Supplementary Materials for

“Analyzing Attention to Scandal on Twitter: Elites Sell what Supporters Buy”

Appendix 1: Representativeness of Twitter attention to scandal among media and elites

The relevance of agenda-setting analyses using Twitter data depends on the extent to which those data are representative of attention in other contexts as well. In the Barberá et al. (2019) supplementary materials, an analysis is conducted to this end focusing on the attention of the public. It compares the topics of tweets from different samples intended to represent the US public with results from the Gallup Most Important Problem question, and finds an intermediate correlation between the two measures of public attention.

But this analysis also incorporates data on elite and the media, opening the question of whether Twitter is representative of their attention also. To test this proposition, I turn to two offline measures of elite and media attention: Congressional floor speeches and Associated Press reports.

Elite attention to scandal on Twitter and on the House/Senate floor

Beginning with my analysis of Congressional floor speeches, I use Congress.gov to search the Congressional Record for entries that contain keywords that hold relevance to each of the scandals. These keywords are “Benghazi,” “IRS,” “NSA,” and “Veteran’s Health Administration.” Entries for speeches that occurred more than two weeks before initial media attention to the scandal are removed, since these are unlikely to refer to the controversies under study, and weekly counts of entries containing each keyword are then collected.

This measure of offline attention contains a few sources of slippage that necessarily make it somewhat noisy. For instance, three of the keywords refer to government agencies that may be referenced not in relation to the scandal (although a quick review of the results indicates that most are). In addition, the measure reflects speeches delivered by all Members of Congress, while the data from Twitter are limited to Republican members.

To compare offline and Twitter attention, the weekly counts of Congressional Record entries with scandal-related keywords are then merged with data representing the mean daily percentage of Tweets devoted to that scandal by Republican MCs in each week. I use these data in two ways. First, I calculate the pairwise correlations between weekly Twitter and offline attention, including only observations from two weeks before each scandal broke and after. Second, to further typify Twitter attention in relation to an offline measure, I conduct a Granger causality test to see "which comes first," attention to scandal online or off? (Lag lengths for the Granger tests are calculated in line with the process described in Appendix 3.)

Table A1 presents the results of these exercises. Each row refers to online and offline attention to one of the four scandals. The first column presents the Pearson's *r* correlation and the 95% confidence interval, while the second column indicates the number of weekly observations analyzed. The third and fourth columns indicate, in simple "yes/no" fashion, whether previous weeks' attention to the scandal on Twitter (on the House or Senate floor) is significantly predictive of attention on the floor (on Twitter) when controlled for by its own lagged values. When significant results are found in one direction but not the other, this indicates that one series is "leading" and the other "following," at least in the short-run.

Table A1. Relationship between elite attention to scandal on Twitter and in Congress

	Correlation between Twitter and floor speech attention		Granger causality tests	
	Pearson's *r*	N	Twitter → Floor	Floor → Twitter
Benghazi	0.52* [0.37, 0.65]	104	Yes	No
IRS Targeting	0.63* [0.48, 0.74]	86	No	Yes
Snowden Leaks	0.43* [0.23, 0.59]	83	No	No
VA Delays	0.34* [0.02, 0.60]	35	No	Yes

Note: * $p < 0.05$

The results indicate an intermediate relationship between attention to scandal on Twitter on the House and Senate floor. Across the four scandals, the correlation between measures has a mean of $r = 0.48$ with a maximum of $r = 0.63$ in the IRS targeting scandal, and a minimum of $r = 0.34$ in the VA delays controversy (although importantly, the latter had relatively few timepoints because the story broke in mid-2014). In terms of ascribing precedence, Twitter attention appears to have preceded offline attention for the Benghazi investigation, but the reverse was true for the IRS targeting and VA delays controversies. From this limited analysis, then, we do not observe that attention in one venue consistently precedes that in another.

Media attention to scandal on Twitter and in other news reporting

Barberá et al. (2019) also construct a measure of media attention to topics on Twitter, drawing from the 36 largest media outlets according to the Pew Research Center, including both mainstream and partisan sources. To assess whether this measure is representative of the news that Americans consume more broadly, I compare the Twitter-based measures of media attention to each scandal to the number of Associated Press reports mentioning keywords relating the scandals. I focus on these reports because, for many outlets, content from the AP wire constitutes a large share of the content they relay to consumers, and outlets have shown to be responsive to the norms and frames that the AP uses in its reporting (Diakopoulos 2019, Djourelova 2020).

To collect AP reports on each of the four scandals, I use NexisUni to conduct a keyword search for reports in the years 2013 and 2014. Because NexisUni allows no more than 1,000 records to be downloaded per query, I had to specify these keywords more narrowly than in the Congressional Record exercise. For the Benghazi investigation, I identify reports containing the terms "Benghazi," "attack," and "Stevens" (the latter being the name of the US ambassador killed in the attack). For the others, I search the terms "VA hospital," "IRS" and "targeting," and "NSA" and "Snowden."

As in the previous exercise, I then create a weekly count of reports containing the terms relating to each scandal or controversy. These counts are compared to the mean daily percentage of tweets devoted the scandal on Twitter for each week. These data are analyzed following the same procedure as the Congressional Record exercise. Table A2 presents the results, with the correlations between weekly Twitter and AP attention to each scandal in the second column, and indicators of Granger causality between them in the third and fourth.

Across the four scandals, there is again an intermediate relationship between attention from outlets on Twitter and in AP wire reports. The mean Pearson's r is a moderate-to-high 0.63, with two scandals showing a very tight relationship (IRS targeting at $r = 0.90$ and VA delays at $r = 0.75$), while Twitter and AP attention to Benghazi had a weak-to-moderate relationship at $r = 0.34$. In terms of ascribing precedence, again we see that attention to Twitter both leads and follows attention in AP reports depending on the scandal, with neither source being a clear leader of the other.

Table A2. Relationship between media attention to scandal on Twitter and in AP reports

	Correlation between Twitter and floor speech attention		Granger causality tests	
	Pearson's *r*	N	Twitter → AP	AP → Twitter
Benghazi	0.34* [0.16, 0.50]	104	Yes	No
IRS Targeting	0.90* [0.86, 0.94]	86	Yes	No
Snowden Leaks	0.53* [0.35, 0.67]	83	Yes	Yes
VA Delays	0.75* [0.56, 0.86]	35	No	Yes

Note: * $p < 0.05$

Appendix 2: Unit root testing for six time series

In this study, I analyzed six time series under assumptions of stationarity. The data in these time series reflected the number of Twitter mentions made by a particular group—the public, media, or elites—in regard to one of four Obama Administration scandals.

To establish the stationarity of these series, I conducted Phillips-Perron Z_τ tests on each of them. The null hypothesis of these tests is that the series contains a unit root, and would therefore not be stationary. To reject this null hypothesis, I must establish that the Z_τ statistics for each series is greater than the critical value at a preestablished level of significance (here, $p < 0.05$) at a comparable or smaller value of T.

The results of my test are presented in Table A3. In all cases, I reject the null hypothesis of a unit root, and treat the series as stationary.

Table A3. Phillips-Perron Z_τ statistics for each series

Topic	Actor	Series Z_τ statistic
Benghazi Investigation	GOP Members of Congress	-18.10
	Public GOP	-16.11
	Mainstream Media	-23.02
IRS Targeting Scandal	GOP Members of Congress	-13.67
	Public GOP	-12.44
	Mainstream Media	-19.78
Snowden Leaks	GOP Members of Congress	-14.42
	Public GOP	-16.21
	Mainstream Media	-17.67
VA Delays Scandal	GOP Members of Congress	-16.29
	Public GOP	-14.92
	Mainstream Media	-22.98

Note: Critical value at $p = 0.05$ is $Z_\tau = -2.87$. Values below this threshold reject null of a unit root.

Appendix 3: Selection of lag lengths

To estimate a vector autoregression (VAR) model, an analyst must specify the number of lags that will be used to explain contemporaneous values of each series. In simple terms, this is a matter of deciding "how far back?" should the model look in determining whether past values explain current values.

To select lag lengths, I run all possible models up to maximum possible lag length. I set this maximum at 14 days, or two weeks, following the lead of Barberá et al. (2019). Then, I use four selection criteria to determine which lag length achieves the strongest balance of parsimony and information loss. The recommendations of these criteria are presented in Table A4. For both VARs, one for each scandal, I take the more conservative of the plurality recommendations (Benghazi and Snowden = 7, IRS and VA = 8) as my lag lengths.

Table A4. Lag length recommendations given by four model selection criteria

	Akaike Information	Hannan-Quinn	Schwarz	Amemiya Prediction
Benghazi VAR	7	3	1	7
IRS VAR	8	3	1	8
Snowden VAR	7	3	3	7
VA Delays VAR	8	3	3	8

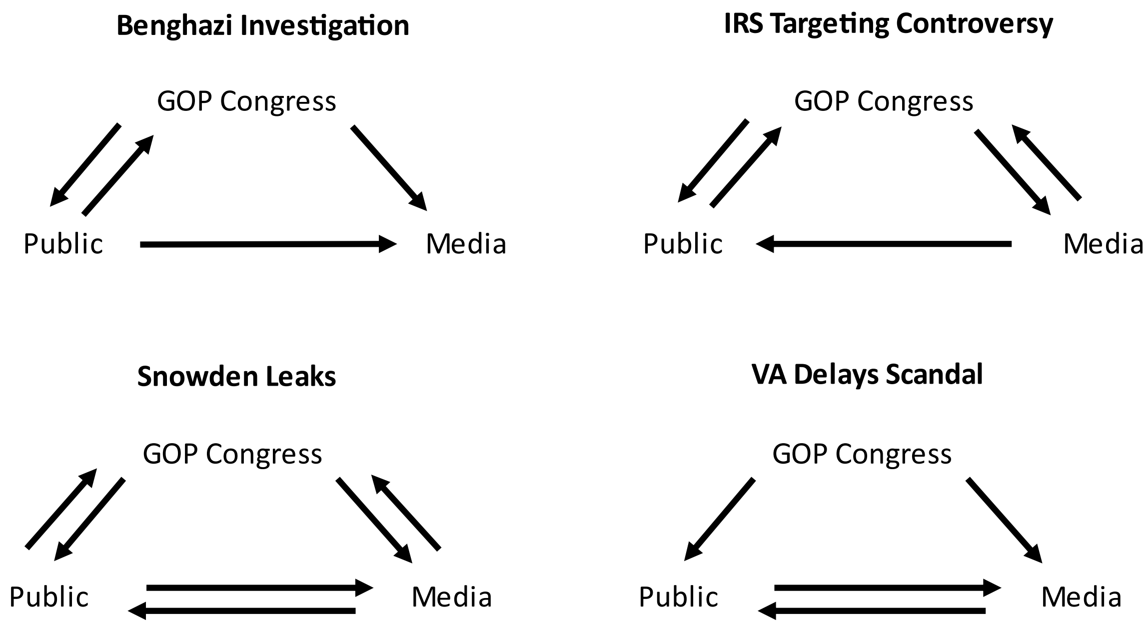
Appendix 4: Reproduction of results with alternative sample

In my analysis for this paper, I used citizen supporters of the GOP to stand in for what theory has referred to generally as "the public." My rationale for doing so was that, to the extent the public leads or follows other actors, such an effect would be seen most clearly among supporters of the political opposition.

Barberá et al. (2019) also track tweet volume among the "attentive public," a random sample of users who follow one or more major news outlet on Twitter. When I reproduce my findings using this sample instead,[1] I find that consistently with expectations, the results show that in the long-run (Table A5), the influence of elites over the public's attention is stronger than any other mechanism studied. The only major difference in results between samples is in the short-run impacts identified by Granger causality testing (Figure A1). In contrast to my findings with the partisan public samples, the relationship between elite and "attentive public" attention to Benghazi and the IRS scandal goes both ways. For the other two scandals, only elites impact the "attentive public" in the short-run, and not vice-versa. This suggests, consistent with Barberá et al.'s original findings, that elites are more responsive to the tweets of their supporters than they are to those of the public at large.

[1] As I did in the main analysis, I test the stationarity of the "attentive public" time series and estimate the VARs to have independently distributed residuals.

Figure A1.
Granger causality of tweet volume in reference to four Obama Administration scandals



Note: Arrows indicate Granger causality at $p < 0.05$.

Table A5. Cumulative responses to a 10-percentage point impulse in Twitter mentions				
Mechanism	**Benghazi**	**IRS Targeting**	**Snowden Leaks**	**VA Delays**
Public → Elites	4.1 [2.8, 5.2]	4.2 [1.0, 5.6]	2.7 [0.6, 4.3]	3.8 [2.0, 4.9]
Public → Media	3.4 [1.7, 4.4]	3.6 [1.4, 4.6]	4.5 [2.6, 5.6]	3.6 [2.1, 4.5]
Media → Public	3.0 [0.7, 3.9]	5.4 [4.0, 6.1]	5.2 [3.8, 5.8]	4.5 [3.2, 5.1]
Media → Elites	3.6 [1.7, 4.8]	5.4 [3.7, 6.1]	4.2 [1.4, 5.4]	*2.4* *[-0.5, 4.0]*
Elites → Media	4.1 [2.5, 4.8]	4.7 [3.0, 5.7]	4.9 [3.1, 5.9]	3.9 [2.4, 4.8]
Elites → Public	4.3 [2.5, 5.2]	9.5 [6.1, 11.2]	9.2 [6.7, 10.9]	3.8 [2.0, 4.7]

Notes: Cell entries indicate cumulative increase in percentage of daily tweets about a scandal by the response group over a 15-day period. 95-percent confidence intervals are in brackets. Upper and lower bounds are not equidistant from estimate because values were reverse transformed from logits. All estimates significant at $p < 0.05$ except media → elites in the VA Delays scandal.