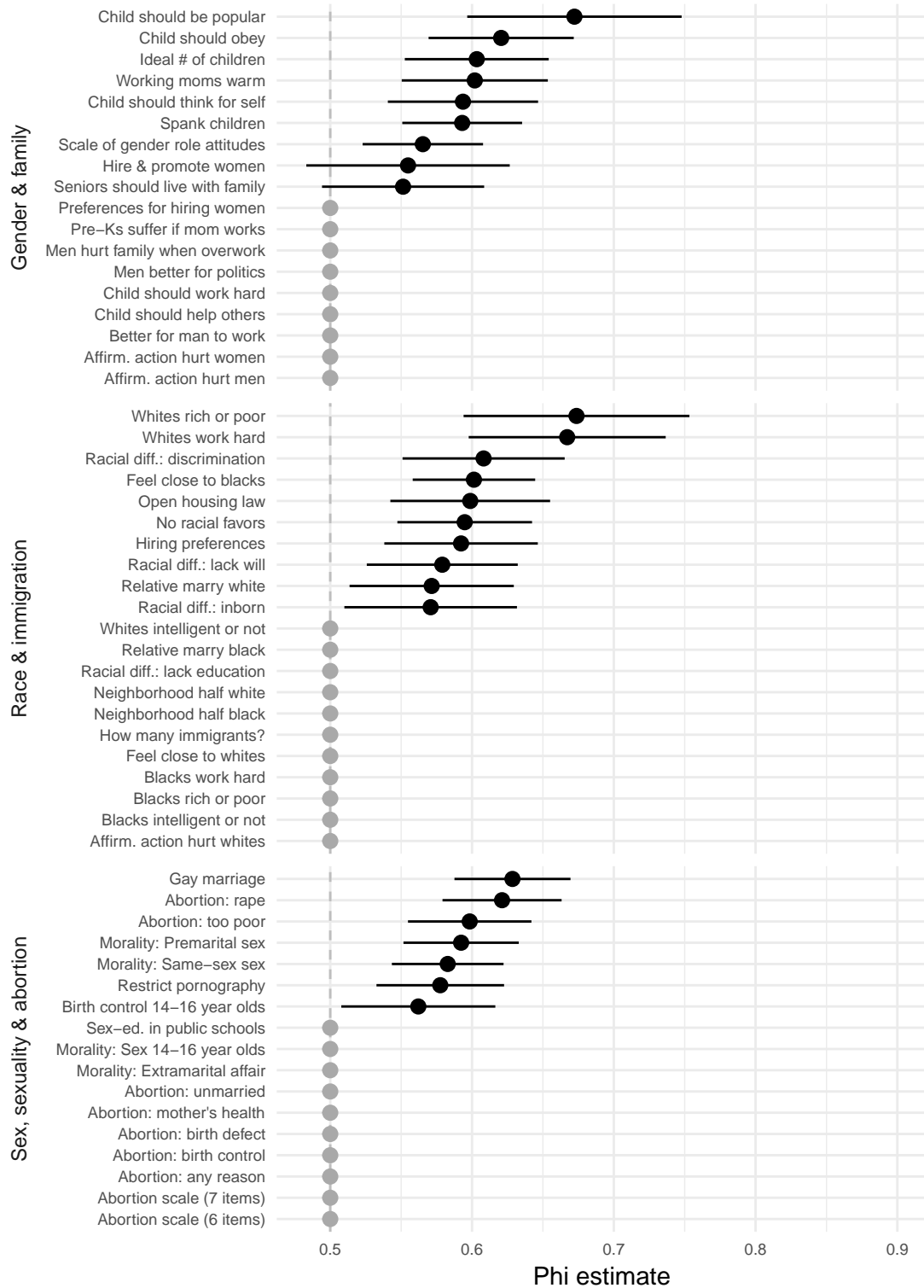individuals' responses or because individuals make non-persisting changes at higher rates than they make

Figure 7: Phi estimates for items about race, gender, sex, sexuality, and abortion.

persisting change. The majority of items that show evidence of active updating have ϕ estimates less than .6, with only five items exceeding .7. Items that show the highest rate of active updating include items

with clearly identified external referents that change (e.g., confidence in the executive branch of the federal government), items that were strongly influenced by the Great Recession of 2008 (a person's assessment of whether their financial situation has changed and ability to get a new job), and facts such as whether someone owns a gun, which are easily recalled and tend to have low levels of measurement error.

Because change can take on a variety of forms, and because the estimate of ϕ varies depending on the direction and relative variance of real change and measurement error, these ϕ estimates do not translate into a clearly understood quantity of change. For reference, however, in simulated models where persisting and non-persisting change are drawn from the same distribution with no measurement error, a ϕ of .6 equates to about twice as many people making non-persisting changes as persisting changes.

It's impossible to talk about all results in these tables, but we highlight some notable results and trends. Starting with Figure 4, religious activity is the only category where all items show evidence of an active updating process, and these items appear to have stronger evidence than other kinds of questions, with $\phi > .6$ for three of the four items. Some items that we might assume would have high evidence of persistence do not. Whether a person had a religious turning point or was "born again" show no evidence of persistence, a surprising finding given that these changes should be non-reversible.

Almost all items dealing with politics, government, and public spending show evidence for an active updating model, with ϕ estimates close to .6. Within items dealing with politics, which political party an individual identifies with shows more evidence than other questions for active updating, while general political ideology and views on the government's role show less evidence.

Within the civil liberties category, persisting change does not appear to be concentrated among either a type of freedom (library, speech, college) or a subject of that freedom (atheist, militarist), as types and subjects see both persisting and non-persisting change. The notable exception here is items about freedom for homosexuals, all of which show evidence of persisting change. Similarly, all but one of the scales show evidence of persisting change, suggesting that removing measurement error increases our ability to detect the true form of change in the data. This lack of a clear pattern for most civil liberties questions raises questions about how and why people update their views on civil liberties.

Within the "confidence in leadership" category, low-visibility institutions with harder-to-identify leadership, such as the scientific community, education, organized labor, and medicine, tend to have no evidence for active updating, while high-visibility institutional leaders such as organized religion, the financial system, and the executive branch of the federal government show evidence of active updating.

As noted previously, gender, race, and sexuality topics in Figure 7 tend to have a higher proportion of settled items compared to other categories. Even for those items that do show evidence of persistence, the ϕ estimates tend to be smaller (less than 0.6), with the exception of items such as support for same-sex marriage and support for abortion in the case of rape. Within items about sex and sexuality, all but two items on abortion appear to be settled.

As noted previously, low measurement error might be a reasonable assumption for some items. Previous studies using different approaches to measuring the reliability of survey reports suggest that some items captured in our study, such as whether a person owns a gun, are measured with a high degree of reliability ($>$ .9) (Hout and Hastings). For other items, such as confidence in the leadership of major companies ($\phi = .58$), reliability might be as low as .5. There is very little correlation ($\rho = .165$) between our ϕ estimates and reliability estimates.

While we cannot control for measurement error in our analysis, we can take steps to mitigate its impact. Appendix A presents results comparing items with more than three scale points to coarsened versions of these question with either two or three scale points. If scales had symmetrical scales with no clear midpoint (e.g., strongly agree/agree/disagree/strongly disagree), we coarsened those scales to two points (agree/disagree). If scales included a clear midpoint, we included that and coarsened responses on either side. For items on large scales such as hours of TV watched, we split responses into greater than the median or less than or equal to the median.

Generally speaking, most items with no evidence of active updating continue to have no evidence of active

updating when coarsened. For items that showed evidence for active updating, the coarsened estimates tended to be very similar to the uncoarsened estimates, suggesting that persisting and non-persisting changes happen about as often between ends of these scales as they do within ends of these scales.

There are a couple notable departures from this general pattern. Several political questions – general political views, views on whether the government should reduce inequality, and views on whether the government should help blacks – show decreased evidence of persisting change when coarsened. This suggests that changes around the midpoint of the scale tend to be measurement error, a finding that's consistent with previous work suggesting that individuals without settled political views tend to choose points around the middle of the response scale (Converse).

Discussion

Appendix A: Coarsened variable estimates

Phi estimates for coarsened items are presented in Figures 8 through 10. If changes between ends of each scale are more persistent than changes within ends of these scales, then ϕ estimates of coarsened item should be higher than estimates for items with more response options.
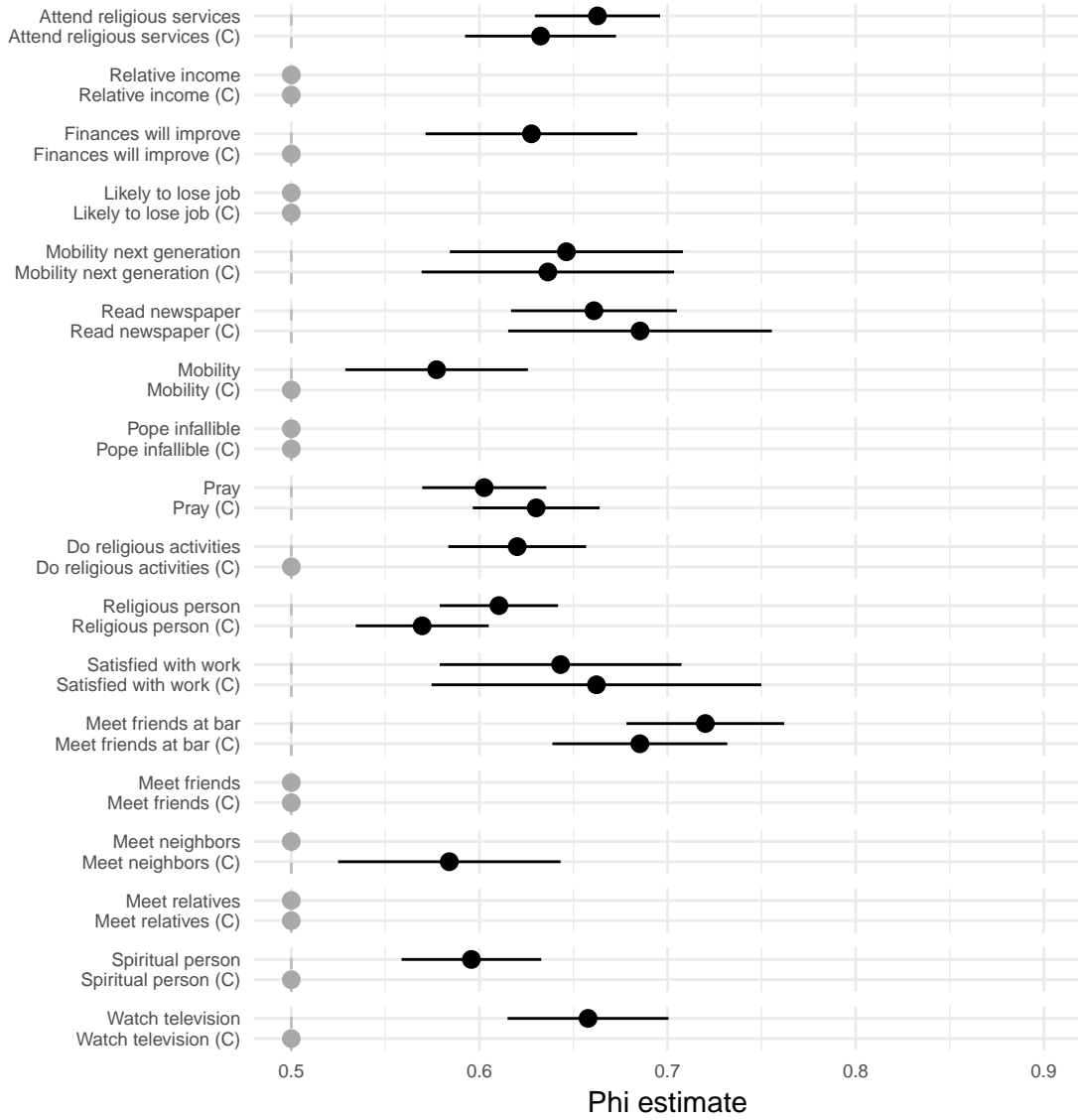
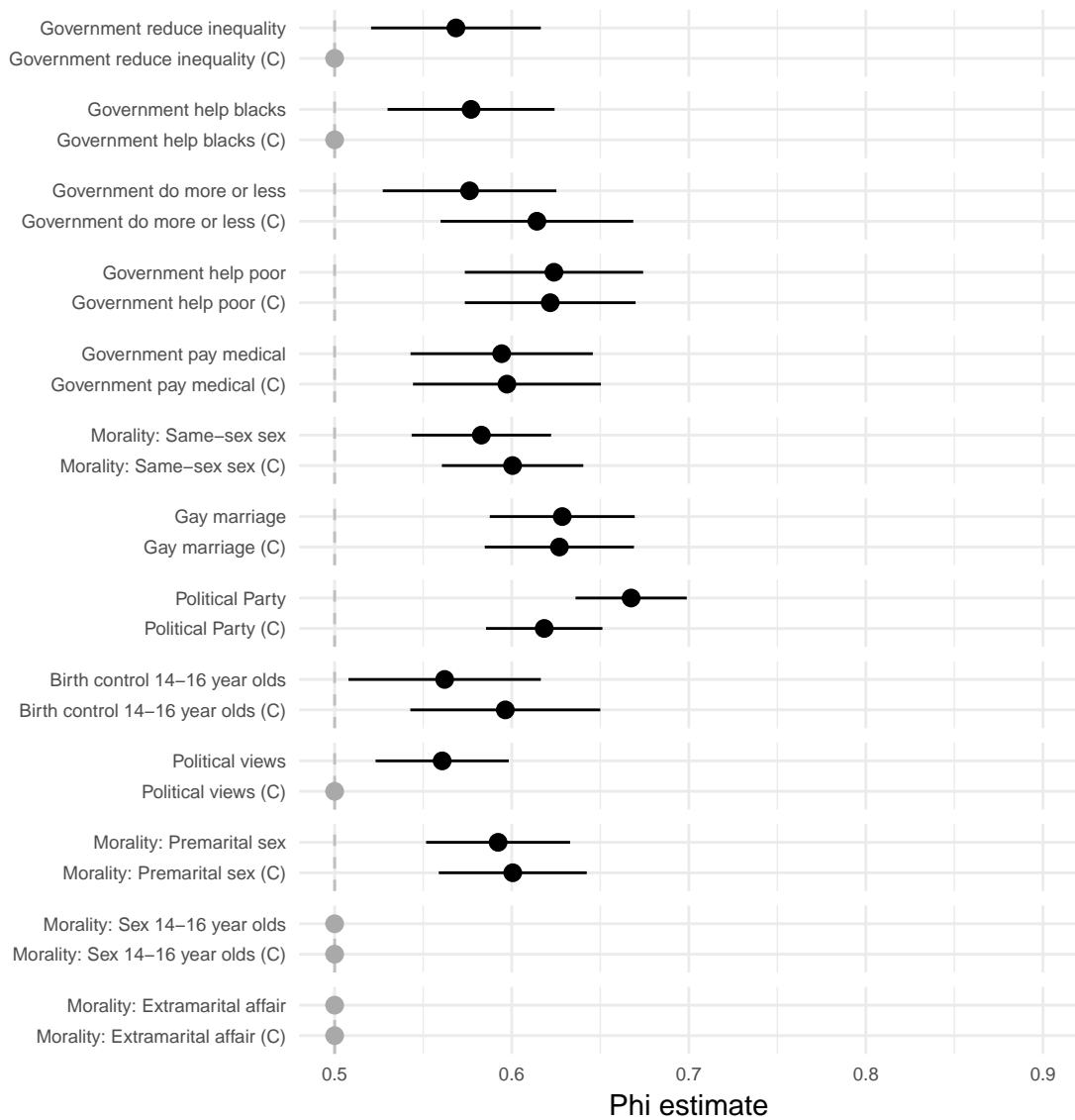Figure 8: Phi estimates for coarsened questions about religion, subjective socioeconomic status, and morale.

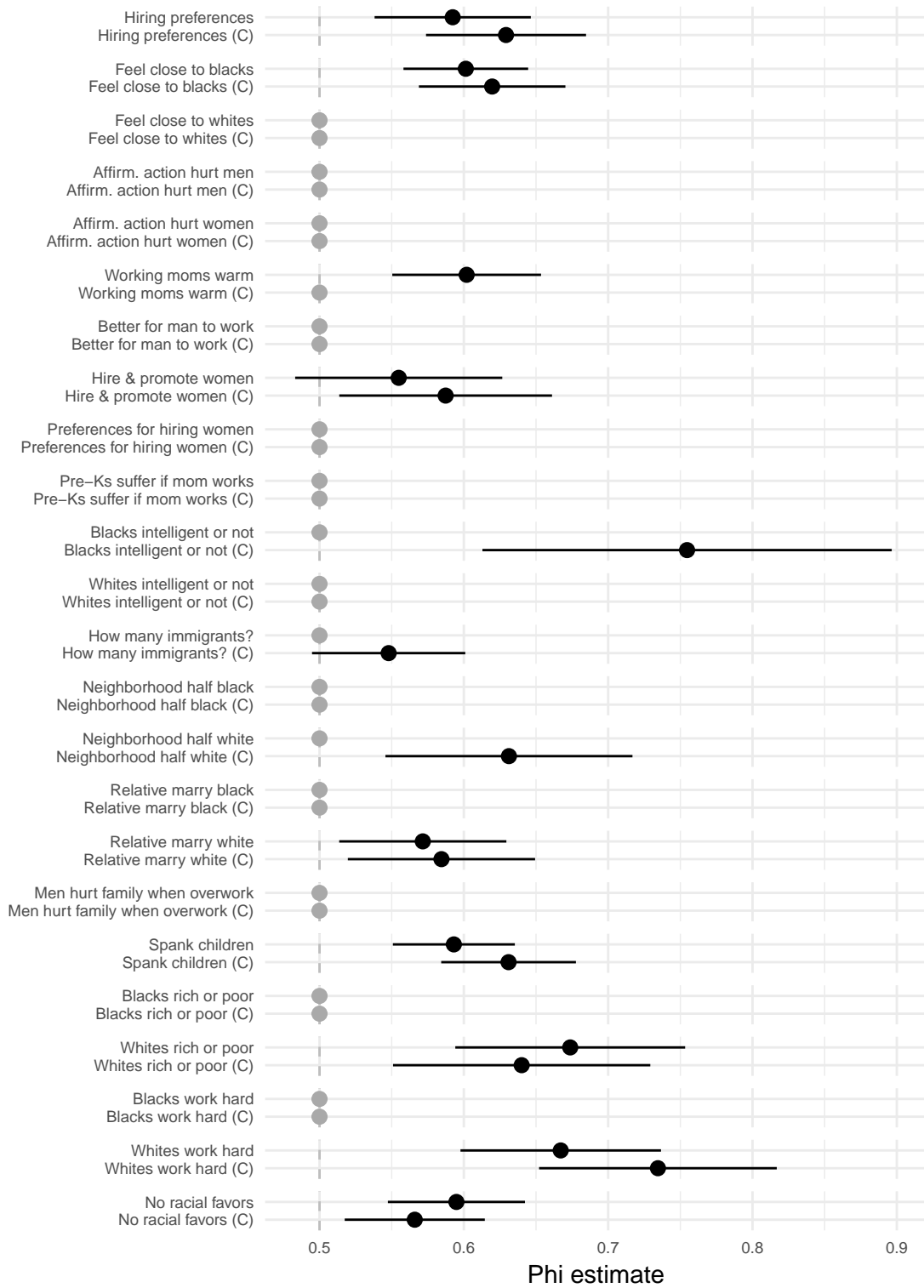Figure 9: Phi estimates for coarsened questions about politics, government, sex and sexuality.

Figure 10: Phi estimates for coarsened questions about race and gender.