

Online Appendix

Issue-Specific Opinion Change: The Supreme Court & Health Care Reform*

Dino P. Christenson
Assistant Professor
Department of Political Science
Boston University
DinoPC@BU.edu

David M. Glick
Assistant Professor
Department of Political Science
Boston University
DMGlick@BU.edu

Contents

1 Online Appendix	1
1.1 Panel Data	1
1.2 Psychometric Properties of Health Care Support	5
1.3 Model Specifications	5
1.4 Predicted Probabilities	9

List of Tables

A1 Sample Demographics and Comparison with Other Surveys	4
A2 Health Care Support	7
A3 Change in Mandate-Specific Health Care Support	8

List of Figures

A1 Sample Demographics by Panel Wave	3
A2 Predicted Probabilities of Support for Mandate	10
A3 Predicted Probabilities of Change in Support for the Mandate	11

*Published in *Public Opinion Quarterly*.

1 Online Appendix

1.1 Panel Data

We collected our data using Amazon.com’s Mechanical Turk (MTurk) service and implemented our surveys online using SurveyGizmo. MTurk is an online marketplace for hiring people to complete simple tasks. We recruited the initial participant pool by posting an open ad or “HIT” on MTurk offering \$1 for an “easy 15 minute survey about politics and health care.” We followed the best practices established in Berinsky, Huber and Lenz (2012) and restricted the posting to Turkers who are over 18, U.S. residents, and had at least a 95% approval rating on their previous MTurk tasks. We excluded participants with IP addresses outside the U.S.

In line with common practice in the literature we also planted screener questions in the survey (Oppenheimer, Meyvis and Davidenko, 2009). First, we asked: “Which position in the federal government do you personally hold?” in a series of factual questions about government officials and institutions. Second, we asked respondents to “Select the ‘somewhat favorable’ button for this row if you are paying attention to this survey” as part of a grid of health care policies they were rating. Finally, we provided a follow up question to a short article that asked if they had read it and recalled the major argument. 99% passed at least one, 89% passed at least two, and 84% passed all three. While some error and the occasional lack of attention are natural in a survey, we wanted to exclude participants who appeared to utterly shirk their duties—randomly clicking through the survey in order to receive a payment. In a recent article Berinsky, Margolis and Sances (2014, p. 747) argue that eliminating people based solely on a single question may be too high of a bar for attentiveness, and thus we kept all those who passed at least one of the three screener questions.

For the second, third, fourth and fifth waves, we sent email invites to all who had successfully completed the previous wave based on code for mass emailing MTurk participants for panel studies (Berinsky, Huber and Lenz, 2012).¹ These emails contained links to a private MTurk HIT that was only available to those we selected to continue. The dates of the data collection (each wave) are indicated in Figure 1 in the manuscript. Our attrition rates were relatively low (see Bartels, 1999). The number of responses per wave was 1242, 944, 856, 751 and 469. In Wave 2 we maintained 76% of the sample, which is lower than the other waves because it was only in the field for 48 hours

¹For more details on the code see http://docs.google.com/View?id=dd4dxgxf_9g9jtdkfc.

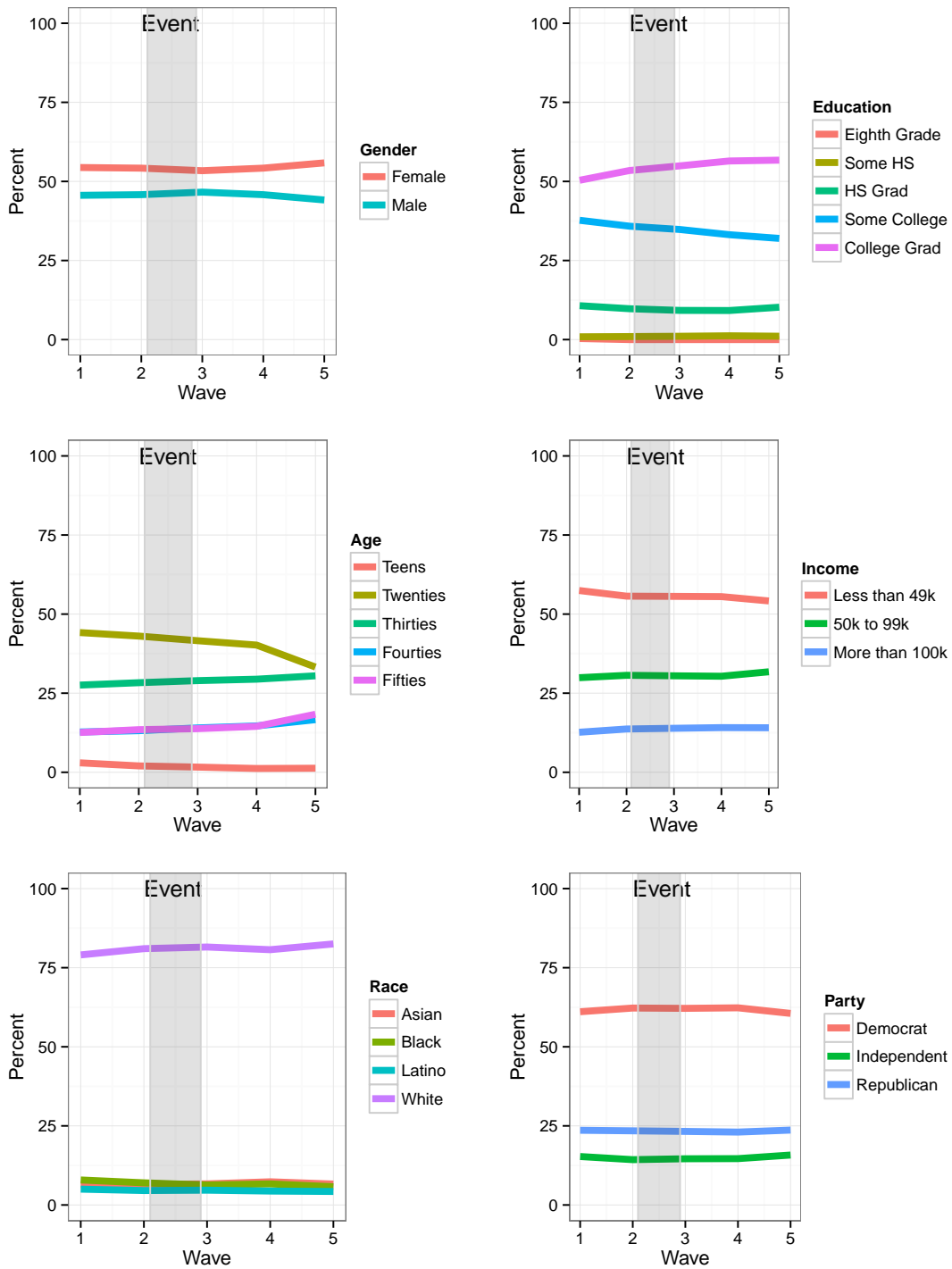
between the Monday when many people expected a decision and the Thursday when the Court released its ruling. We were able to maintain 90% of our respondents from waves 2 to 3 and 88% from waves 3 to 4. Most notably, 469 individuals responded in the fifth wave, which is over 60% of the fourth wave sample that we conducted more than three months earlier.

Figure A1 shows stability in our panel’s demographics over time. While there were some deviations in the fifth wave, collected three months later, our data suggest that panel attrition was essentially more or less equal across the categories of respondent traits, such as race, gender, partisanship, and income. Interestingly, slight trends are evident in age and education, with 18 and 19 year olds and those with some college education consistently falling out slightly more often than the older and more educated, respectively. It is somewhat intuitive to expect that across five months the older and more educated would be more reliable survey respondents.

Naturally, our use of Mechanical Turk to recruit a sample for our panel study raises some questions. While still a new technology, MTurk is becoming an increasingly popular tool for participant recruitment in the social sciences. For example, in the last couple of years, the very top political science journals have published studies that rely on MTurk samples (Huber, Hill and Lenz, 2012; Grimmer, Messing and Westwood, 2012; Arceneaux, 2012; Healy and Lenz, 2014; Dowling and Wichowsky, 2014). All of these studies use the MTurk respondent pool for survey experiments, which makes our application of MTurk to a panel an important extension. It is nevertheless not a radical one since we focus on within subjects change. MTurk provides speed, flexibility, and cost advantages that matter more in panel studies than in survey experiments. Ours is a study that would not be possible with a nationally representative sample and thus joins Druckman, Fein and Leeper (2012) and Gaines et al. (2007) as an example of a useful convenience panel study.

According to Berinsky, Huber and Lenz (2012), MTurk samples offer better demographics than other convenience samples (especially student samples) published in the *American Political Science Review*, *American Journal of Political Science*, and *Journal of Politics* between 2005 and 2010. Our sample essentially matches the expectations established in Berinsky, Huber and Lenz (2012). It is not as representative as the field’s best national probability samples (it is younger and more liberal) but more representative than other convenience samples. Table A1 imitates their demographic summary. It compares our MTurk sample to their MTurk sample, to a high quality internet

Figure A1: Sample Demographics by Panel Wave



Note: Event refers to the period between the second and third waves when the Supreme Court ACA decision was announced.

Table A1: Sample Demographics and Comparison with Other Surveys

Variable	Our Sample	Internet		Face to Face	
		BHL MTurk	ANES-P 2008-09	CPS 2008	ANES 2008
% Female	54.4	60.1	57.6	51.7	55.0
% White	79.0	83.5	83.0	81.2	79.1
% Black	7.9	4.4	8.9	11.8	12.0
% Hispanic	5.0	6.7	5.0	13.7	9.1
Age (years)	33.4	32.3	49.7	46.0	46.6
Party ID (mean 7 pt.)	3.2	3.5	3.9		3.7
Ideology (mean 7 pt.)	3.3	3.4	4.3		4.2
Education	50% Coll Grad 37% Some Coll	14.9 yrs	16.2 yrs	13.2 yrs	13.5 yrs
Income (median)	30-49K	45K	67.5K	55K	55K

Note: Traits for our sample from wave 1 (N=1242), “BHL MTurk” = Berinsky, Huber and Lenz (2012), ANES-P = American National Election Panel Study (Knowledge Networks), CPS = Current Population Survey, ANES = American National Election Study), CPS and ANES are weighted. Data from all columns other than “Our Sample” reproduced from Table 3 in Berinsky, Huber and Lenz (2012).

panel study, to the 2008-2009 American National Election Panel Study (ANES-P) conducted by Knowledge Networks, and to two gold standard traditional surveys, the 2008 Current Population Survey (CPS) and the 2008 ANES. Overall, our sample appears to closely mirror the population with a few expected deviations.

One of the samples that Berinsky, Huber and Lenz (2012) compare MTurk to is the American National Election Study Panel survey (ANES-P). They conclude that the MTurk sample is “slightly less educated” than the 2008 ANES-P and that both “somewhat underrepresent low education respondents” (p. 357-358). On some dimensions the MTurk respondents are closer to the national population than are the participants in the ANES-P though they are also younger and more liberal. Additionally, their analysis shows that MTurk respondents’ voting registration and participation rates are “more similar to the nationally representative samples than is the ANES-P” and, on issues central to this paper: “the MTurk responses match the ANES well on universal health care about 50% of both samples support it whereas those in the ANES-P are somewhat less supportive at 42%.” Their general conclusion suggests that MTurk should not be mistaken for a nationally representative sample, but that it not dramatically skewed or fundamentally flawed as applied to within subject changes either. “All told, these comparisons reinforce the conclusion that the MTurk sample does not perfectly match the demographic and attitudinal characteristics of the U.S. population but

does not present a wildly distorted view of the U.S. population either. Statistically significant differences exist between the MTurk sample and the benchmark surveys, but these differences are substantively small. MTurk samples will often be more diverse than convenience samples and will always be more diverse than student samples.”

As in any panel study there is the possibility that the treatment effects are somewhat different than in the population at large due to panel learning. Pertaining to our hypotheses, however, there is nothing about being in the panel that would make respondents sensitive on the mandate but not on broader reform. Moreover, any potential heightened attention to the decision from being on the panel simply means more people may have been exposed to the treatment making our estimate closer to the average treatment effect on the treated.

1.2 Psychometric Properties of Health Care Support

The scale of general health care support is made of three items that are strongly interrelated, reliable and when combined make up a unidimensional scale. The coefficient of reliability, Cronbach’s α is .82, suggesting that the items have relatively high internal consistency. The item-test correlations range from .84 to .88 with an average inter-item covariance of .48. The factor analysis provides an eigenvalue for the first factor of 1.71, which is much higher than the subsequent two factors, -.12 and -.16. All items load onto the first factor at greater than .7 (.72, .79, .75, respectively). Principal component analysis arrives at the same conclusions with eigenvalues of 2.24, .43 and .34 for the first three components. Subsequent parallel analysis further indicates that only one component should be retained. The results overwhelmingly demonstrate that the variable is an appropriate operationalization of general support for health care reform.

1.3 Model Specifications

We chose a dichotomous variable construction over an ordinal one for the mandate-specific support variable above (see Table 1). We did so primarily because post-model diagnostics suggest that the ordinal measure does not consistently meet the proportional odds assumption. Thus, given the potential for bias in the parameter estimates, we utilized the more simplistic measure. While the dichotomous measure disregards some information in the ordinal measure, we show here that the substantive conclusions drawn from the results of the dichotomous specification in Table 1 are

essentially no different than those gathered from an ordinal construction. Moreover, because the ordinal measure is a simple four-point scale, two points are unsupportive and two points are supportive, which makes the collapse to a dichotomous measure straightforward. Table A2 presents the results from a random effects (individual) ordinal probit model with the same covariate specifications as above. The direction and significance of the results are virtually unchanged suggesting the robustness of the results.

We also chose to use least squares for the 7-point (symmetric) summative index dependent variable of directional change in Table A3. Again, here, post-model diagnostics suggest that the ordinal measure does not consistently meet the proportional odds assumption, and therefore we use the least squares results above. However, we can still show that the substantive conclusions are generally the same when using an ordinal logit specification as when using least squares, which lends some additional confidence to the results. The one difference appears to be that strength of partisanship meets statistical significance in the model below, though in both cases the variable is negative.

Table A2: Health Care Support

	Mandate
Ideology	-0.196* (0.042)
Democrat	0.593* (0.157)
Republican	-0.945* (0.159)
Trust in Government	0.545* (0.080)
Media Attention	0.224* (0.066)
Age	-0.005 (0.004)
Education	0.245* (0.094)
Female	-0.340* (0.106)
Black	-0.080 (0.150)
Latino	0.107 (0.243)
Income	0.199* (0.053)
Wave 2	0.024 (0.058)
Wave 3	0.535* (0.061)
Wave 4	0.695* (0.064)
Wave 5	0.534* (0.075)
τ_1	1.777* (0.452)
τ_2	3.163* (0.453)
τ_3	4.819* (0.455)
Pooled N	4260
Unique N	1241

* $p < 0.05$

Note: Dependent variable is ordinal support for the health care mandate (Mandate). Coefficients, with standard errors in parentheses, are from ordinal probit with random group intercepts for individuals.

Table A3: Change in Mandate-Specific Health Care Support

	Change in Support Directional
Ideological Distance	-0.029 (0.044)
Strength of Partisanship	-0.109* (0.045)
Egocentric Reform Support	0.234* (0.107)
Legitimacy	0.014 (0.018)
Decision Knowledge	-0.038 (0.055)
Trust in Government	0.062 (0.102)
Media Attention	0.099 (0.090)
Age	0.010 (0.006)
Education	-0.015 (0.104)
Female	-0.052 (0.142)
Black	-0.032 (0.299)
Latino	0.460 (0.334)
Income	-0.167* (0.069)
τ_1	-5.408* (0.873)
τ_2	-3.786* (0.705)
τ_3	-1.877* (0.663)
τ_4	1.271 (0.659)
τ_5	3.164* (0.673)
τ_6	5.370* (0.774)
N	856

* $p < 0.05$

Note: Dependent variable is the change in mandate-specific health care support from Wave 2 to Wave 3. Coefficients, with standard errors in parentheses, in the Directional model is from ordinal logit regression.

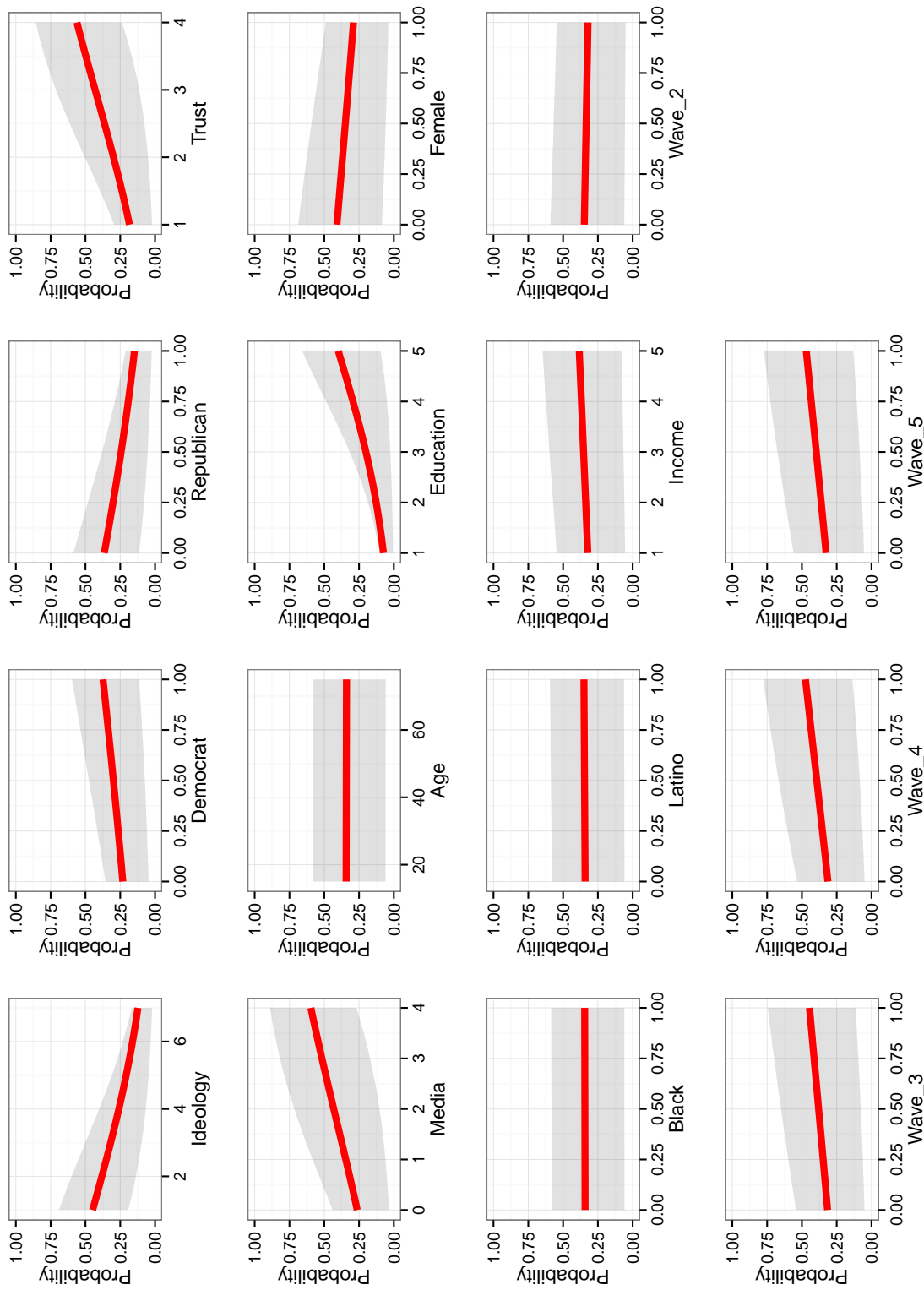
1.4 Predicted Probabilities

Models from tables 1 and 2 are on nonlinear (logit) scales and therefore the marginal effects of these covariates are less obvious. To that end, we present graphs of the changes in predicted probabilities for each of these models below.

Figure A2 plots the average marginal probability of support for the health care mandate (dichotomous) from the generalized (logit) linear mixed effects regression in Table 1. For each graph we hold the predictor of interest at one value within its range, the other covariates at their actual value, and calculate the linear prediction. We do so for all the individuals (random intercepts) across the full range of each of the covariates. That is, in order to get predicted probabilities for the generalized linear mixed effects—where, contrary to a standard logit model, the random terms also affect the results—we calculate conditional probabilities for every individual and average across them. We present the average predicted probability at each level with confidence bands for the first and last quartiles. Thus the predicted probabilities in Figure A2 give us a good estimate of the conditional expectations at each level of the covariate, and do so on the scale of the model: the probabilities of support for the mandate. For example, we see in the Ideology subgraph that the average probability of support for the mandate decreases by about .33 as ideology changes from strongly liberal to strongly conservative.

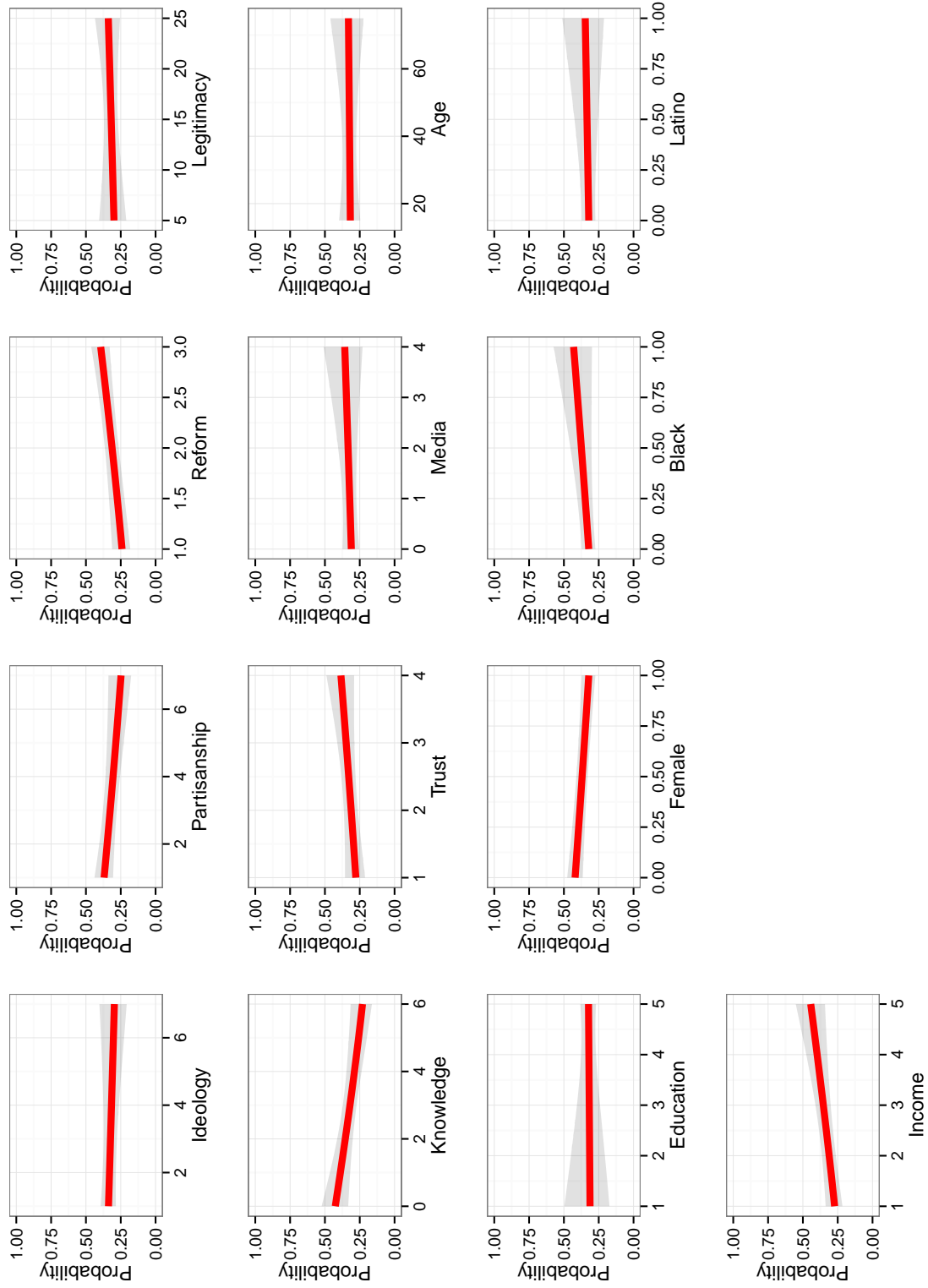
Figure A3 plots the predicted probabilities of change in support for the mandate (dichotomous) from the logit model in Table 2. Here we plot the effect of each value of a covariate, holding the other covariates at their mean (continuous) and mode (dichotomous), on the probability of change in support with shading for the 95% confidence intervals. For example, we see in the Knowledge subgraph that the predicted probability of change in support for the mandate decreases by about .2 as individuals go from uninformed to very informed about the health care decision.

Figure A2: Predicted Probabilities of Support for Mandate



Note: Average predicted probability of support for the mandate with shading for the those between the first and last quartiles. Parameters are from the generalized (logit) linear mixed effects regression in Table 1.

Figure A3: Predicted Probabilities of Change in Support for the Mandate



Note: Predicted probability of change in support for the mandate for each covariate holding the others at their mean (continuous) and mode (dichotomous) with shading for the 95% confidence intervals. Parameters are from the logit model in Table 2.

References

- Arceneaux, Kevin. 2012. "Cognitive Biases and the Strength of Political Arguments." *American Journal of Political Science* 56(2):271–285.
- Bartels, Larry M. 1999. "Panel Effects in the American National Election Studies." *Political Analysis* 8(1):1–20.
- Berinsky, Adam J., Gregory A. Huber and Gabriel S. Lenz. 2012. "Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk." *Political Analysis* 20(3):351–368.
- Berinsky, Adam J., Michele F. Margolis and Michael W. Sances. 2014. "Separating the Shirkers from the Workers? Making Sure Respondents Pay Attention on Self-Administered Surveys." *American Journal of Political Science* 58(3):739–753.
- Dowling, Conor M and Amber Wichowsky. 2014. "Attacks without Consequence? Candidates, Parties, Groups, and the Changing Face of Negative Advertising." *American Journal of Political Science* .
- Druckman, James N., Jordan Fein and Thomas J. Leeper. 2012. "A Source of Bias in Public Opinion Stability." *American Political Science Review* 106(2):430–54.
- Gaines, Brian J, James H Kuklinski, Paul J Quirk, Buddy Peyton and Jay Verkuilen. 2007. "Same Facts, Different Interpretations: Partisan Motivation and Opinion on Iraq." *Journal of Politics* 69(4):957–974.
- Grimmer, Justin, Solomon Messing and Sean Westwood. 2012. "How Words and Money Cultivate a Personal Vote: The Effect of Legislator Credit Claiming on Constituent Credit Allocation." *American Political Science Review* 106(4):703–719.
- Healy, Andrew and Gabriel S Lenz. 2014. "Substituting the End for the Whole: Why Voters Respond Primarily to the Election Year Economy." *American Journal of Political Science* 58(1):31–47.
- Huber, Gregory A., Seth J. Hill and Gabriel S. Lenz. 2012. "Sources of Bias in Retrospective Decision-Making: Experimental Evidence on Voters Limitations in Controlling Incumbents." *American Political Science Review* 106(4):720–741.
- Oppenheimer, Daniel M, Tom Meyvis and Nicolas Davidenko. 2009. "Instructional Manipulation Checks: Detecting Satisficing to Increase Statistical Power." *Journal of Experimental Social Psychology* 45(4):867–872.