Online Appendix: Chief Justice Roberts' Health Care Decision Disrobed: The Microfoundations of the Supreme Court's Legitimacy

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

List of Tables

List of Figures

1 Online Appendix

1.1 Panel Data Details

We utilized Amazon.com's Mechanical Turk (MTurk) service to recruit participants to an online panel survey which we implemented using Survey Gizmo. MTurk is an online marketplace for hiring people to complete simple tasks. It is becoming an increasingly popular tool for recruiting participants for social scientific surveys and experiments (Berinsky, Huber and Lenz, 2012; Buhrmester, Kwang and Gosling, 2011) and results based on MTurk samples have been published in our field's top journals (Huber, Hill and Lenz, 2012; Grimmer, Messing and Westwood, 2012, in the *APSR*) (Arceneaux, 2012; Healy and Lenz, 2014; Dowling and Wichowsky, 2014, in the *AJPS*). Though these studies all use MTurk for experiments, our extension to a panel study focused on within subjects change is a natural and important extension. Indeed, we believe the cost savings, speed, and flexibility that MTurk offers have much larger comparative advantages for panel studies and that MTurk can support some panel studies that may otherwise be impossible (see e.g. Druckman, Fein and Leeper, 2012; Gaines, Kuklinski and Quirk, 2007). While Berinsky, Huber and Lenz (2012) directly test MTurk by replicating experiments, they also compare its properties to high quality panel populations and highlight its potential for panel studies that focus on within subject changes. In our case, we are interested in both the within subject changes in response the Court's decision as well as an experimental treatment.

We recruited the initial participant pool by posting an open ad or "HIT" on Mechanical Turk offering \$1 for an "easy 15 minute survey about politics and health care." Following Berinsky, Huber and Lenz (2012) we restricted the posting to MTurk users, over 18, with at least a 95% approval rating on their previous tasks, and to U.S. residents. Consistent with with common practices we included a few screener questions in the survey (Oppenheimer, Meyvis and Davidenko, 2009). First, we asked respondents "Which position in the federal government do you personally hold?" in a series of factual questions about government officials and institutions. Second, we asked them 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 inattention is natural and realistic in a survey, we wanted to address a concern that is especially salient for online surveys where participants may be inclined to shirk through the survey by randomly clicking boxes in order to receive a payment. In their recently published article (a working paper when we ran our survey) Berinsky, Margolis and Sances (2014, p. 747) argue that eliminating people based on a single screener may set too high of a bar for attentiveness. Balancing this new advice and common practice at the time, we kept all those who passed at least one of the three screener questions in order to eliminate the most egregious shirkers from the panel. We also excluded participants with IP addresses outside the U.S. For the second, third, and fourth waves we sent email invites to all who had successfully completed the previous wave. These emails contained links to a private HIT that was only available to those we selected to continue.

The numbers of completed responses in the four waves were 1,242, 944, 856, and 751. The first wave ran from June 19th to June 21st and thus finished a week before the decisions were released. The second wave ran from Tuesday June 26th (the day after many had speculated the decisions would be released) until the morning of the 28th (just hours before the release). The third wave ran from the morning of Friday the 29th (one day after the release) through the holiday weekend and ended July 5th. The final wave ran from July 20th through July 30th (three weeks after the decisions).

MTurk samples have attractive properties relative to student and adult convenience samples published in the *American Political Science Review*, *American Journal of Political Science*, and *Journal of Politics* between 2005 and 2010 and are comparable to high quality internet panel samples (Berinsky, Huber and Lenz, 2012; Buhrmester, Kwang and Gosling, 2011). 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 outperforms the typical convenience samples. Table A1 follows their work in comparing our MTurk sample to their MTurk sample, to a high quality internet 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.[1]

[1]We can further validate our sample by essentially replicating the results in Bartels and Johnston (2013) in Table A3 using static data.

Most relevant perhaps is the comparison between MTurk demographics and the ANES panel study. Berinsky, Huber and Lenz (2012, pp. 357-358) write that the MTurk sample is "slightly less educated" than the 2008 American National Election Study Panel survey (ANESP) and that both "somewhat underrepresent low education respondents." In other words, the MTurk respondents are less over-educated relative to the ANES than the participants in the ANESP. They also find that MTurk respondents' voting registration and participation rates are "more similar to the nationally representative samples than is the ANESP." Furthermore, they compare samples on policy views. They find that "the MTurk responses match the ANES well on universal health careabout 50% of both samples support itwhereas those in the ANESP are somewhat less supportive at 42%." Finally, they summarize MTurk relative to nationally representative samples as follows: "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." In sum, Berinsky, Huber and Lenz show that the MTurk sample deviates from the gold standard surveys in important ways and should never be considered a valid substitute for making population estimates. On the other hand, they also show that many deviations are relatively minor and that on some important metrics the MTurk sample is more similar to the ANES than the ANES panel.

Given our research design, data about the panel's evolution is as important as information about its initial population. Our analysis of panel attrition does not suggest any significant reason for concern. The panel appears to exhibit virtually random attrition. Figure A1 shows the evolution of our panel's demographics. Particularly in the first four waves, collected in a one-month period around the Supreme Court decision (highlighted as the decision in the figure), our panel demographics remained stable as people fell out of the panel. Thus 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. Despite some very slight trends in age and education, with 18 and 19 year olds and those with some college education falling out more often than the older and

Figure A1: Panel traits by wave

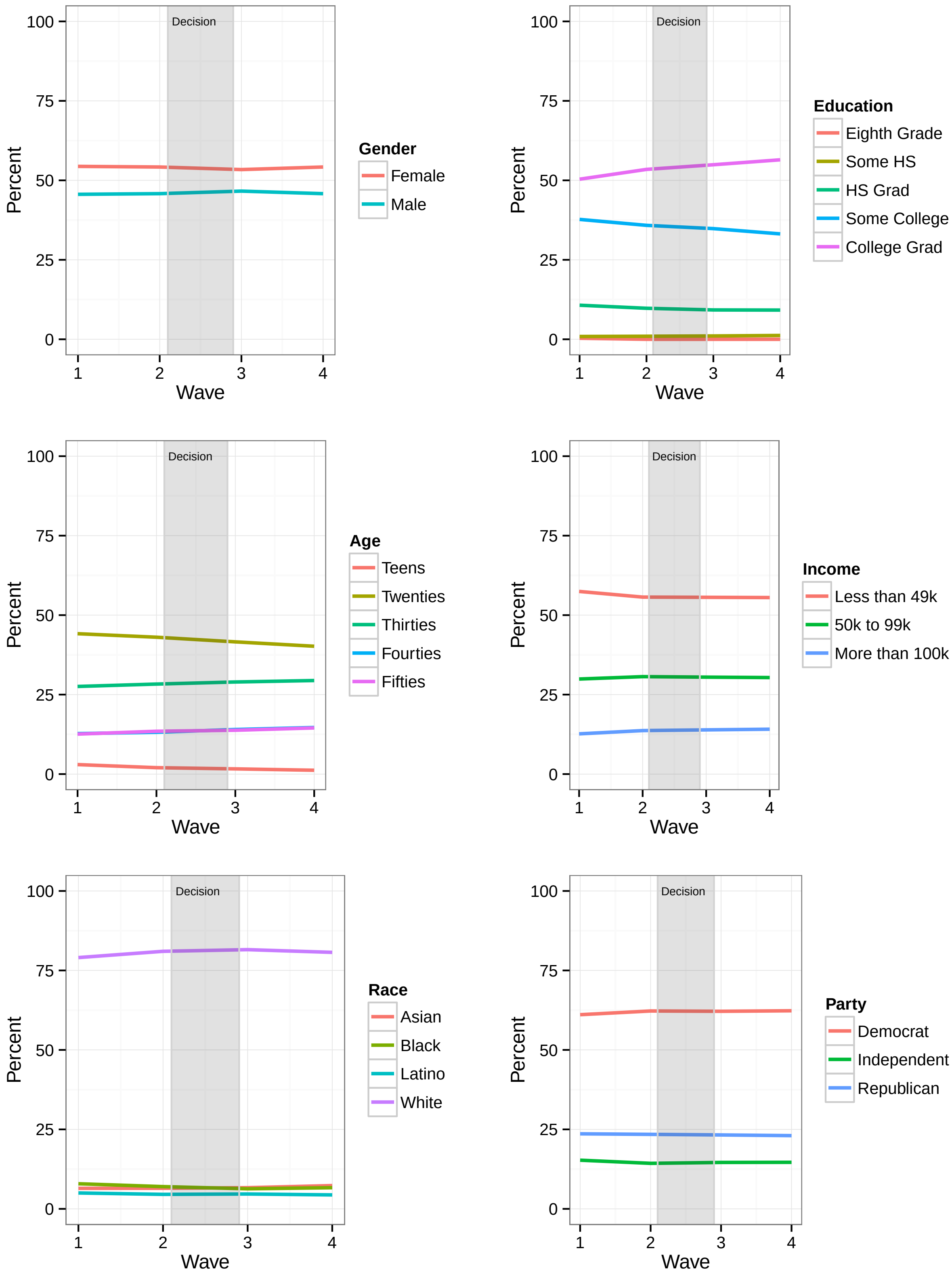


Table A1: Sample Demographics and Comparison With Other Surveys

		Internet Samples		Face to Face Samples	
Variable	**Our Sample**	**BHL Turk**	**ANES-P 08-09**	**CPS 08**	**ANES 08**
% 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% Col Grad 37% Some Col	14.9 yrs	16.2 yrs	13.2 yrs	13.5 yrs
Income (median)	30-49K	45K	67.5K	55K	55K

Traits for our sample from wave 1 (N=1242), BHL Turk = 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).

more educated, respectively—it is somewhat intuitive to expect that the older and more educated would be more reliable survey respondents—the data suggest that there is little danger in the panel attrition in MTurk samples across the typical demographics.

1.2 The Political Courts Treatment

A portion of the sample was assigned to read the article shown in Figure A2. The as-if random assignment is detailed in the body of the paper with further information in the next section of the Appendix. The article was created entirely from the language and formatting of the actual CBS news article (Crawford, 2012). We kept the details most related to Roberts' non-legalistic or political rationale and trimmed some of the context and extra details to reduce the article length to a single page.

1.3 As-If Random Assignment

Table A2 shows that even though we could not randomize exposure to treatment, our two groups are balanced in important ways such that this convenient treatment is unlikely to bias the results. As the comparisons in the table show, the two groups are comparable in terms of demographics. Only the percent of white respondents even approches conventional significance levels and a p-value of .16

Figure A2: The Treatment Condition

July 1, 2012 1:29 PM

PRINT TEXT

Roberts switched views to uphold health care law

comments 492 Like 4.4k Tweet 963 +1 59 Share 23 More +

By Jan Crawford

The justices of the U.S. Supreme Court sit for their official photograph Oct. 8, 2010, at the Supreme Court in Washington. (AFP/Getty Images)

(CBS News) Chief Justice John Roberts initially sided with the Supreme Court's four conservative justices to strike down the heart of President Obama's health care reform law, the Affordable Care Act, but later changed his position and formed an alliance with liberals to uphold the bulk of the law, according to two sources with specific knowledge of the deliberations

Roberts then withstood a month-long, desperate campaign to bring him back to his original position, the sources said. Ironically, Justice Anthony Kennedy - believed by many conservatives to be the justice most likely to defect and vote for the law - led the effort to try to bring Roberts back to the fold.

The conservatives refused to join any aspect of his opinion, including sections with which they agreed, such as his analysis imposing limits on Congress' power under the Commerce Clause, the sources said.

Instead, the four joined forces and crafted a highly unusual, unsigned joint dissent. They deliberately ignored Roberts' decision, the sources said, as if they were no longer even willing to engage with him in debate.

After the historic oral arguments in March, the two knowledgeable sources said, Roberts and the four conservatives were poised to strike down at least the individual mandate. There were other issues being argued - severability and the Medicaid extension - but the mandate was the ballgame

Over the next six weeks, as Roberts began to craft the decision striking down the mandate, the external pressure began to grow. Roberts almost certainly was aware of it. As Chief Justice, he is keenly aware of his leadership role on the Court, and he also is sensitive to how the Court is perceived by the public. There were countless news articles in May warning of damage to the Court - and to Roberts' reputation - if the Court were to strike down the mandate. Leading politicians, including the President himself, had expressed confidence the mandate would be upheld.

Some informed observers outside the Court flatly reject the idea that Roberts buckled to liberal pressure, or was stared down by the President. They instead believe that Roberts realized the historical consequences of a ruling striking down the landmark health care law. There was no doctrinal background for the Court to fall back on - nothing in prior Supreme Court cases - to say the individual mandate crossed a constitutional line.

Roberts then engaged in his own lobbying effort - trying to persuade at least Justice Kennedy to join his decision so the Court would appear more united in the case.

Latest from CBS News

- Colo. evacuees visit devastated neighborhoods
- Mid-Atlantic power outages could last days
- Roberts switched views to uphold health care law
- New elderly care program controls health costs
- Report: Romney's family vacation is mandatory

more headlines from CBSNEWS.COM

Table A2: Treatment vs. Control Comparison

Variable	**Control**	**Treatment**	**P**
Female	53%	56%	.539
White	81%	86%	.164
Black	6%	6%	.896
Age	34.603	32.379	.136
Education (Mean of 1-5)	4.431	4.466	1.000
Ideology (Mean of 1-7)	3.230	3.172	.886
Party ID (Mean of 1-7)	3.166	3.276	1.000
Legal Analysis Was Primary Factor	59%	54%	.322
Legal Analysis To Be Primary Factor	52%	53%	.755
ACA Ruling Factual Knowledge (Mean of 1-6)	2.984	2.793	.449
N	740	116	

P-Values from t-tests for binary variables and Kolmogorov-Smirnov tests for nonbinary variables.

and even then the difference is only a few percentage points. Variables are not perfectly balanced in actual randomizations anyway and we are able to control for race in our models. More importantly, the primary ways we would expect race to matter connect to partisanship and ideology and those variables are highly balanced. The two year age gap is the largest difference between the treatment and control groups, though even it is not statistically significant and unlikely to affect how one reacts to information about the Supreme Court.

More importantly, those in the treatment and control groups answered a question similarly about which factor they thought would be dominant in the Justices' minds: "Which of these do you think will play the most important role in the Supreme Court Justices' decisions on the health care reform case?" 52% percent of respondents in the control and 53% in the treatment chose "the Justices' analysis and interpretation of the law" while the rest chose other options such as "whether a Justice was appointed by a Republican or Democratic president" or "national politics." This balance is critical; it demonstrates that prior to the treatment the two groups had indistinguishable expectations about the study's issues and that neither group was initially more cynical on average.

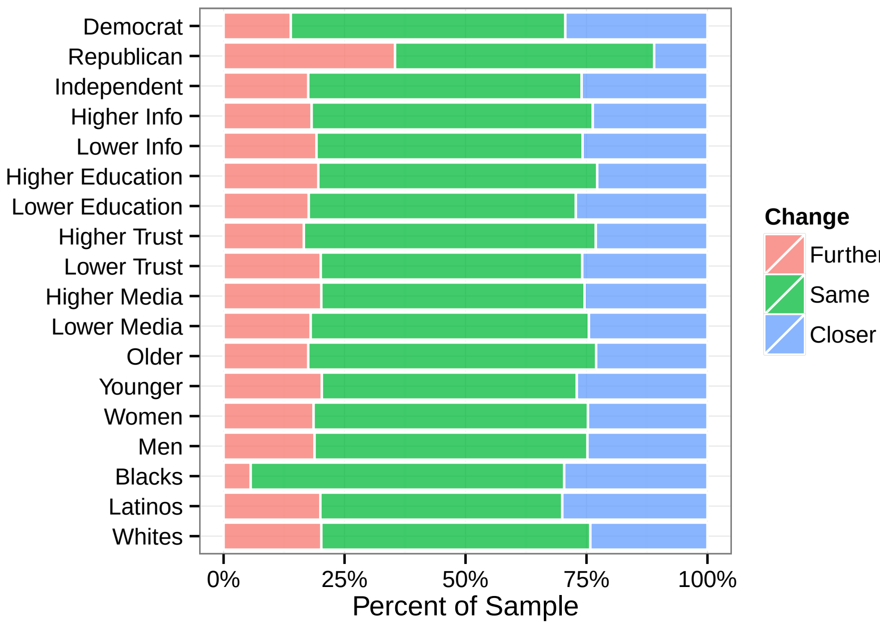
As we discuss in the text, we can also address the concern that selection was quasi-random but answering the questions a few days later made the groups unequal in important ways. Having a couple of extra days of exposure to stories about the decisions, to partisan spin, and other factors could make our late responders (the treated) different. We are able to demonstrate that these concerns are minor ones at most. One might expect that those extra days meant more time to incorporate information. We compare the two groups using a six question index of challenging factual questions about the ACA decision (e.g., its author, the dissent's position, the Medicaid expansion). If those in the treated group were affected by being immersed in the post-decision environment, or if they were less enthusiastic about the issues, they do not show it. As Table A2 shows, they knew as much about what the Court did as those in the control. We can also test whether the treated were more cynical about the Court's motives or the politics of the decision than those in the control. We asked a question about the primary factor in the Justices' decision that parallels the one described previously. Here, in Wave 3 we replaced "will play" with "played" and asked the question before the treatment. Again, there is no difference between the treated and control; they were equally confident in the Court's legality. As best we can tell, going well beyond basic demographic balance, our opportunistic assignment mechanism worked very well.

1.4 Individual Level Change & Aggregate Stability

Figure A3 displays how many people increased, decreased, and did not change their ideological congruence assessments after the decisions by a variety of attributes which may affect legitimacy (Bartels and Johnston, 2013). Not surprisingly, Democrats were more likely to increase their congruence assessments and Republicans were more likely to decrease them. Black respondents also increased congruence, though there is certainly a partisan overlap here. Additionally, people with more political trust, more information, and older people were more stable and thus less likely to reassess the Court in both directions. All of these results are consistent with the Court's generally high level of aggregate stability. The people most likely to get new information about the Court (high information people) are also less likely to update their views about its ideology.

Moreover, as long as a case has a partisan dimension, many individual level changes will appear to cancel each other out in the aggregate, a central finding of ours that we discuss in the body of the paper. As we see above, Democrats and Republicans update their beliefs about their ideological proximity to the Court in opposite directions and in roughly equal numbers in reaction to the decision. Thus the cross-sectional measures of legitimacy and ideological proximity appear stable even though individuals' perceptions have changed. That is, the analysis in the paper and elaborated on here is the first to explain why the previous literature has captured both evidence of stability (largely in observational studies at the aggregate level) and change (largely in experimental studies at the individual level). On salient and politicized cases individuals from both sides of the debate move in opposite directions in roughly equal number.

Figure A3: Bar Plots of Change in Ideological Proximity to Court



1.5 Static Models

In Table A3 we present the results of models similarly specified to those in the text, but with a static legitimacy score for the dependent variable. When we look at a static dependent variable of legitimacy from the 3rd wave cross-section, the results are more similar to what Bartels and Johnston (2013) have found with regards to Court awareness and political trust as well as ideological congruence. We note, however, that there are a couple of differences. In particular, our measures of media awareness, though similarly specified to that in the extant literature, as well as gender and Latino do not meet with traditional levels of statistical significance. Despite less racial and informational heterogeneity in our sample, these results suggest the robustness of our key findings

Table A3: OLS Models of Supreme Court Legitimacy (Static)

	I	II
Intercept	8.671^{***}	8.652^{***}
	(0.697)	(0.698)
Roberts Flip	-0.416^{*}	
	(0.241)	
Δ Ideological Proximity	-0.551^{***}	-0.550^{***}
	(0.053)	(0.053)
Republican	-0.504^{**}	-0.512^{**}
	(0.246)	(0.247)
Independent	-0.381^{**}	-0.377^{**}
	(0.186)	(0.186)
Supreme Court Awareness	0.282^{***}	0.279^{***}
	(0.050)	(0.050)
Political Trust	0.974^{***}	0.962^{***}
	(0.117)	(0.117)
Media Differential	0.015	0.015
	(0.023)	(0.023)
Age	−0.004	−0.002
	(0.007)	(0.007)
Education	0.351^{***}	0.347^{***}
	(0.123)	(0.123)
Female	0.043	0.035
	(0.170)	(0.171)
Black	-0.650^{*}	-0.640^{*}
	(0.342)	(0.342)
Latino	0.032	0.073
	(0.393)	(0.392)
R^2	0.216	0.213
Adj. R^2	0.205	0.203
Num. obs.	856	856

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

above, and the appropriateness of the MTurk sample for this study.

1.6 Parsimonious Specifications

We noted above in the text that several of the static variables above did not influence the change in legitimacy, even though they are correlated with a static measure of legitimacy, the latter of which is shown in the appendix section immediately above this one. We present here more parsimonious specifications of the change in Court legitimacy. That is, in Table A4 we show the results from three regression models of legitimacy with combinations of our key independent variables: ideology, the politicization story and the interaction between the two. The results are very similar in size, strength and significance to the fully specified models in the text.

Table A4: OLS Models of Change in Supreme Court Legitimacy (Parsimonious)

	I	II	III
Intercept	-0.037	0.000	-0.119^{*}
	(0.071)	(0.075)	(0.067)
Roberts Flip	-0.558^{***}	-0.586^{***}	
	(0.194)	(0.203)	
Δ Ideological Proximity	0.396^{***}		0.472^{***}
	(0.058)		(0.055)
Roberts Flip * Δ Proximity	0.644^{***}		
	(0.172)		
R^2	0.102	0.010	0.079
Adj. R^2	0.099	0.009	0.078
Num. obs.	856	856	856

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

1.7 Court & Case Awareness

In line with some previous research, we test here an additional interactive hypothesis. Based on Gibson and Caldeira (2011) we expected political awareness to moderate the effect of political exposure on changes in legitimacy. The coefficient on the interaction term in Model A5 is insignificant and so we are unable to reject the null in this case. We are unable to find support here for the potentially moderating effect of political information. Before jumping to conclusions on this point, however, we note that a number of factors peculiar to this study may have contributed to this null finding. First, our MTurk sample was particularly well-informed and educated, much more so than the population. Additionally, the respondents may have learned about the Court as a mere function of being in a panel study of this nature. In both cases the homogeneity of the sample on these variables may have dampened any potential effects from political awareness. Finally, the focus of this study on a singular case may underestimate the role of media attention and political awareness at times when a case before the Court is not so salient.

While the models throughout the paper control for a *Court Awareness* variable that follows Bartels and Johnston (2013), we find similar null results using a series of questions which test nuanced knowledge of the Court's health care opinions in Table A6. Even this measure, comprised

Table A5: OLS Model of Change in Supreme Court Legitimacy (Court Awareness Interaction)

	I
Intercept	-0.469
	(0.580)
Roberts Flip	-0.950
	(0.979)
Δ Ideological Proximity	0.420^{***}
	(0.056)
Republican	-1.038^{***}
	(0.202)
Independent	-0.208
	(0.147)
Supreme Court Awareness	0.030
	(0.043)
Political Trust	0.032
	(0.094)
Media Differential	0.002
	(0.018)
Age	-0.003
	(0.006)
Education	0.100
	(0.099)
Female	-0.059
	(0.138)
Black	-0.038
	(0.276)
Latino	0.066
	(0.316)
Roberts Flip * Court Awareness	0.045
	(0.107)
R^2	0.119
Adj. R^2	0.106
Num. obs.	856

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

of questions focusing on the actual health care case (an index of the number of correct answers to factual questions about what the Court ruled), does little to affect the direct results. As Table A6 shows, this variable has no effect in the change model and our key results are robust to it.

1.8 Measuring Legitimacy

Following previous work in the literature (e.g. Gibson, Caldeira and Spence, 2003; Gibson and Caldeira, 2009) we explore the psychometric properties of the legitimacy index to validate our measures. We find that the responses to the five items in our legitimacy index are positively

Table A6: OLS Models of Change in Supreme Court Legitimacy (Case Awareness)

	I
Intercept	-0.578
	(0.560)
Roberts Flip	-0.567^{**}
	(0.193)
Δ Ideological Proximity	0.353^{***}
	(0.058)
Republican	-1.021^{***}
	(0.200)
Independent	-0.208
	(0.146)
Supreme Court Awareness	0.039
	(0.042)
Case Awareness	-0.002
	(0.055)
Political Trust	0.043
	(0.094)
Media Differential	0.006
	(0.018)
Age	-0.003
	(0.006)
Education	0.101
	(0.099)
Female	-0.031
	(0.137)
Black	-0.114
	(0.275)
Latino	0.069
	(0.314)
Roberts Flip * Δ Proximity	0.627^{***}
	(0.171)
R^2	0.133
Adj. R^2	0.118
Num. obs.	856

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

correlated, with an average inter-item covariance of 0.23 for each Wave, which is extremely close to the "decent level" of inter-correlation of .26 for survey data noted by Gibson, Caldeira and Spence (2003). Reliability is high, with a Cronbach's α of 0.79 for each Wave. The response with the weakest relationship to the total item set is the first one, a rather strong statement about whether the Court should be "done away with" if it starts making decisions respondents disagree with. However, the factor loadings for the orthogonal solution still show that all the variables are moderately to strongly correlated with the first factor. The first factor loadings are .35, .70, .64, .84 and .83, respectively for Wave 2, and .39, .68, .58, .86, .85, respectively for Wave 3. Though the reliability is stronger across the measures in our data, the many similarities in the factor loadings between our data and that of Bartels and Johnston (2013) provide further support for the representativeness of our sample since it appears to mirror the nationally represenative Annenberg sample and suggest that the measures are extremely reliable. In all, our measures appear to have properties that are closely in line with previous research and also indicative of the underlying construct of institutional support.

Undoubtedly, some measures have stronger relations to each other and to the underlying concept than others. For example, as Gibson and Nelson (2013) note, the "trust" items are particularly well correlated with one another, which is also true in our data at $r = .80$ and $r = .82$ in Waves 2 and 3, respectively. They argue that these two items may be picking up "specific support" and thereby contaminating the measure of "diffuse support" in the Bartels and Johnston work. In response to this possibility, we test whether our results are contingent upon the exclusion of these items. Specifically, we create a new three item index which excludes the items concerning whether the Court "can usually be trusted to make decisions that are right for the country as a whole," and "in the best interests of the American people." This new three item index only includes the items concerning whether "we should do away with the Court," whether the Court "favors some groups," and whether the Court is "too political." We estimate the model which comprises our main findings (the effects of ideological updating, the effect of the Roberts article and their interaction) using this new index. Since this new index does not suffer the potential contamination concerns of specific support that Gibson and Nelson believe is conveyed in the trust items, should our results remain consistent across the dependent variable specifications we can be extremely confident in the hypothesized effects of the independent variables on the underlying concept of diffuse support at the heart of both indices, regardless of exactly how precisely it is measured.

The analysis in Table A7 shows that our results are robust to this more conservative measure. The Table presents five models. The first two (I and II) use the 5 item index and the second two (III and IV) use the three item index. For each index we present both the simple "additive" index and a "standardized" (mean centered and rescaled 0-1) version of the index. Thus the first model in the Table shows the results in the orginal submission of the manuscript. Looking at the third model, the three item additive index, we find good reason to rest easy. All three hypothesized effects are

Table A7: OLS Models of Change in Supreme Court Legitimacy (Varying Index and Scale)

	5 Item Index		3 Item Index		Principal Component
	Additive	Standardized	Additive	Standardized	PCA
	I	II	III	IV	V
Intercept	−0.577	−0.038	−0.066	−0.007	−0.421
	(0.559)	(0.037)	(0.396)	(0.044)	(0.344)
Roberts Flip	−0.566**	−0.038**	−0.495***	−0.055***	−0.337**
	(0.193)	(0.013)	(0.136)	(0.015)	(0.119)
Δ Ideological Proximity	0.353***	0.024***	0.178***	0.020***	0.228***
	(0.058)	(0.004)	(0.041)	(0.005)	(0.036)
Republican	−1.021***	−0.068***	−0.484***	−0.054***	−0.663***
	(0.200)	(0.013)	(0.142)	(0.016)	(0.123)
Independent	−0.209	−0.014	−0.077	−0.009	−0.124
	(0.146)	(0.010)	(0.103)	(0.011)	(0.090)
Supreme Court Awareness	0.039	0.003	0.020	0.002	0.028
	(0.040)	(0.003)	(0.028)	(0.003)	(0.024)
Political Trust	0.043	0.003	0.008	0.001	0.036
	(0.093)	(0.006)	(0.066)	(0.007)	(0.057)
Media Differential	0.006	0.000	0.004	0.000	0.005
	(0.018)	(0.001)	(0.013)	(0.001)	(0.011)
Age	−0.003	0.000	−0.002	0.000	−0.001
	(0.006)	(0.000)	(0.004)	(0.000)	(0.004)
Education	0.100	0.007	0.025	0.003	0.071
	(0.098)	(0.007)	(0.069)	(0.008)	(0.060)
Female	−0.030	−0.002	−0.052	−0.006	−0.017
	(0.137)	(0.009)	(0.097)	(0.011)	(0.084)
Black	−0.114	−0.008	−0.059	−0.007	−0.077
	(0.275)	(0.018)	(0.194)	(0.022)	(0.169)
Latino	0.069	0.005	−0.004	0.000	0.047
	(0.313)	(0.021)	(0.222)	(0.025)	(0.193)
Roberts Flip * Δ Proximity	0.627***	0.042***	0.332**	0.037**	0.399***
	(0.171)	(0.011)	(0.121)	(0.013)	(0.105)
R^2	0.133	0.133	0.081	0.081	0.143
Adj. R^2	0.119	0.119	0.067	0.067	0.130
Num. obs.	856	856	856	856	856

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

still substantial, in the posited direction and highly significant. This important robustness check shows that our results are not driven by the potentially contaminated "trust" items.

Looking across the models we see substantial differences in the coefficients. This is to be expected as the coefficients necessarily change in line with the different scales of the legitimacy indices. There are, of course, a number of reasons to standardize variables and/or aggregate them in particular manners. Foremost for studies of legitimacy, standardization helps to ensure that variables contribute evenly to a scale when items are used in an index. However, standardizing does not affect our model results, as we show in the standardized columns in Table A7. All results are consistent in direction, significance and relative magnitude across the simple additive and standardized indices. Indeed, this is typically the case when the individual measures that make up the index have similar means, standard deviations, and the same outcome categories, as is the case in our data: $[2.30, 3.06]$, $[.65, .80]$, likert scale agreement, respectively.

We go a step further by using the predicted scores from principal component analysis of the five items as the dependent variable. While we should not expect these scores to give us drastically different results than the simple summative index given the high internal consistency of the legitimacy items and the similar structural properties across them, the potential for a latent dimension from the indicators still merits investigation. The primary concern here is that a simple additive index assumes that each item has an equal effect on the outcome. This assumption is not always tenable, in which case we can make use of principal components analysis to obtain component scores for each respondent. We use these scores or weights as the estimate of each respondent's underlying institutional support. Model five in Table A7 presents the results from this dependent variable construction. Again, we find the expected differences in the coefficients due to the different scale of the dependent variable, but, crucially, the sign, significance and relative magnitude of the effects are the same as the simple summative index. The principal component analysis therefore bolsters the findings from our simpler summative index models. In sum, we find that different constructions of the dependent variable—including variations in the included items, standardization, and principal component analysis—all lead to the virtually identical results and the same substantive conclusions pertaining to our hypotheses.

References

Arceneaux, Kevin. 2012. "Cognitive Biases and the Strength of Political Arguments." *American Journal of Political Science* 56(2):271–285.

Bartels, Brandon L. and Christopher D. Johnston. 2013. "On the Ideological Foundations of Supreme Court Legitimacy in the American Public." *American Journal of Political Science* 57(1):184–199.

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.

Buhrmester, Michael, Tracy Kwang and Samuel D. Gosling. 2011. "Amazon's Mechanical Turk A New Source of Inexpensive, Yet High-Quality, Data?" *Perspectives on Psychological Science* 6(1):3–5.

Crawford, Jan. 2012. "Roberts Switched Views to Uphold Health Care Law." *CBS News: July 1, 2012* .

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 and Paul J. Quirk. 2007. "The logic of the Survey Experiment Reexamined." *Political Analysis* 15(1):1–20.

Gibson, James L. and Gregory A. Caldeira. 2009. "Knowing the Supreme Court? A Reconsideration of Public Ignorance of the High Court." *Journal of Politics* 71:421–441.

Gibson, James L. and Gregory A. Caldeira. 2011. "Has Legal Realism Damaged the Legitimacy of the U.S. Supreme Court." *Law and Society Review* 45(1):105–219.

Gibson, James L., Gregory A. Caldeira and Lester Kenyatta Spence. 2003. "The Supreme Court and the U.S. Presidential Election of 2000: Wounds, Self-Inflicted or Otherwise?" *British Journal of Political Science* 33:535–556.

Gibson, James L and Michael J. Nelson. 2013. "Is the U.S. Supreme Court's Legitimacy Grounded in Performance Satisfaction and Ideology." *Presented at the 2013 American Political Science Association Conference (Chicago, IL)* .

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 ElectionYear 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.