REAL-TIME REACTIONS TO A 2012 PRESIDENTIAL DEBATE
A METHOD FOR UNDERSTANDING WHICH MESSAGES MATTER

AMBER E. BOYDSTUN*
REBECCA A. GLAZIER
MATTHEW T. PIETRYKA
PHILIP RESNIK

Abstract How effective are presidential candidates at engaging viewers during debates? To answer this question, we designed a mobile app and conducted a large-scale national study of individual college students’ real-time reactions to the first general election debate of 2012. Previous studies have relied either on real-time but small-sample individual dial reactions or on large-scale public reactions to debates in their entirety, after the fact, and without consideration of specific statements or events within the debates. By contrast, our approach allowed us to collect moment-by-moment data from a large and diverse group of participants in natural settings. The resulting data make it possible to answer questions previously believed to be outside the bounds of systematic inquiry. Here, we explain the method and provide some key findings that illustrate the payoff of our approach. Our study suggests that collecting large-scale, real-time data is feasible and valuable for advancing research on a host of public opinion phenomena.

Amber E. Boydstun is an assistant professor of political science at the University of California–Davis, Davis, CA, USA. Rebecca A. Glazier is an assistant professor of political science at the University of Arkansas at Little Rock, Little Rock, AR, USA. Matthew T. Pietryka is an assistant professor of political science at Florida State University, Tallahassee, FL, USA. Philip Resnik is a professor in the Department of Linguistics and the Institute for Advanced Computer Studies at the University of Maryland, College Park, MD, USA. The authors are deeply indebted to the hundreds of instructors and thousands of students who participated in the 2012 React Labs: Educate project. Timothy Jurka’s masterful skills helped make the project possible in the first place. All-night post-debate data analysis sessions were made possible—and hilarious—by Timothy Jurka, Debra Leiter, Jack Reilly, and Michelle Schwarze. The authors thank Ben Highton for very helpful comments on the manuscript and Drew Stephens for his invaluable technical support. Funding support for this project was provided by a Presidential Studies Center Grant from the University of Arkansas at Little Rock. *Address correspondence to Amber E. Boydstun, University of California–Davis, One Shields Ave., Davis, CA 95616, USA; e-mail: aboydstun@ucdavis.edu.

doi:10.1093/poq/nfu007
© The Author 2014. Published by Oxford University Press on behalf of the American Association for Public Opinion Research. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com
Presidential debates serve a singular role in U.S. elections. Debates uniquely provide candidates unmediated access to a large and diverse audience (Trent and Friedenberg 2008), including marginally attentive citizens (Pfau 2003) and undecided voters (Geer 1988) who use debates to learn about the candidates (Blais and Perrella 2008; Holbrook 1999; Lemert 1993). Indeed, debates are the most visible, widely watched events of a presidential campaign (Benoit, Hansen, and Verser 2003; Schroeder 2008). Yet, despite the importance of debates, we know little about exactly which candidate cues tend to resonate positively with viewers and, just as important, which cues provoke negative affect.

Examining the effects of debate cues requires the ability to track a large sample of viewers’ responses to debates in real time in a natural environment. Toward this aim, we designed a mobile app for use during the first 2012 debate, providing real-time reactions with a level of scale and detail not previously possible. Here, we describe the method we developed and its implementation, along with presenting several key findings that illustrate its value over existing methods.

### Studying Debate Reactions

Past debate research, although impressive in many ways, has been unable to measure the effect of specific candidate messages on individual attitudes. Most mainstream polls collect aggregate data only after a debate has finished (e.g., Holbrook 1999; Shaw 1999), making individual-level conclusions impossible. And most large-scale individual-level research on debates also relies on post-debate evaluations (e.g., Abramowitz 1978; Geer 1988; Hillygus and Jackman 2003; Steeper 1978). Whether surveys are cross-sectional (e.g., Lanoue 1992; Sigelman and Sigelman 1984) or panel designs (e.g., Kraus and Smith 1977; Tsfati 2003), the data cannot differentiate between the effects of the debate itself and other influences, such as media coverage of the debates (Brubaker and Hanson 2009; Fridkin et al. 2007). Moreover, these studies cannot isolate which candidate messages are influencing viewers. Recent work indicates that researchers cannot trust survey respondents to self-report accurately even whether they watched the debate (Prior 2012). Thus, while past research has contributed greatly to our understanding of debate effects (Bartels 2006; Benoit, Hansen, and Verser 2003; Geer 1988; Holbrook 1999), scholars have often been reduced to educated guesswork about which specific candidate cues produce these effects.

A handful of innovative studies have used dial testing to collect real-time data but have been limited by costs and logistical complications associated with specialized hardware, small sample sizes (Kirk and Schill 2011; McKinney, Kaid, and Robertson 2001), and other challenges to external validity such as artificial focus-group settings (Ramanathan et al. 2010) and, in the case of Kirk and Schill’s landmark study (2011), priming from the CNN
moderator (Moore 2008). Furthermore, dials provide poor measures of participant engagement. Participants are often repeatedly reminded to respond, and a dial can simply be maintained at a non-midpoint position. Dials can thus differentiate between degrees of favorability and unfavorability but cannot tell us reliably when a cue has engaged citizens enough to evoke a response.

Collecting Debate Viewer Responses via Mobile App

Our app brings together traditional survey methodology with the moment-by-moment data characteristic of dial-test methods, but it runs on mobile devices, making it possible to utilize a much larger participant pool. Access is via the mobile device’s browser. Thus, no “app store” download is required, and it can be used from any smartphone, tablet, or computer.

As figure 1 illustrates, four reactions are available: Agree, Disagree, Spin, and Dodge (we consider only the first two here, leaving Spin and Dodge reactions for later analysis). To register a reaction, the user taps (or clicks) the person to whom they are reacting, followed by a reaction button. All reactions therefore include both a target (Moderator, Obama, or Romney, order randomized by participant) and a reaction type (Agree, Disagree, Spin, or Dodge), making clear precisely how and to whom a debate viewer is reacting. Viewers’ ability to react on their own initiative allows us to track not only participants’ affect but also when they have passed a minimal threshold of effort to take action—even action as small as a click. If a candidate can get a viewer to click—analogous to other forms of minimal political engagement (Shulman 2009; White 2010)—it may represent the first rung in a “ladder of engagement” (Karpf 2010, 16) leading to more substantive mobilization.

This mobile-app methodology allows us to collect data from a large and diverse group of debate viewers reacting in their natural environments outside the lab (e.g., in their own homes or at debate-viewing parties). Responses are viewer initiated and virtually instantaneous, thereby allowing us to capture and analyze unmediated viewer reactions as opposed to digested opinions (Brubaker and Hanson 2009; Fridkin et al. 2007; Tsfati 2003).

Data and Methods

We invited political science instructors (through professional listservs) to offer their classes extra credit to watch the debate while using our app. In return, we

---

1. Soledad O’Brien’s instructions included: “A couple of things that we’re interested to see is how is negativity going to play? Because we know from research, negative go negative [sic], dial testers tend to turn down the dial. They hate it.” This prompt likely influenced focus-group responses: During the vice presidential debate, responses ranged only from neutral (50) to very positive (100), with no negative reactions (Moore 2008).
provided instructors with debate-related teaching resources, including lecture slides and slides summarizing our initial findings the morning following the debate (Boydston et al. forthcoming). The resulting 3,340-participant sample was more comparable to national population means in terms of gender, income, race, party identification, and religion than we would find in any single-campus study (see Boydston et al. [forthcoming] for details). The major demographic difference, of course, was in age. Participants used the app to complete a pre-debate survey, including standard demographic and attitudinal questions and questions about issue priorities, before accessing the main screen that allowed them to react to the debate through the Agree, Disagree, Spin, and Dodge clicks. In results presented below, we focus our discussion on net positive engagement, a measure of the average number of Agree clicks.
minus Disagree clicks per viewer targeted at a given candidate in the five seconds following that candidate’s discussion of a given topic or frame.\(^2\)

To identify candidate messages, we performed a content analysis of the debate transcript.\(^3\) We divided the transcript into quasi sentences (i.e., separate clauses; see Boydstun, Glazier, and Pieryka [2013]), which were manually time-stamped. Each quasi-sentence was coded for the candidate speaking, the primary topic (using the Policy Agendas Topics codebook),\(^4\) and the primary frame (moral, constitutional/legal, economic, safety, bureaucratic/logistical, political, effectiveness, patriotism, and not codable; see Boydstun, Glazier, and Phillips [2013]).\(^5\)

Results

Combining our time-stamped, coded transcript data and our data set of participants’ real-time reactions and individual-level variables enables us to investigate questions that have previously been addressed only through educated guesswork.\(^6\) We illustrate the feasibility and benefits of our approach here by examining a research area that has long interested scholars of political communication: the development and control of political agendas through agenda building and frame building. Agenda building (or agenda setting) is the process by which policy problems become topics of political discussion (Erbring, Goldenberg, and Miller 1980; McCombs and Shaw 1972). With finite agenda space, topics get attention at the necessary expense of other topics. Likewise, political actors use frame building (or issue framing) to emphasize one aspect of a topic over competing aspects (Chong and Druckman 2007; Nelson, Clawson, and Oxley 1997). Through agenda and frame building, candidates can define “what politics is about” (Schattschneider 1960), a powerful tool

2. For our analyses of the effects of candidate cues on viewer engagement, we include any five-second rolling window in which the candidate discussed the topic. The unit of analysis is the participant-second. Since our study had 3,340 participants and the debate lasted 5,443 seconds, our data set contains a total of 18,179,620 observations (3,340 participants × 5,443 seconds), noting that absence of a reaction is also an observation in our statistical analysis. In order to prevent participants who logged into the app late and/or left early from biasing downward our standard errors, we drop 5,205,421 participant-second observations where no clicks had yet registered or where no additional clicks would be registered for that participant, leaving us with just under thirteen million observations.

3. Online appendix 1 contains our complete codebook. Online appendix 2 shows topic and frame summary statistics by candidate.


5. Inter coder reliability was strong. Based on a randomly sampled 75 quasi-sentences, coders registered 94.6 percent agreement in topic codes (Cohen’s kappa = 0.924) and 85.1 percent agreement in frame codes (Cohen’s kappa = 0.764).

6. See online appendix 3 for details on how we validated our method of synchronizing response times with the transcript time stamps.
for building coalitions and gaining votes (e.g., Jones and Baumgartner 2005; Baumgartner and Jones 2009; Kingdon 1995; McCombs and Shaw 1972).

Prior agenda-building and frame-building research emphasizes a critical three-part question, what Iyengar and Valentino (2000) call “the classic shorthand of message learning theory—who says what to whom?” (110). This mantra reminds us that we must attend to the entirety of a candidate’s message: messenger, message (e.g., which topic is being discussed? which frame is being used?), and audience. Within the debate literature, however, data limitations have precluded answering questions about how message sources and specific message cues influence viewers generally, or how responses might differ across viewers. Our methodological approach allows us to illustrate how variation in each element—messenger, message, and audience—contributed to agenda- and frame-building effects in the first presidential debate of 2012. Below, we interweave findings relevant to all three of these elements through our discussion of agenda building, frame building, and audience priority.

AGENDA BUILDING

We focus here on the central discussions of the economy, health care, and foreign affairs from the first 2012 general debate. We find that some messages were uniformly more resonant with viewers than others, even given variation in messenger and audience. Figure 2 displays net positive engagement (Agree clicks minus Disagree clicks) with each candidate by response topic. This figure shows that both candidates fared best among their base supporters and independents when discussing foreign affairs, although discussing foreign affairs also yielded the worst net results for Obama among Republicans.

One prescriptive interpretation of these results could be that to maximize net positive engagement with independents and their respective bases, both candidates should have emphasized foreign affairs. Yet, Obama’s discussion of foreign affairs may have worked in Romney’s favor, as foreign affairs was the only topic where Romney surpassed Obama in terms of net positive engagement among independents. Thus, from a heresthetics perspective (Riker 1996), Romney was advantaged by shifting the agenda toward foreign affairs, whereas Obama held the relative advantage on economic, health, and other topics. Prior data could not differentiate between audience responses

---

7. Our economy and foreign-affairs categories each contain three Policy Agendas topics. From a citizen’s point of view, macroeconomics, labor (jobs), and banking discussions are all central to the most pressing question of the first 2012 general debate: the economy. Likewise, discussions of defense, foreign trade, and international affairs all shift viewers’ focus from domestic to foreign affairs.

8. According to Welch modified two-sample t-tests, Romney’s advantage over Obama on foreign affairs is not statistically significant at $p < .05$ (two-tailed) ($t = -0.8$, df = 3,978), but Obama’s advantage over Romney on the other topics is statistically significant, as is the difference between Obama’s advantage over Romney on each of these topics minus Romney’s advantage over Obama on foreign affairs.
to, for instance, foreign-affairs versus health-care messages. Our data, however, can show such fine distinctions, and our findings here reveal a tension between a candidate’s pursuit of absolute net positive engagement and his desire to keep the agenda away from topics where the opponent has a relative advantage.

FRAME BUILDING

Figure 2 also presents viewers’ net positive engagement in response to each candidate’s use of different frames. These data reveal that Democrats and independents responded most favorably to Obama’s messages when he used safety and political frames, and all viewer groups registered some of the
lowest net positive engagement when he used patriotism and constitutional/legal (henceforth legal) frames. (See online appendix 4 for moment-by-moment illustrations from the debate.) Conversely, each partisan group reacted more positively to Romney when he used patriotism frames—and Republicans still more when he used legal frames—relative to his use of other frames. These findings underscore the importance of the messenger: The least resonant frames for Obama were actually most resonant for Romney.9

The data also illustrate the importance of viewer party identification, most clearly through reactions to legal frames. Again, net positive engagement indicates that Republicans responded particularly well to Romney’s use of legal frames. In contrast, for both candidates, legal frames were among the least effective for Democrats and independents. Thus, while some frames—coming from particular candidates—resonate across viewers of all political stripes, responses to other frames are conditioned by partisanship.

AUDIENCE PRIORITY

Our fine-grained data also allow us to examine how audience characteristics beyond partisanship influence reactions. For example, viewers may respond differently to economic messages based on how strongly they prioritize the economy (Iyengar et al. 2008; Holbrook et al. 2005). Examining only candidate statements in response to moderator questions about the economy, we model viewers’ responses as a function of candidate agenda building and frame building. The results are presented in table 1.10

Each model in table 1 is a pooled cross-sectional time-series logit, in which the unit of analysis is participant-second. The response variables (Agree, Disagree) equal 1 if a participant registered that response over the previous five-second span, 0 otherwise. Our first key explanatory variable is the priority the viewer attached to the economy. The pre-debate survey asked participants to prioritize the economy using a continuous slider ranging from “Not Important” to “Very Important,” mapped to a value between 0 and 1. We also interact this viewer economic priority value with the count of seconds that the focal candidate discussed an economic topic (agenda-building models) or used an economic frame (frame-building models) in the preceding five-second

9. According to Welch modified two-sample t-tests, Romney’s advantage over Obama in terms of agreement among independents in response to patriotism frames is statistically significant (t = 7.5, df = 32,482).
10. Focusing on economic questions in this way allows us to hold constant the content of the moderator’s prompt and, thus, to better identify how viewers react to candidates’ discussion of economic topics and frames, relative to their use of other topic and frame responses that are potentially less relevant to the question.
As a candidate spends more time on economic topics/frames, and therefore less time discussing others, these count variables increase. Thus, the interaction term tests whether viewers’ reactions to economy-oriented messages were conditioned by the viewers’ own economic prioritization.

Table 1. Viewer Priority Conditions Reactions. Cell Entries Are Model Estimates (standard errors in parentheses).

**Part A. Agenda-Building Models.** Pooled cross-sectional time-series logistic regressions of audience reactions on candidate agenda-building behaviors and audience characteristics.

<table>
<thead>
<tr>
<th>Party ID</th>
<th>Obama Agree</th>
<th>Obama Disagree</th>
<th>Romney Agree</th>
<th>Romney Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Democrat</td>
<td>0.72 (0.077)</td>
<td>−2.03 (0.176)</td>
<td>−1.12 (0.092)</td>
<td>1.45 (0.107)</td>
</tr>
<tr>
<td>Republican</td>
<td>−1.35 (0.090)</td>
<td>1.73 (0.164)</td>
<td>1.00 (0.099)</td>
<td>−2.10 (0.138)</td>
</tr>
<tr>
<td>Economics priority</td>
<td>0.00 (0.002)</td>
<td>0.01 (0.004)</td>
<td>0.00 (0.002)</td>
<td>−0.01 (0.002)</td>
</tr>
<tr>
<td>Economics topic</td>
<td>−0.05 (0.012)</td>
<td>−0.01 (0.039)</td>
<td>−0.05 (0.014)</td>
<td>−0.13 (0.013)</td>
</tr>
<tr>
<td>Economics priority × Economics topic</td>
<td>0.04 (0.014)</td>
<td>0.04 (0.042)</td>
<td>0.05 (0.015)</td>
<td>0.09 (0.015)</td>
</tr>
<tr>
<td>Intercept</td>
<td>−4.07 (0.174)</td>
<td>−8.77 (0.417)</td>
<td>−4.99 (0.206)</td>
<td>−4.99 (0.232)</td>
</tr>
<tr>
<td>Var(intercept)</td>
<td>0.78 (0.035)</td>
<td>1.86 (0.058)</td>
<td>1.08 (0.037)</td>
<td>1.27 (0.042)</td>
</tr>
<tr>
<td>σ_u</td>
<td>1.48 (0.026)</td>
<td>2.54 (0.074)</td>
<td>1.71 (0.032)</td>
<td>1.89 (0.039)</td>
</tr>
<tr>
<td>ρ</td>
<td>0.40 (0.008)</td>
<td>0.66 (0.013)</td>
<td>0.47 (0.009)</td>
<td>0.52 (0.010)</td>
</tr>
<tr>
<td>AIC</td>
<td>450,125</td>
<td>102,489.8</td>
<td>365,746.8</td>
<td>252,560.3</td>
</tr>
</tbody>
</table>

(Continued)

11. Each model was restricted to moments following statements by the focal candidate. For example, the Romney models focus only on Romney-targeted clicks after he made a statement and ignore the relatively few clicks focused on Romney after Obama has made a statement. The models control for viewer party identification.
12. Note that participants’ conceptions of “the economy” may not match the Policy Agendas codebook. Attributing issue priorities to citizens based on restricted question wording is a serious
problem (Wlezien 2005) that ideally should be verified through cross analysis of multiple open-ended survey items (Jennings and Wlezien 2011). Unfortunately, having only asked participants to prioritize a few topics, we cannot verify that participants’ perceptions of the economy match our categorization.

Our grouping of macroeconomics, labor (jobs), and banking into a single economic category helps address this concern, as citizens’ economic evaluations often depend on employment considerations (Haller and Norpoth 1997; Niemi, Bremer, and Heel 1999) and media reports (Hetherington 1996), which in 2012 tended to emphasize the banking sector’s role in shaping the economy. Regardless, statistically significant effects of participants’ self-reported economic priority on what we categorize as economic cues point to issue priority as a conditioning factor.

Table 1. Continued


<table>
<thead>
<tr>
<th></th>
<th>Obama Agree</th>
<th>Obama Disagree</th>
<th>Romney Agree</th>
<th>Romney Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n = 1,739,564</td>
<td>n = 1,739,564</td>
<td>n = 1,815,663</td>
<td>n = 1,815,663</td>
</tr>
<tr>
<td>Party ID</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Independent</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Democrat</td>
<td>0.72 (0.077)</td>
<td>–2.02 (0.175)</td>
<td>–1.12 (0.092)</td>
<td>1.45 (0.107)</td>
</tr>
<tr>
<td>Republican</td>
<td>–1.35 (0.090)</td>
<td>1.73 (0.163)</td>
<td>1.00 (0.099)</td>
<td>–2.10 (0.138)</td>
</tr>
<tr>
<td>Economics priority</td>
<td>0.00 (0.002)</td>
<td>0.01 (0.004)</td>
<td>0.01 (0.002)</td>
<td>–0.01 (0.002)</td>
</tr>
<tr>
<td>Economics frame</td>
<td>–0.04 (0.010)</td>
<td>–0.04 (0.031)</td>
<td>–0.06 (0.014)</td>
<td>–0.05 (0.014)</td>
</tr>
<tr>
<td>Economics priority ×</td>
<td>0.06 (0.011)</td>
<td>0.03 (0.034)</td>
<td>0.03 (0.015)</td>
<td>–0.01 (0.016)</td>
</tr>
<tr>
<td>Economics frame</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>–4.20 (0.169)</td>
<td>–8.71 (0.389)</td>
<td>–5.05 (0.201)</td>
<td>–5.34 (0.229)</td>
</tr>
<tr>
<td>Var(intercept)</td>
<td>0.79 (0.035)</td>
<td>1.86 (0.058)</td>
<td>1.08 (0.037)</td>
<td>1.27 (0.042)</td>
</tr>
<tr>
<td>σ_u</td>
<td>1.48 (0.026)</td>
<td>2.54 (0.074)</td>
<td>1.71 (0.032)</td>
<td>1.89 (0.039)</td>
</tr>
<tr>
<td>ρ</td>
<td>0.40 (0.008)</td>
<td>0.66 (0.013)</td>
<td>0.47 (0.009)</td>
<td>0.52 (0.010)</td>
</tr>
<tr>
<td>AIC</td>
<td>450,069.5</td>
<td>102,499.9</td>
<td>365,550.8</td>
<td>252,656.6</td>
</tr>
</tbody>
</table>

Note.—The unit of analysis is participant-second. The response variables equal 1 if a participant registered the corresponding response over the previous five-second span and 0 otherwise. Models were run only for candidate responses to economic questions. Democratic (Republican) participants include identifiers and independents leaning Democratic (Republican).
Demonstrating the importance of individuals’ issue priorities, the agreement models show positive, statistically significant coefficients associated with the interaction between viewers’ economic priority and the candidate’s discussion involving economic topics and frames: People’s tendency to click “Agree” in response to economic discussion increased with their economic prioritization. In contrast, in three of four disagreement models, the coefficient associated with the interaction is small and statistically indistinguishable from 0, suggesting that well-executed agenda- and frame-building cues may effectively draw issue publics (Converse 1964) into the debate, producing agreement without necessitating disagreement. In sum, our analysis illustrates our methodology’s potential to yield detailed insight into specific audience reactions, such as how viewers’ economic prioritization conditions receptiveness to economic discussion.

Conclusion
A novel mobile app enabled us to collect real-time data from a large, diverse population reacting at their own initiative in a natural environment. This approach overcame many limitations of prior large-N debate studies and small-N dial testing, allowing us to investigate effects of specific candidate cues on viewer engagement—and, thus, to ask questions previous data have lacked the granularity to answer. Our brief results illustrate the potential of a mobile-app research design for tracking political behavior in real time. Naturally, much work remains in order to develop an understanding of how—and why—specific cues prompt positive and negative reactions from citizens. Applied across multiple presidential debates and election years, the app-based approach could be used to test specific theoretically derived hypotheses, thereby advancing our understanding of debate effects and their underlying mechanisms. The broader promise of this methodology is greater still. Mobile apps could be used to study a host of public-opinion phenomena, from tracking response latency, to measuring real-time reactions to a major policy speech or media coverage of an unfolding crisis, to deploying experimental studies across geographically diverse populations.

Supplementary Data
Supplementary data are freely available online at http://poq.oxfordjournals.org/.

13. On the other hand, a significant interaction in Romney’s agenda-building disagreement model shows that the greater the priority viewers placed on the economy, the more likely they were to disagree with Romney’s responses about the economy. This finding may suggest that viewers attuned to the economy were more likely to react negatively to Romney’s comments in the context of the mixed economic climate (Vavreck 2009) or his personal wealth (Adams 2012).
References


