Cynicism, Conspiracies, and Contemporaneous Conditions Moderating Experimental Treatment Effects

Katherine Levine Einstein *
Assistant Professor
Department of Political Science
Boston University
kleinst@bu.edu

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

January 29, 2015

Abstract

Survey experiments are widely used in political science. Among other things, they promise internally valid estimates of a wide variety of causal effects. One specific threat to this assumption is that participants might bring other information based on real-world conditions to these experiments that could mediate posited experimental effects. We investigate this possibility by comparing the results of an experiment concerning conspiracy theories and trust in government conducted at a fairly routine time to results from the same experiment collected in a high-scandal period. We find that the results from the high-scandal period deviate from the original results in ways theoretically consistent with the real world affecting the experiment. Substantively, our findings show that cynicism and conspiracy theories are linked in a vicious cycle in which macro conditions affect conspiratorial thinking. More generally, these findings, especially when combined with previous research, suggest an important threat to the interpretation of at least some survey experiments' results.

6,141 words

*Authors names are listed alphabetically. Working paper prepared for the 2013 APSA meetings. We'd like to thank Brendan Nyhan and Doug Kriner for insightful comments.

In recent years, survey experiments have become increasingly popular in political science research. Their main promise is internally valid identification of direct causal effects. While survey experiments have enabled many important contributions to a variety of literatures, some prominent recent scholarship raises important questions about how cleanly these experiments can parse effects and how other stimuli (including experiments) can contaminate such designs and confound estimation and interpretation (Gaines, Kuklinski and Quirk, 2007; Druckman and Leeper, 2012; Transue, Lee and Aldrich, 2009, see also(Barabas and Jerit, 2010)). In one common survey experimental technique, researchers expose participants to an information treatment, often in a question prologue or in a simulated news article, and then assess that exposure's effect via comparison to a control group. The assumption underlying this design is that differences in the dependent variable identify the main effect of the experimental information. One specific threat to this assumption is the possibility that conditions and/or prevailing news in the real world—including scandals, elections, and wars— mediate experimental effects such that at least some findings are actually interaction effects between the treatment and the environment. In this paper, we directly investigate this issue in the substantive context of conspiracy theories. We simultaneously show that environmental conditions moderate survey experiments and that macro factors affect conspiracy attitudes. As we explain below, our investigation of these substantive questions is inseparable from, and takes advantage of, our methodological queries and contributions.

While observational research—particularly scholarship focused on media frames—frequently takes potential priming from external events into account (Mutz, 1994; Kinder, 1998; Hopkins, 2010), these issues' implications for experiments remain understudied and seemingly under-appreciated. Indeed, while the idea that public opinion is fickle is widely accepted in survey research, standard practice in reporting experimental results does not include discussing potential interactions with contemporaneous conditions. A quick look at the survey experiments papers (about 15 in total) published in the APSR, AJPS, and JOP in 2014 suggests that doing so is exceeding rare (see Figure A1 in the Appendix). None of the

excellent articles we looked at discussed what was going on in the world when the data were collected and/or whether the experimental results could be anything but stable direct treatment effects. Many discuss the implications that participant pools and/or experimental designs have for generalizability, but not whether or not external conditions could make the reported findings time bound interactions.

This is not to impugn any of these studies. They all may have been conducted at times in which the real would was unlikely to moderate the treatments, and/or at least some were replicated in multiple time periods. The point is simply that standard practice includes interpreting treatment effects as independent main effects that will manifest across space and time. Leading researchers, reviewers and editors appear unconcerned about these issues moderating or confounding findings. Our results below show that rethinking this standard practice may be wise.

A small amount of existing work suggests that survey experimentalists should take these concerns seriously and that they demand further examination. Neely's (2007) article, attempts to replicate the question wording effects reported in Burden and Klofstad (2005) and gets different results. He attributes the difference to the fact that the original question wording manipulation was conducted shortly after 9-11 in a high anxiety and pro-Republican environment. Similarly, but using spatial variation, Kriner and Shen (2014) find that being from a state that has disproportionately lost residents to war mediates responses to an experimental casualties treatment. Finally, less directly related to naturally occurring environmental conditions, Druckman, Fein and Leeper (2012) note the paucity of work on "pre-treatment" effects and use a panel-experiment design to show that an injection of information at time T1 conditions the effect of an experiment at time T2 in what they deem the first "conclusive evidence of a pre-treatment dynamic." (see also Transue, Lee and Aldrich (2009) on contamination from other experiments).

We offer a direct and explicit test of the proposition that contemporaneous conditions only indirectly related to experiments mediate effects. We do so by investigating important

and understudied questions regarding macro factors and conspiracy politics. We intentionally exploit exogenous events to field the same survey experiment in two different environments that are theoretically pertinent to the treatment of interest. Specifically, we compare the results of an experiment about conspiracy theories and trust in government (Einstein and Glick, 2014) conducted at a fairly routine time to results from the same experiment when it was conducted in a high-scandal period only months later. In brief, we find that people are more likely to believe conspiracy claims when exposed to them at a time when other government scandals are salient. These findings, especially when combined with those in Neely (2007) and Kriner and Shen (2014), have big implications for conducting, reporting, and interpreting experimental research.

Along with their potential impact in the broad field of experimental methods, our findings also have important substantive implications. We show that cynicism and conspiracy theories are linked in a vicious cycle and in doing so contribute new insights to the burgeoning conspiracy politics literature (e.g. Oliver and Wood, 2014; Uscinski, Klofstad and Atkinson, 2014; Banas and Miller, 2013; Berinsky, 2013; Lewandowsky et al., 2013; Mulligan and Habel, 2013; Einstein and Glick, 2014). Most of the existing scholarship explores the connection between individual-level attributes and conspiracy beliefs. It links traits like partisanship, knowledge, and general cynicism to conspiracy support (Uscinski, Klofstad and Atkinson, 2014; Berinsky, 2013). This very fruitful focus on micro variables and individual psychology, however, overlooks the possibility that broader macro factors might also drive conspiracy beliefs. In a rare example of macro approaches, Uscinski and Parent (2014) discover that power asymmetries and perceptions of threat—both domestic and foreign—might help to explain conspiracy belief. We focus on a different environmental factor: political scandals. The expectation that a cynical and scandal-ridden climate affects conspiracy attitudes follows from a novel theoretical idea which we term *the cycle of cynicism*. Previous research has uncovered a strong link between experimental exposure to conspiracy theories and diminished trust in government (Einstein and Glick, 2014). In this article, we suggest

that the causal arrow could point in the opposite direction as well. Exogenous events—like political scandals—could spur lower levels of trust. In turn, declining trust could lead to more conspiracy beliefs.

1 Scandals and Conspiracies

Our study revolves around exposure to a conspiracy theory in a high-scandal climate. Before proceeding, we briefly define both concepts. A high-scandal environment is fairly intuitive: it is a context in which political figures act (or are purported to act) contrary to expected behavior norms *and* the media covers said transgression(s) extensively. Take the example of the 2013 controversy surrounding the Internal Revenue Service (IRS), the scandalous context used in this article. We consider this event to be a scandal because it featured IRS bureaucrats using tactics that many Americans found unacceptable. More importantly, the IRS controversy was covered extensively and treated as a scandal by the media starting in May 2013 (Nyhan, 2013). The fact that later news suggests the original scandal coverage may have been overblown is less important for the research design than the way the incident was framed right after the news broke.

Our definition of a political conspiracy theory follows previous scholarship. In contrast with factual misinformation—which comprises essentially any fact that is incorrect—conspiracy theories are flawed and unfalsifiable causal explanations that contradict a widely endorsed mainstream account of events (Brotherton, French and Pickering, 2013; Keeley, 1999; Coady, 2006; Aaronovitch and Langton, 2010). While conspiracy theories can encompass myriad domains, we are primarily interested in *government* conspiracies, in which conspiracists claim that the government is using an "official story" to deliberately mislead the public.

1.1 Scandals and Conspiracy Beliefs

We argue that, in addition to the important individual-level attributes identified in previous research, *macro-level variables* like a high-scandal climate could have a profound effect on belief in conspiracies. Scandals which engender a high cynicism environment might have a reinforcing impact on belief in conspiracy theories. Jon Stewart perfectly articulates this point on *the Daily Show* on the day after the IRS scandal broke in May 2013:

> Well, congratulations, President Barack Obama, conspiracy theorists who generally can survive in anaerobic environments have just had an algae bloom dropped on their...heads, thus removing the last arrow in your pro-governance quiver: skepticism about your opponents....This has, in one seismic moment, shifted the burden of proof from the tin-foil behatted to the government.

The intuition is straightforward. Scandals diminish overall trust in government and make other similar claims, including conspiracy theories, seem more plausible. Thus, we put forward **Hypothesis 1**: *All else equal, we should observe greater belief in a conspiracy theory during a scandal-ridden climate than during a low-scandal period, even if the conspiracy claim is unrelated to the existing scandal(s).*

It follows, then, that the way that individuals process conspiracy theories may be different depending upon the prevalence of political scandals. In previous research, Einstein and Glick (2014) find evidence for a *stop and think* mechanism. Individuals who were experimentally prompted to give a specific conspiracy allegation additional thought were less likely to support the allegation. A high cynicism climate should diminish this mechanism's ability to temper the effects of conspiracy claims. In a scandal-laden environment, individuals exposed to a conspiracy theory might think about it and instead find it plausible—as in *Hypothesis 1*. This leads us to **Hypothesis 2**: *We should observe less rejection of conspiracy claims via cognitive processing during periods of high political scandal.*[1]

[1]As Einstein and Glick (2014) note, it is very difficult to cleanly disentangle cognitive processing from social desirability effects. While we believe that the weight of the evidence is on the side of the stop and

Finally, if *Hypothesis 1* proves accurate, and declining trust does indeed spur more conspiracy belief during high-scandal periods, it also follows that conspiracy exposure should have a *lessened* effect on trust in more scandalous environments. Having already been diminished by ongoing political scandals, trust in government may be so low in scandal-heavy climates that it cannot be pushed downward any further by modest conspiracy exposure. This leads us to **Hypothesis 3**: *In a scandal-laden environment, conspiracy theories should have a more modest effect on trust in government than in a less scandal-heavy climate.* Taken together, we term these three hypotheses the *Vicious Cycle of Cynicism*: scandals and/or conspiracies reduce trust in government, which in turn makes the next scandal or conspiracy more believable.

2 Data and Methods

To test our hypotheses, we compare new data collected in May 2013 to data from January 2013 reported in Einstein and Glick (2014). The key conspiracy theory treatment and experimental manipulations are identical but the political climates were markedly, and intentionally, different. May 2013 was, to say the least, challenging for the Obama administration. The President was beset by a nonstop onslaught of scandals, including accusations that his administration had concealed information about the killing of Americans (including an ambassador) in Benghazi, inappropriately obtained Associated Press reporter phone records, and used the Internal Revenue Service to unfairly target conservative political organizations for additional scrutiny.

Figure 1 illustrates the sharp increase in political scandals in May 2013. Using a LexisNexis search we plot the number of mentions of "political scandal" per month in US newspapers from January-June 2013. While mentions of political scandals hovered below 200 per month between January and April 2013, they roughly tripled in May as the IRS, and

think mechanism, it is also plausible that, in high-scandal times, people are more willing to admit aligning with the conspiracy theorists.

AP phone tapping controversies emerged and new Benghazi allegations developed (see also Nyhan (2013)).

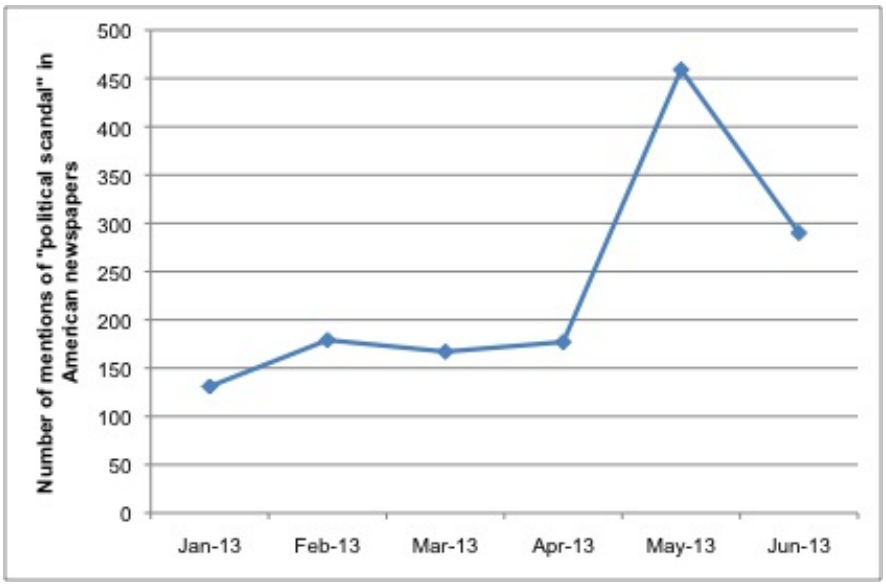


Figure 1: Number of mentions of "political scandal" in US newspapers, January-June 2013. Source: LexisNexis search.

As the figure shows, the climate in which we fielded the experiment in May differed from the climate in January (which was representative of other months) when the original Einstein and Glick data were collected. This sharp difference allows us to test our hypotheses concerning the effects of a high-political-scandal climate on the relationship between conspiracies and trust in govenrment.

2.1 BLS Conspiracy and Experimental Manipulations

Hypothesis 1 implies a simple comparison. It predicts that conspiracy belief will increase in periods of high political scandal. We therefore need a measure of conspiracy belief separate from Benghazi, the IRS, and the AP phone tapping scandals. We focus on the claim—most prominently promulgated by former General Electric CEO Jack Welch (but also by Donald Trump and others)—that the Bureau of Labor Statistics (BLS) manipulated unemployment data for political reasons. This conspiracy claim was actually fairly widespread during the fall

2012 presidential election—when Welch made his comments—-but faded from mainstream political consciousness by the time the experiments were run in January and May 2013. This conspiracy is a good one to use because it is unrelated to the scandals that were salient in May and because it shares important attributes with other conspiracy theories (see Einstein and Glick, 2014).

Our measure of conspiracy belief also mirrors that in the January experiment: "Do you think that recent monthly employment data from the Bureau of Labor Statistics are always calculated as accurately as possible or are they politically manipulated?" Respondents were then offered two options: (1) "Calculated as accurately as possible;" or (2) "Politically manipulated." Figure 2 illustrates the newspaper article we use as our conspiracy: it describes Jack Welch's comments about purported BLS data manipulation, while including both economic data and a compelling rebuttal to the cynical claim. (Einstein and Glick (2014) include a longer discussion of how the treatment article was assembled from real news reports, the choices that went into it, and how it prioritizes realism and external validity.) Using our belief measure and Welch article, we evaluate H1 by assessing whether belief in the BLS conspiracy after exposure to Welch's claim increased, stayed the same, or decreased between January and May 2013.

Our second hypothesis relates to the finding in Einstein and Glick (2014) that people who were exposed to the explicit conspiracy allegation were actually *less* likely to say the data were manipulated than people in control conditions. They attributed this difference to cognitive processing. H2 predicts that, in a higher scandal climate respondents will now find it more plausible in light of the myriad accusations made against the government. To investigate this hypothesis, we compare people who read the Welch article to an article that looked the same and presented the same economic data, but did not include any reference to Welch's conspiracy comments or the response to them (Figure A2 in the appendix).

Finally, we also use the conspiracy exposure in Figure 2 to evaluate *Hypothesis 3*, which anticipates a smaller (or a nonexistent) conspiracy exposure effect on trust in a scandal-

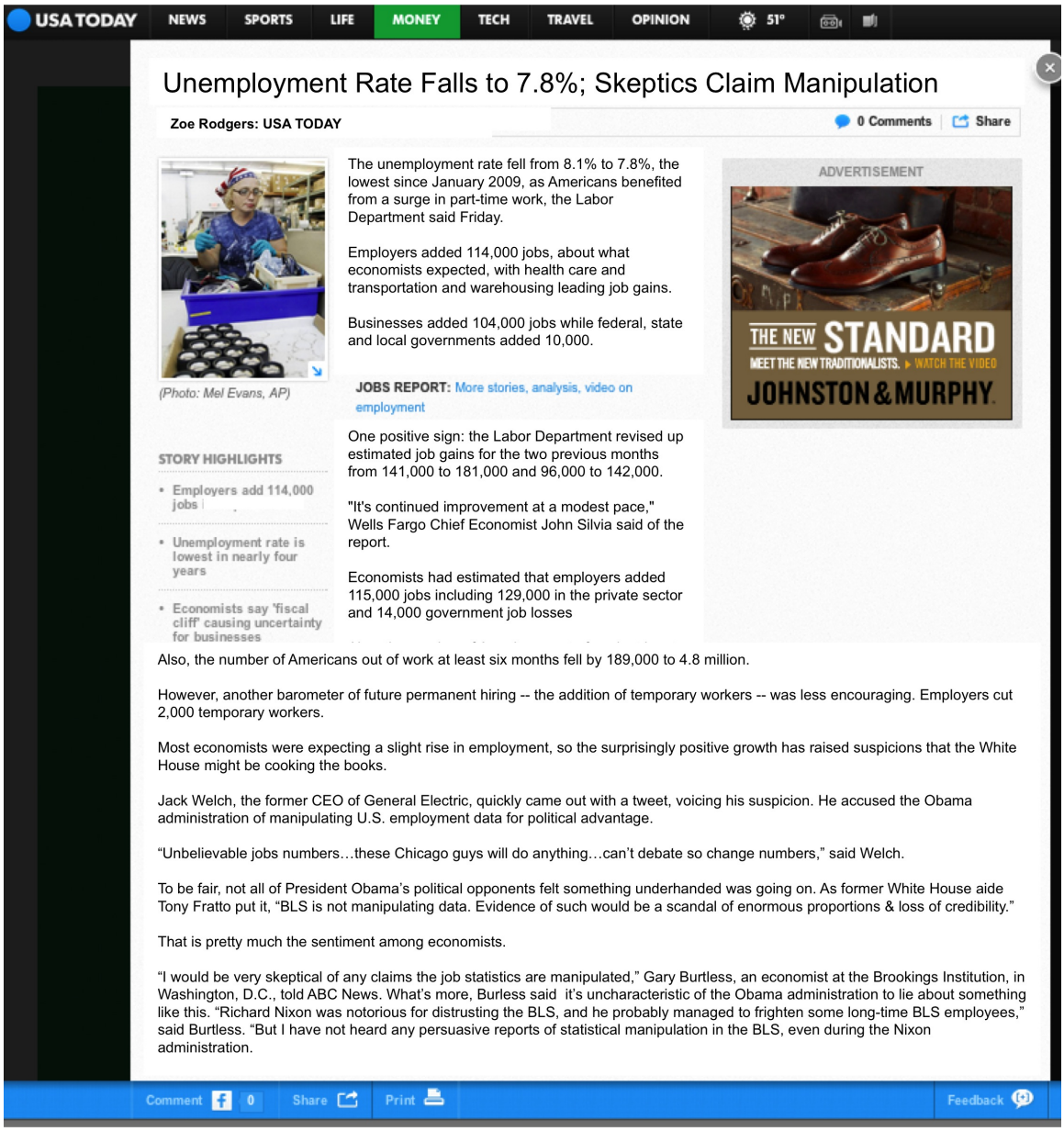


Figure 2: Experimental Exposure to a Conspiracy Theory

heavy climate. To assess this hypothesis, we compare people who read the conspiracy article and were NOT asked about their beliefs in it to a control group who neither read an article nor were asked about their conspiracy beliefs. Consistent with Einstein and Glick (2014), we do not ask the treatment group a conspiracy belief question here because incorporating a question about beliefs affects trust responses. Our trust measure—the predicate to H1 and the dependent variable in H3—is derived from a widely used Gallup poll question, which asks respondents to rate their confidence in multiple government institutions, ranging from federal to local, on a four point scale. It reads "below is a list of institutions in American society. Please indicate how much confidence you have in each one." The four options are "very confident," "somewhat confident," "not so confident," "not confident at all." We provide a variety of government institutions, including the U.S. Census Bureau, the Food and Drug Administration, the President, Congress, and local police. Asking about institutions separately rather than asking about trust in general is important (and rare (see also Gadarian and Albertson, 2012)) because we should expect different impacts on trust in different institutions depending on how tightly coupled they are with the BLS.

2.2 Data Collection

We collected data using participants recruited with Amazon's Mechanical Turk (MTurk). As with our other methodological choices, this matches the approach in the low-scandal period experiment to which we compare our May results.[2] MTurk is an online crowdsourcing marketplace. Survey experiments using MTurk data have recently been published in journals such as the *American Political Science Review* and *American Journal of Political Science* (Huber, Hill and Lenz, 2012; Grimmer, Messing and Westwood, 2012; Arceneaux, 2012; Healy and Lenz, 2014; Dowling and Wichowsky, 2014). As is typical with MTurk samples, our median participant is younger and more liberal than the average American. While MTurk samples are less representative than the highest quality national samples, they are

[2]We also insured that participants in the May experiment had not participated in January one.

often more representative than other convenience (and especially student) samples and they have been used to replicate classic results (Berinsky, Huber and Lenz, 2012; Buhrmester, Kwang and Gosling, 2011). Table A1 (in the appendix) appends demographic information about the January and May subject groups onto the demographics table in Berinsky, Huber and Lenz (2012). Our participants were paid .75, which is consistent with mainstream rates on MTurk. We restricted participation to those in America who had at least a 95% approval rate on at least 50 HITs per the recommendations in (Berinsky, Huber and Lenz, 2012).

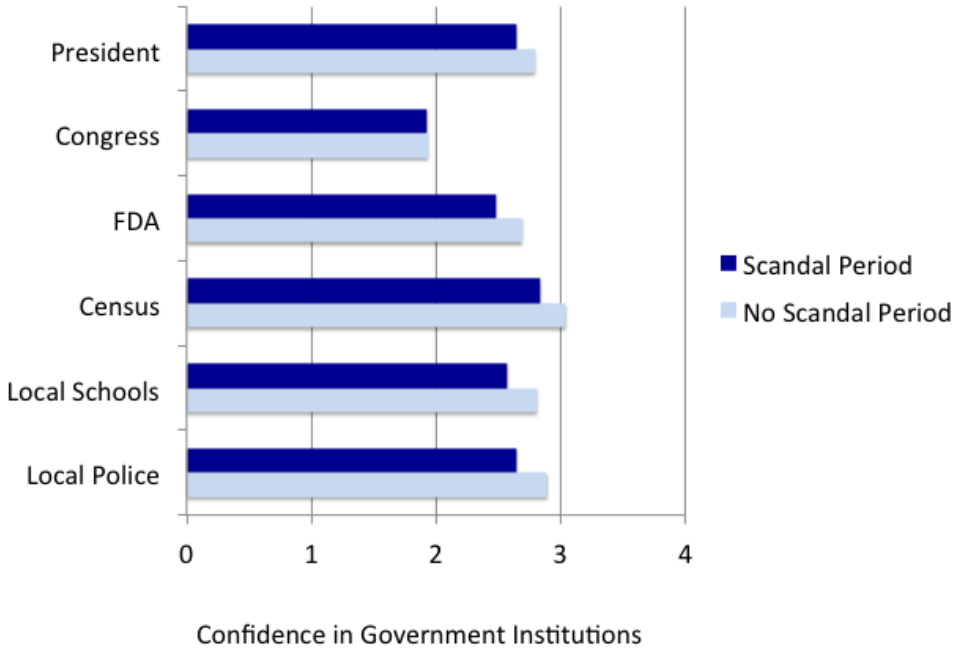
3 Results

The assumption that the May period was a high-scandal, high-cynicism, and thus low-confidence period is central to all of our hypotheses. We therefore begin by simply verifying that confidence in government was indeed lower. Figure 3 displays baseline trust in each period for a variety of institutions, including the Presidency, Congress, the FDA, the U.S. Census Bureau, local schools, and local police. These confidence data were derived from the control group in each experiment. These respondents were neither asked a question about conspiracy belief nor exposed to the BLS conspiracy before assessing confidence. Trust was indeed consistently lower in the period around the IRS, Benghazi and AP scandal coverage. While the differences are not dramatic (and in some cases not statistically significant, as illustrated in Table A3), the relative effects are constant across institutions. These results are thus consonant with our empirical predictions: the scandal-heavy climate in May was at least associated with—if not a direct cause of—lower levels of trust in government.

3.1 Hypothesis 1: Conspiracy Belief

Our most basic and important empirical question is whether this high-scandal, low-trust environment spurred greater belief in the BLS conspiracy. As described above, our dependent variable is the response to a question about whether the BLS statistics are politically

Figure 3: Confidence in government institutions in control group (no articles and no conspiracy questions) by high and low-scandal period

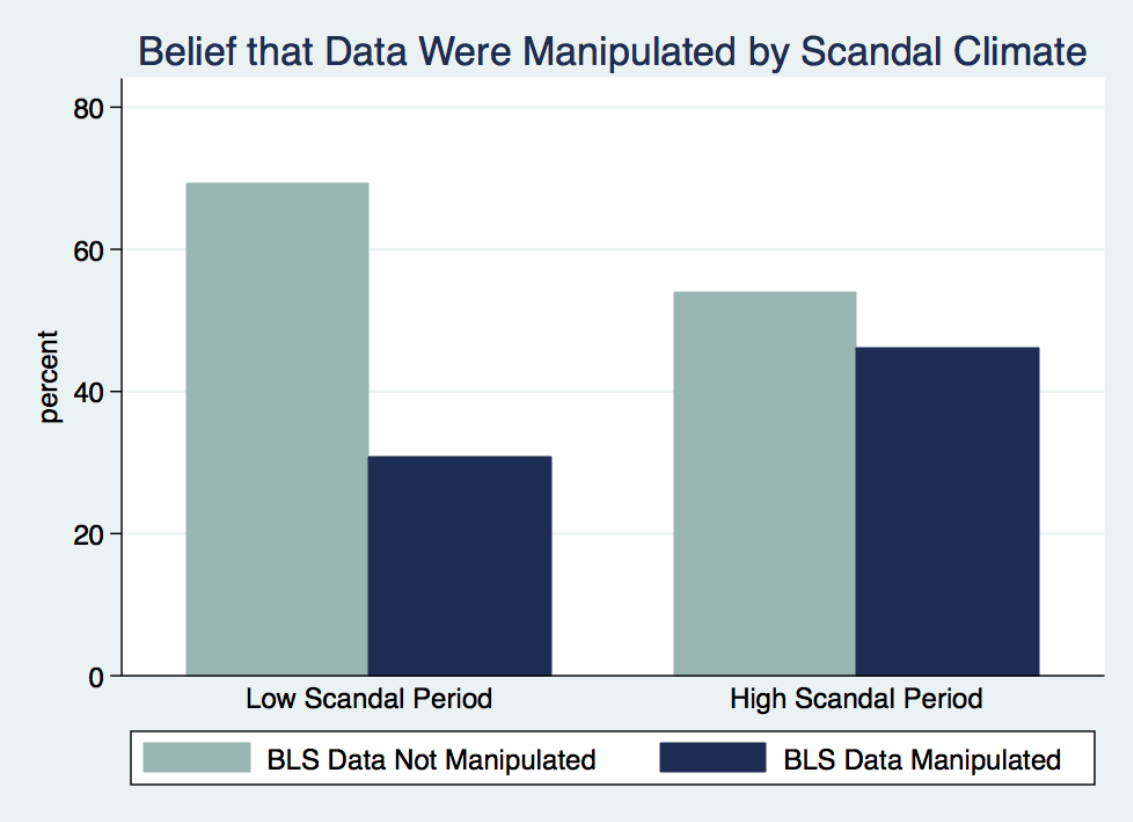


Ns= 243 low-scandal period, 109 high-scandal period. Low scandal period data from (Einstein and Glick, 2014).

manipulated or calculated as accurately as possible. Figure 4 illustrates this comparison. In January 2013, roughly 30 percent of respondents who reach the Welch allegations believed that BLS data were politically manipulated. In May, during the high-scandal period, that figure was nearly 50 percent.

While we believe the simple comparison in the table sufficiently tests our hypothesis, we also estimated the effect of the scandal environment more rigorously in Table 1 by controlling for other variables (partisanship, age, political knowledge, and racial resentment) that could plausibly affect belief in the conspiracy claim. Many of these also capture places where the MTurk sample deviates from the national population (see Einstein and Glick, 2014, for more info). The key variable in this probit model is the *High-Scandal Period* indicator which

Figure 4: Percent of respondents who believe that BLS data were manipulated after reading the conspiracy article by experiment



Ns= 221, 102 respectively. Low scandal period data from (Einstein and Glick, 2014).

differentiates the May data (high-scandal) from the January experiment (low-scandal). The model shows a large and significant positive effect on expressed conspiracy belief as a result of being a participant in the high-scandal period version of the experiment. More substantively, holding other variables at their means, moving from the low- to high-scandal period increases the predicted probability of saying BLS data are manipulated by 25%. Finally, the model also shows that, unsurprisingly, Republican partisanship (higher on a seven-point scale) and racial resentment also increase avowed belief in the conspiracy claim.

3.2 Hypothesis 2: Stop and Think

Hypothesis 2 predicts that, in the high cynicism period, we should not observe the same *lower* admitted conspiracy belief after the conspiracy treatment relative to the control article that

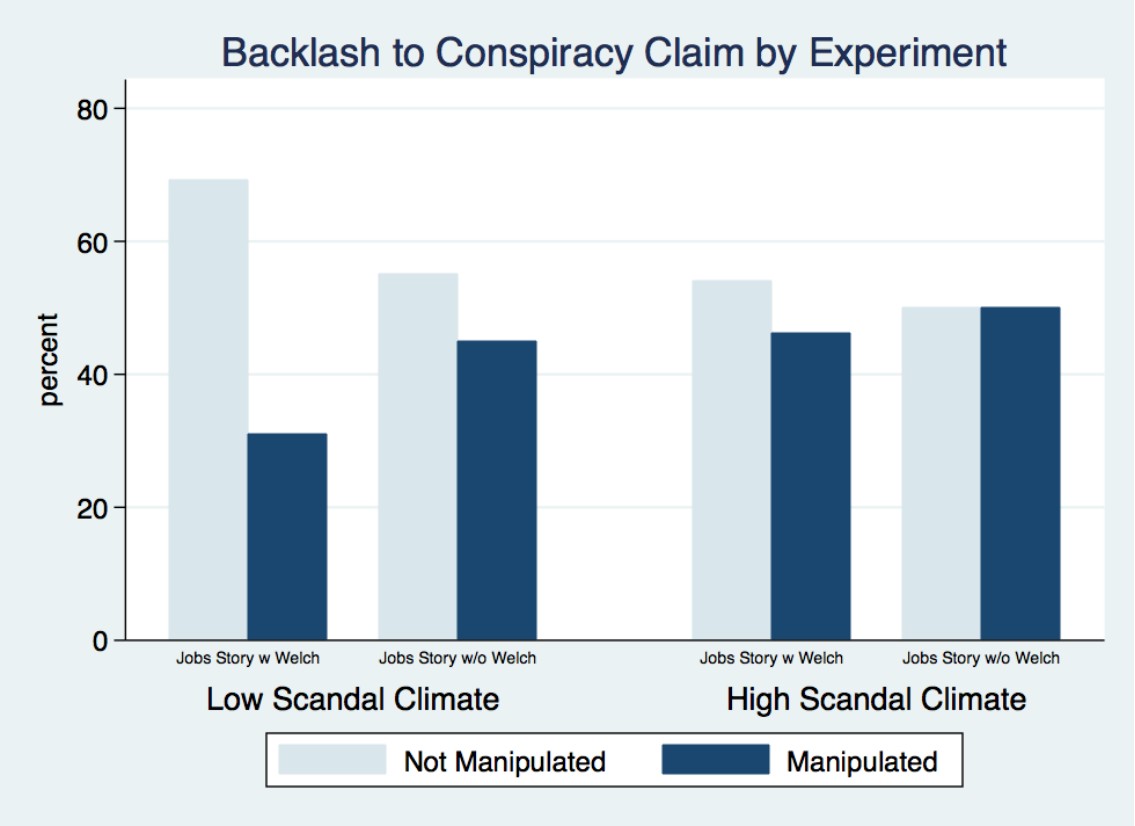
Table 1: Probit models for believing that data were manipulated for those who read the Welch story by scandal climate.

(1) VARIABLES	Manipulated
High Scandal Period	0.69***
	(0.25)
Age	0.00
	(0.01)
Partisanship	0.23***
	(0.05)
White	-0.17
	(0.22)
Female	-0.12
	(0.18)
PolKnowledge	0.02
	(0.11)
RacialResent	0.32***
	(0.10)
Constant	-2.42***
	(0.43)
Observations	288

Standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

manifested in the January experiment. We test this expectation in Figure 5. Here, we compare people who read the article with the Welch claim to those that read the article with the same jobs data but without the conspiracy allegations or counter-argument.

Figure 5: Percent of respondents who believe that BLS data were manipulated after reading article (with and without conspiracy claim) by experiment



Ns= 136, 138, 102, 108 respectively. Low scandal period data from (Einstein and Glick, 2014).

The figure provides strong support for the prediction that the stop and think effect erodes in high-scandal times. The left side reproduces the January results: fewer respondents agreed that the BLS data were manipulated after being forced to confront a specific allegation. Conversely, the right side shows that this effect vanishes when respondents are confronted with an identical conspiracy allegation during the high-scandal period. We corroborate these graphical findings with regressions and statistical controls, as with *Hypothesis 1*. More details are available in the appendix in Table A2.

Our interpretation is that this shift between January and May is a consequence of the

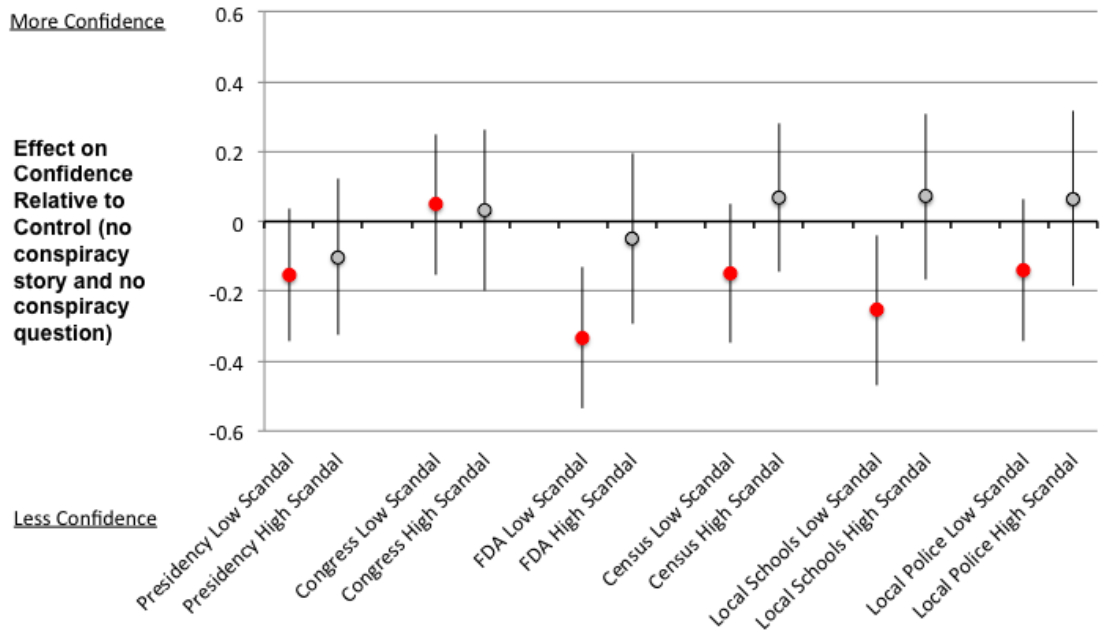
conspiracy claim becoming more plausible in the scandal-ridden period, and that this is strong evidence of the real world spilling over into a survey experiment. When respondents are confronted with both a specific conspiracy allegation and a question about their belief in that conspiracy, they are forced to cognitively engage with that claim and determine whether they believe it to be plausible. In January, when there were few ongoing political scandals, participants did not find the specific BLS allegation particularly believable. Conversely, by May, the combination of the other controversies rendered the BLS claim far more plausible.

3.3 Hypothesis 3: Confidence in Government

The final hypothesis—*Hypothesis 3*—concerns trust in government. Here, we expect that the conspiracy claim in the high cynicism period will have a smaller effect on trust in government institutions than in the low-scandal period. We anticipate less impact on confidence as a consequence of diminishing marginal effects and/or the fact that people have already either reduced their confidence or steeled themselves in response to the other scandal coverage. Ongoing scandals would have already lowered confidence in government to such an extent that conspiracy exposure would have, at best, a very modest effect. We estimate this effect by comparing those who were assigned the Welch article but not asked about conspiracy beliefs to those in the pure control group who neither read an article nor were asked about beliefs.

We already know (see Figure 3 above) that trust in government is lower during the scandal-heavy period. Figure 6 turns, then, to the effect of the Welch article relative to our control groups for both the high- and low-cynicism periods. It illustrates coefficient estimates and 95 percent confidence interval estimates for six government institutions. The coefficients are outputs from two sets of seemingly unrelated regressions—one for the high- (Table A5) and one for the low-scandal period (Table A4)—and include our standard control variables. The figure plots the coefficients relative to the control for each condition and each institution. Negative estimates mean the Welch claim reduced confidence.

Figure 6: Comparison of trust effects between high and low-scandal environment across institutions. Estimated effect of reading story without being asked about belief relative to baseline confidence in control group.



Points are coefficients from seemingly unrelated regression models of confidence with controls for partisanship, age, political knowledge and racial resentment. Bars indicate 95 % confidence intervals on estimates. Low scandal period data from (Einstein and Glick, 2014).

Figure 6 shows that, consistent with *Hypothesis 3*, the Welch story reduced confidence in the low-scandal period, but not during the high-scandal environment. The estimates in the high-scandal period are all very close to zero, whereas nearly all of the estimates in the low-scandal period are substantially negative. Table A6 in the appendix confirms these results by including both the high- and low-scandal periods in the same model with a variable indicating the IRS scandal and an interaction for the scandal period and the Welch article. The model reveals that the Welch article had an effect when controlling for the scandal environment, but not when it was interacted with the high-scandal period.

4 Discussion: Substantive Importance, Replication and Experimental Methodology

Combined, our results make substantive and methodological contributions that are inseparable from each other. The dramatically different treatment effects speak to the importance of macro variables on conspiracy attitudes. This is notable in a literature dominated by individual psychological mechanisms. Our cycle of cynicism theoretical framework leads to unfortunate conclusions about substantive outcomes of interest to scholars of political behavior, democracy, and public policy. Moreover, our method for assessing the effects of unrelated scandals on conspiracy beliefs relies on, and points to, under-appreciated realities of ostensibly controlled experiments. We join a select few other studies in showing that social scientists should be attentive to the potential of real world events to alter experimental outcomes. Our results suggest that experiments do not unfold in a vacuum, but interact with the outside world in theoretically reasonable ways.

The results described above raise two questions. The first is whether these findings indicate meaningful substantive effects or are simply evidence of a failure to replicate the original (Einstein and Glick, 2014) experiment. The second question concerns the much wider methodological implications.

4.1 Replication

Do the two experiments together tell a compelling and consistent theoretical story about conspiracy theories, beliefs, and confidence in government? Or, have we simply failed to replicate the January results due to random variation? Clearly the replicability question is crucial so we address it first.

The first argument in favor of real differences rather than a simple failure to replicate is based on the findings' fit with theoretical expectations. We intentionally chose to conduct an experiment focused on conspiracies in this scandal-laden environment because theory

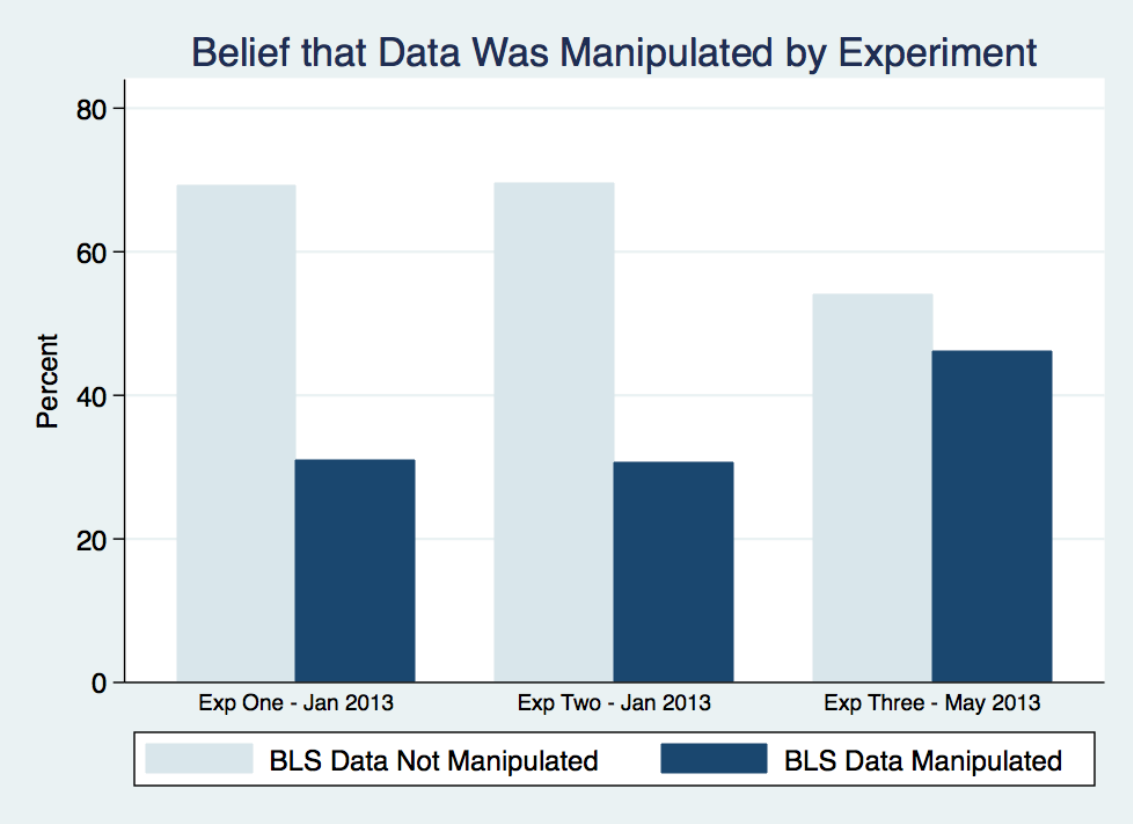
suggested we should find different results in the event of a real world spillover. As we showed earlier in Figure 1, the environment in May—when the media's focus on scandals was overwhelming—was significantly different than in January. Moreover, all of the three pieces fit together in a way consistent with theory, which should at least assuage doubts that the different findings from January to May are just the result of random variation. Indeed, our work mirrors Neely (2007) on these points. His evidence for the importance of context on the effects of question wording is also based on theoretically coherent (based on anxiety) differences between a question asked shortly after 9/11 and the same question asked four years later.

In addition, the samples in the Einstein and Glick (2014) experiment offer us further opportunity to demonstrate that our results are not simple a failure to replicate. Indeed, the January results are actually from two separate experiments on two different MTurk samples. Importantly, the Welch article treatment was repeated in two separate January samples and combined into most of the results reported in the original article. Here, to address the replication question it is helpful to separate them. Figure 7 depicts reported belief in the conspiracy for people in the Welch article condition across three samples (two low-scandal (January) and one high-scandal (May)). It reveals that the percent of respondents believing the data were manipulated was almost identical in the first two, and dramatically different in the third—the high-scandal period. In other words, the baseline against which we are comparing our spillover results has already been replicated. Moreover, Figure 8 illustrates that—as with these belief results—the stop-and-think effect was also replicated across the two January experiments before the results deviated in May.[3] Across both January experimental conditions, exposing respondents to both an article about the conspiracy and a question about their beliefs reduced their avowed belief in the BLS conspiracy. The left

[3]The replication of the backlash between the two experiments is not perfect. The second January experiment did not include the jobs story without the Welch article. Instead, as the figure shows, we compare jobs-with-Welch to jobs-without-Welch from one January experiment one to jobs-with-Welch to no article in the second January experiment two. Both show similar levels of people being less likely to believe the conspiracy after being exposed to the specific allegation than they are without it.

side of the figure shows this in the first low-scandal experiment, while the right illustrates the same effect in the second. Both parts of this figure stand in sharp contrast to the high scandal period findings which reveal no such "backlash" type effect.

Figure 7: Percent of respondents who believe that BLS Data were manipulated after reading article by experiment

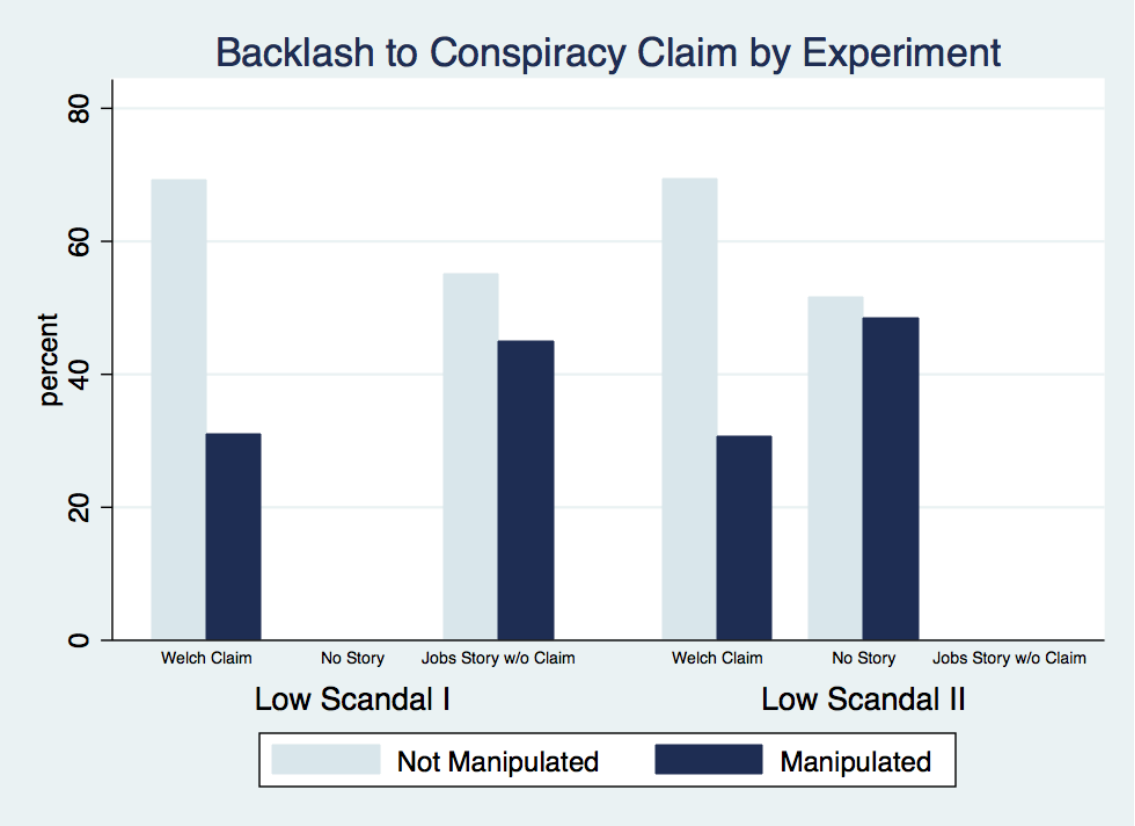


Ns= 136, 85, 102 respectively. Low scandal period data (Exp one and two) from (Einstein and Glick, 2014).

4.2 Methodological Lessons

While the findings contribute to the growing substantive literatures concerned with conspiracy politics and trust, our broader contribution is methodological. The evidence we provide for the real world leaking into an ostensibly controlled experiment has broad implications for one of the most popular methods in political science and sister disciplines. Our most fundamental finding is that contemporaneous real world conditions can dramatically shape experimental results. As we noted in the introduction, this work parallels similar find-

Figure 8: Replication of backlash in the low-scandal experiments



N = 136, 138, 89, 95 respectively. low-scandal I included jobs story without Welch claim but did not include the no story with conspiracy belief question condition. low-scandal II was the opposite. Thus, replication relies on comparing the Welch article to two different types of controls. Data from (Einstein and Glick, 2014).

ings concerning question wording (Neely, 2007) and spatial variation in experimental effects (Kriner and Shen, 2014). This is important because these results tell a strong story together. Combined, they provide evidence that salient and relevant real world conditions mediate experimental findings in areas ranging from party-ID question wording, to war casualties, to reading about a conspiracy theory.

These findings do not invalidate the results from hundreds (or more) published experiments, but they do suggest that experimental effects are mutable depending on the political environment. This reality has important implications for such designs going forward and for the interpretation of prior findings. We hope our work at least leads to a broader discussion about best practices. To start this conversation, we offer three concrete recommendations

which we suspect at least some researchers are already doing. First, we suggest researchers at least discuss any potentially relevant environmental conditions and how these external factors might affect experimental results results. For example, experiments about racial priming should look at the national salience of racial issues (e.g., mentions of immigration or affirmative action on national news networks) and consider how this context shapes their results. Second, we recommend that researchers consider delaying experiments when needed. For example, if one is going to field an experiment about foreign policy attitudes and an international crisis breaks out, researchers might consider waiting until the situation calms (or, at least, rerun their experiment in more typical times). Third, we hope our work inspires other theoretically driven replication studies in which our colleagues intentionally time experiments to take advantage of real world conditions. Indeed, the flip side of delaying an experiment in the event of a crisis might be to have an experiment at the ready should a real world crisis emerge. On the one hand we show that the real world can confound experiments, but on the other, as our work and Kriner and Shen (2014) show, the fact that pertinent exogenous conditions mediate experimental treatments can produce great opportunities for researchers.

References

Aaronovitch, David and James Langton. 2010. *Voodoo Histories: The Role of the Conspiracy Theory in Shaping Modern History*. Wiley Online Library.

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

Banas, John A and Gregory Miller. 2013. "Inducing Resistance to Conspiracy Theory Propaganda: Testing Inoculation and Metainoculation Strategies." *Human Communication Research* 39(2):184–207.

Barabas, Jason and Jennifer Jerit. 2010. "Are Survey Experiments Externally Valid?" *American Political Science Review* 104(02):226–242.

Berinsky, Adam .J. 2013. "Rumors, Truths, and Reality: A Study of Political Misinformation." *Unpublished Working Paper (V3.1)* .

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.

Brotherton, Robert, Christopher C French and Alan D Pickering. 2013. "Measuring Belief in Conspiracy Theories: The Generic Conspiracist Beliefs Scale." *Frontiers in psychology* 4.

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.

Burden, Barry C. and Casey A. Klofstad. 2005. "Affect and Cognition in Party Identification." *Political Psychology* 26(6):869–886.

Coady, David. 2006. *Conspiracy Theories: The Philosophical Debate*. Ashgate Publishing, Ltd.

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.

Druckman, James N. and Thomas J. Leeper. 2012. "Learning More From Political Communication Experiments: Pretreatment and its Effects." *American Journal of Political Science* 56(4):875–896.

Einstein, Katherine and David M. Glick. 2014. "Do I think BLS data are BS? The Consequences of Conspiracy Theories." *Political Behavior* .

Gadarian, Shana and Bethany Albertson. 2012. "Anxious Politics: The Effect of Anxiety on Political Trust." *APSA 2012 Annual Meeting Paper* .

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

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.

Hopkins, Daniel J. 2010. "Politicized Places: Explaining Where and When Immigrants Provoke Local Opposition." *American Political Science Review* 104(01):40–60.

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.

Keeley, Brian L. 1999. "Of Conspiracy Theories." *The Journal of Philosophy* pp. 109–126.

Kinder, Donald R. 1998. "Communication and Opinion." *Annual Review of Political Science* 1(1):167–197.

Kriner, Douglas L. and Francis X. Shen. 2014. "Reassessing American Casualty Sensitivity The Mediating Influence of Inequality." *Journal of Conflict Resolution* 58:1174–1201.

Lewandowsky, Stephan, Werner GK Stritzke, Alexandra M Freund, Klaus Oberauer and Joachim I Krueger. 2013. "Misinformation, Disinformation, and Violent Conflict: From Iraq and the War on Terror to Future Threats to Peace." *American Psychologist* 68(7):487.

Mulligan, Kenneth and Philip Habel. 2013. "The Implications of Fictional Media for Political Beliefs." *American Politics Research* 41(1):122–146.

Mutz, Diana C. 1994. "Contextualizing Personal Experience: The Role of Mass Media." *The Journal of Politics* 56(03):689–714.

Neely, Francis. 2007. "Party Identification in Emotional and Political Context: A Replication." *Political Psychology* 28(6):667–688.

Nyhan, Brendan. 2013. "The Scandal Attention Cycle." *Columbia Journalism Review* .

Oliver, J Eric. and Thomas J. Wood. 2014. "Conspiracy Theories and the Paranoid Style (s) of Mass Opinion." *American Journal of Political Science* 58(4):952–966.

Transue, John E., Daniel J. Lee and John H. Aldrich. 2009. "Treatment Spillover Effects Across Survey Experiments." *Political Analysis* 51(1):78–91.

Uscinski, Joseph E, Casey Klofstad and Matthew Atkinson. 2014. "Why Do People Believe in Conspiracy Theories? The Role of Informational Cues and Predispositions." *Paper presented at the 2014 American Political Science Association Annual Meeting* .

Uscinski, Joseph E and Joseph M Parent. 2014. *American Conspiracy Theories*. New York University Press.

Appendix

4.3 2014 Survey Experiments' Lack of Discussions of Real World Effects.

We conducted a somewhat cursory look at all of the papers using survey experiments (or something close to them) published (in the journals, not online firsts) in the APSR, AJPS and JOP in 2014. Our goal was simply to understand standard practice for discussing (or not discussing) real world context and survey experiments. Figure A1 lists all of the papers that we found and looked at. In our readings, none of these papers discussed what was going on in the world at the time the experiments were conducted. We note that some (e.g. Malhotra and Margalita and Healy and Lenz) did mention that their main effects replicated in multiple experiments conducted at different times, and at least one (Boudreau and MacKenzie) was intentionally conducted right before an election and did implicitly note context. Our sole purpose for this analysis is to document that standard practice does not include discussing potential environmental confounds and/or interactions with treatments. It is likely that nothing was going on that would have affected many of these papers. What matters for our purposes is that the standard presentation of survey experiments treats results and independent main effects and scholars, reviewers, and editors do seem overly concerned with the potential for the real world to threaten validity or interpretation.

4.4 Design and Data

Figure A2 shows the control article we used for estimating backlash. It includes the good news about economic data that is in the Welch conspiracy article but does not include anything about the conspiracy. [4]

[4]The January experiment also included a second control article that described positive economic statistics that had nothing to do with the BLS, but rather, concerned the Oregon craft beer market. In all article conditions, respondents were asked two questions about the clarity of the data presented in the assigned article and about the media's use of statistics in general. We placed these questions prior to our items exploring belief in conspiracy to moderately mask the purpose of the study and distract respondents from thinking about the experimental manipulation.

Figure A1: All survey experiments (or similar designs) we found published in the APSR, AJPS, and JOP in 2014. To our knowledge, none discussed the potential methodological concerns our paper focuses on.

Journal	Title	Authors
APSR	Selling Out?: The Politics of Navigating Conflicts between Racial Group Interest and Self-interest	White, Laird, and Allen
APSR	The Political Mobilization of Ethnic and Religious Identities in Africa	McCauley
AJPS	Preferences for International Redistribution: The Divide over the Eurozone Bailouts	Mechtel, Hainmueller and Margalit
AJPS	Partisans in Robes: Party Cues and Public Acceptance of Supreme Court Decisions	Nicholson and Hansford
AJPS	Partisanship in a Social Setting	Klar
AJPS	Substituting the End for the Whole: Why Voters Respond Primarily to the Election-Year Economy	Healy and Lenz
AJPS	Informing the Electorate? How Party Cues and Policy Information Affect Public Opinion about Initiatives	Boudreau and MacKenzie
AJPS	The Conditionality of Vote-Buying Norms: Experimental Evidence from Latin America	Ocantos, de Jonge and Nickerson
AJPS	The Power of Partisanship in Brazil: Evidence from Survey Experiments	Samuels and Zucco Jr.
JOP	Expectation Setting and Retrospective Voting	Malhotra and Margalita
JOP	Self-Fulfilling Misperceptions of Public Polarization	Ahler
JOP	Linking Issue Stances and Trait Inferences: A Theory of Moral Exemplification	Clifford
JOP	Understanding the Party Brand: Experimental Evidence on the Role of Valence	Butler and Powell
JOP	Priming under Fire: Reverse Causality and the Classic Media Priming Hypothesis	Hart and Middletona
JOP	Visual Political Knowledge: A Different Road to Competence?	Prior

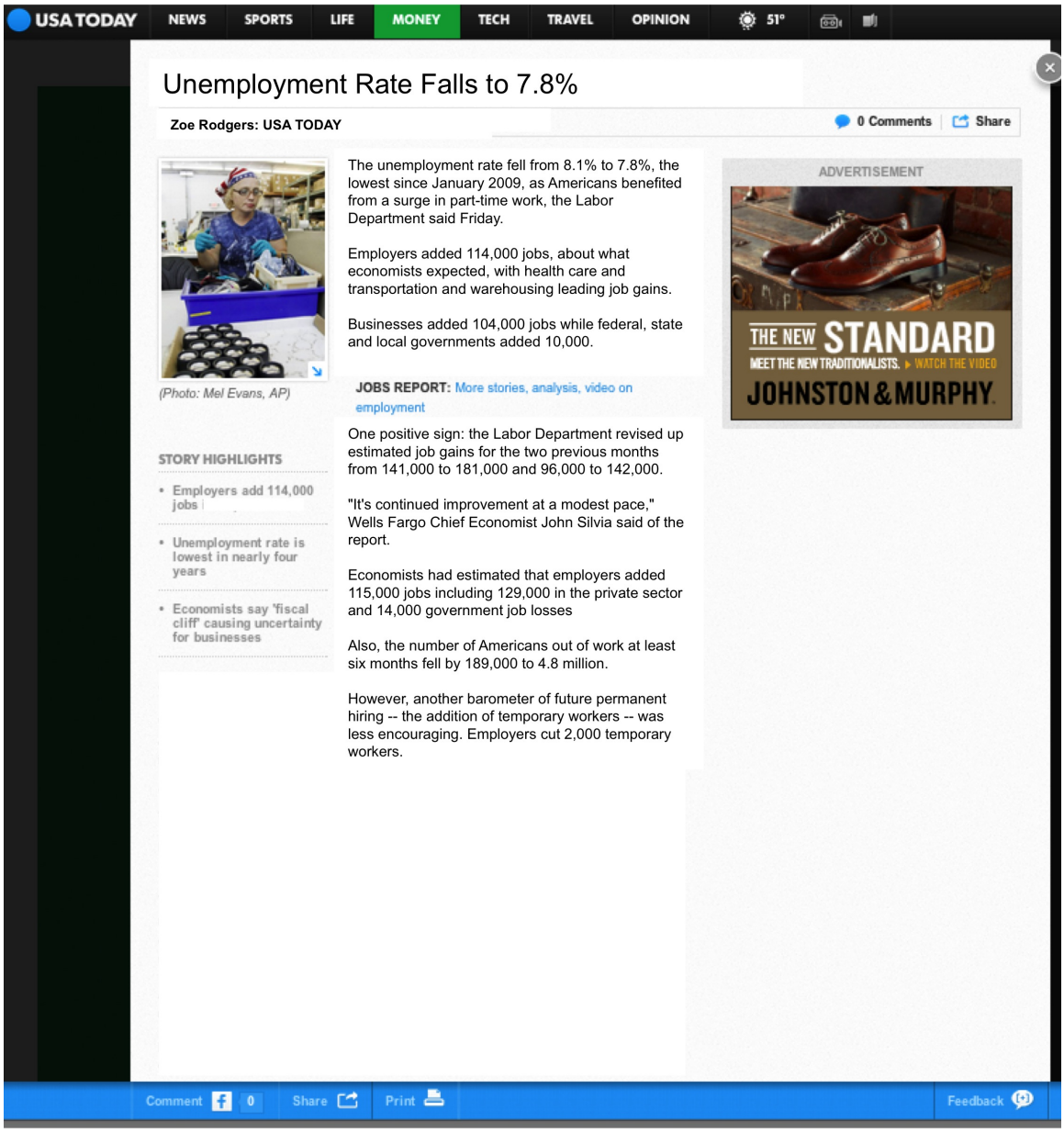


Figure A2: Experimental story: Straight jobs data without Welch

Table A1: Sample Demographics by Experiment: The table compares our sample demographics to those found in another MTurk study and to other high quality surveys.

			Internet Samples		Face to Face Samples	
Variable	**Jan '13**	**May '13**	**BHL Turk**	**ANES-P 08-09**	**CPS 08**	**ANES 08**
% Female	47.2	50.4	60.1	57.6	51.7	55.0
% White	80.3	82.2	83.5	83.0	81.2	79.1
% Black	7.6	8.3	4.4	8.9	11.8	12.0
% Hispanic	6.0	5.9	6.7	5.0	13.7	9.1
Age (years)	35.1	34.3	32.3	49.7	46.0	46.6
Party ID (7 pt.)	3.1	3.0	3.5	3.9		3.7
Ideology (7 pt.)	3.2	3.3	3.4	4.3		4.2
Education	15.3 yrs	15.3	15.3 yrs	16.2 yrs	13.2 yrs	13.5 yrs
Income (median)	30-49K	30-49K	45K	67.5K	55K	55K

"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 those corresponding to our experiments are reproduced from Table 3 in Berinsky, Huber and Lenz (2012).

Table A1 compares the demographics of our low (Jan) and high (May) scandal period experiments to other survey samples including the MTurk sample that was used to replicate classic experiments and to high quality representative samples. Our two MTurk samples were very similar to each other and deviated from non Turk samples in the expected ways.

4.5 Stop and Think

Table A2 supports the backlash figure and discussion in the text. It shows that the Welch claim has a negative effect on conspiracy beliefs relative to the baseline of the jobs article without the Welch claim, but that the interaction of the Welch claim and high-scandal period have a positive effect on belief that essentially cancels out the main backlash effect. In other words, the model confirms what the tabulation shows. There is only a stop and think type backlash in the low-scandal period.

Table A2: Probit models for believing that data were manipulated as a function of exposure to the Welch story, IRS scandal period, and controls

EQUATION	VARIABLES	(1) Manipulated
Manipulated	High Scandal	0.29
		(0.22)
	WelchClaim	-0.68***
		(0.18)
	High Scandal x Claim	0.58**
		(0.26)
	Age	-0.00
		(0.01)
	Partisanship	0.19***
		(0.04)
	White	-0.05
		(0.17)
	Female	0.04
		(0.13)
	PolKnowledge	0.14*
		(0.09)
	RacialResent	0.37***
		(0.07)
	Constant	-1.99***
		(0.35)
	Observations	443

Standard errors in parentheses

$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

4.6 Confidence

Table A3 estimates the effect of the low/high-scandal variable on baseline levels of confidence in government. Each column corresponds to a different institution and the confidence scores are from respondents who neither read an article, but were asked about their conspiracy beliefs. The sign on the IRS variable - indicating the high-scandal period - was always negative consistent with lower confidence in government but it was only significant for a couple of institutions.

The coefficients and confidence intervals in Figure 6 are derived from the models summarized in Tables A4 and A5. These models estimate the effect on confidence in institutions that results from reading the conspiracy claim. In both, the baseline category is the true control group (no article and no conspiracy questions) and the main variable of interest is the Welch NotAsked variable which indicates that the respondent was in the group that read the conspiracy article but were not asked about their belief that the data were manipulated. The coefficients on this variable indicate the "real world" effect of the conspiracy claim on confidence. Table A6 estimates similar effects without splitting the sample. Instead it includes the IRS variable for the high-scandal group and the interaction between IRS and Welch NotAsked.

Table A3: Seemingly unrelated regression models of baseline (control group) confidence by period

VARIABLES	(1) President	(2) Congress	(3) FDA	(4) Census	(5) Local_Schools	(6) Local_Police
High Scandal	-0.11	-0.22*	-0.23*	-0.04	-0.12	-0.14
	(0.12)	(0.13)	(0.12)	(0.12)	(0.12)	(0.12)
Age	0.00	-0.00	-0.00	0.00	-0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Partisanship	-0.29***	-0.04	-0.06**	-0.08***	0.00	0.00
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
PolKnowledge	-0.02	-0.17***	0.00	0.11**	0.04	0.06
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
RacialResent	-0.09*	0.08	0.04	-0.02	0.06	0.06
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Constant	3.95***	2.31***	2.89***	3.15***	2.61***	2.51***
	(0.21)	(0.22)	(0.22)	(0.21)	(0.21)	(0.22)
Observations	312	312	312	312	312	312
R^2	0.34	0.06	0.03	0.07	0.02	0.03

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table A4: Trust effect on confidence in responses collected before the IRS scandal

VARIABLES	(1) President	(2) Congress	(3) FDA	(4) Census	(5) Local_Schools	(6) Local_Police
Conspiracy w/o Belief Q	-0.15	0.05	-0.33***	-0.15	-0.25**	-0.14
	(0.10)	(0.10)	(0.10)	(0.10)	(0.11)	(0.10)
	0.11	0.63	0.00	0.14	0.02	0.18
Age	0.00	-0.01	-0.00	-0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	0.81	0.15	0.24	0.40	0.71	0.38
Partisanship	-0.30***	-0.04	-0.08***	-0.08***	-0.03	-0.00
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
	0.00	0.23	0.01	0.01	0.34	0.94
PolKnowledge	-0.03	-0.18***	0.05	0.12**	0.01	0.07
	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)
	0.53	0.00	0.34	0.02	0.91	0.18
RacialResent	-0.11**	0.05	0.07	-0.08	0.03	0.07
	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)
	0.04	0.33	0.23	0.14	0.63	0.23
Constant	4.05***	2.43***	2.83***	3.42***	2.77***	2.47***
	(0.20)	(0.22)	(0.22)	(0.22)	(0.23)	(0.22)
	0.00	0.00	0.00	0.00	0.00	0.00
Observations	296	296	296	296	296	296
R^2	0.38	0.07	0.07	0.08	0.02	0.02

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table A5: Trust effect on confidence in responses collected during the IRS scandal

VARIABLES	(1) President	(2) Congress	(3) FDA	(4) Census	(5) Local_Schools	(6) Local_Police
Conspiracy w/o Belief Q	-0.10	0.03	-0.05	0.07	0.07	0.07
	(0.11)	(0.12)	(0.12)	(0.11)	(0.12)	(0.13)
	0.37	0.79	0.69	0.53	0.55	0.61
Age	-0.00	0.00	-0.00	0.00	-0.00	0.01*
	(0.00)	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)
	0.68	0.57	0.38	0.53	0.57	0.05
Partisanship	-0.24***	-0.05	-0.03	-0.08**	0.00	0.04
	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)
	0.00	0.16	0.37	0.02	0.90	0.26
PolKnowledge	-0.00	0.03	-0.10	0.01	-0.09	-0.02
	(0.09)	(0.10)	(0.10)	(0.09)	(0.10)	(0.11)
	0.97	0.75	0.33	0.95	0.35	0.88
RacialResent	-0.09	0.10	0.02	-0.02	0.00	0.02
	(0.06)	(0.06)	(0.07)	(0.06)	(0.07)	(0.07)
	0.15	0.12	0.72	0.78	0.99	0.78
Constant	3.78***	1.71***	2.73***	3.06***	2.78***	2.15***
	(0.24)	(0.25)	(0.26)	(0.23)	(0.26)	(0.27)
	0.00	0.00	0.00	0.00	0.00	0.00
Observations	195	195	195	195	195	195
R^2	0.28	0.02	0.01	0.04	0.01	0.03

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table A6: Combined Trust (IRS*WelchNotAsked interaction

VARIABLES	(1) President	(2) Congress	(3) Supremecourt	(4) Bureaucracy	(5) Local_Services	(6) Churches	(7) Corporations
Conspiracy w/o Belief Q	-0.16	0.04	0.00	-0.24***	-0.20**	0.11	0.02
	(0.10)	(0.10)	(0.10)	(0.08)	(0.09)	(0.13)	(0.10)
	0.11	0.70	1.00	0.00	0.02	0.41	0.86
High Scandal	-0.10	-0.17	0.02	-0.14	-0.19*	-0.09	-0.11
	(0.11)	(0.12)	(0.11)	(0.09)	(0.10)	(0.14)	(0.11)
	0.35	0.15	0.89	0.13	0.06	0.54	0.32
(High Scandal)*(Conspiracy...)	0.05	-0.02	0.07	0.27**	0.28**	-0.18	-0.04
	(0.15)	(0.16)	(0.15)	(0.13)	(0.13)	(0.19)	(0.15)
	0.75	0.90	0.64	0.03	0.04	0.35	0.79
Age	-0.00	-0.00	-0.00	-0.00	0.00	0.01***	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	0.84	0.28	0.55	0.36	0.20	0.00	0.72
Partisanship	-0.28***	-0.04*	-0.03	-0.07***	0.00	0.10***	0.09***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)
	0.00	0.05	0.17	0.00	0.97	0.00	0.00
PolKnowledge	-0.02	-0.14***	0.02	0.05	0.01	-0.08	-0.04
	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)
	0.70	0.00	0.59	0.15	0.73	0.14	0.39
RacialResent	-0.09**	0.09**	0.02	-0.01	0.03	0.16***	0.15***
	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.05)	(0.04)
	0.01	0.03	0.61	0.79	0.45	0.00	0.00
Constant	3.97***	2.19***	2.81***	3.11***	2.66***	1.07***	1.28***
	(0.17)	(0.18)	(0.17)	(0.14)	(0.15)	(0.22)	(0.17)
	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Observations	491	491	491	491	491	491	491
R^2	0.34	0.04	0.01	0.07	0.02	0.12	0.10

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1