Second screen effects on perceived candidate performance during a televised candidate debate

Dominic Nyhuis* Thorsten Faas†

Version: December 2017

Abstract

The second screen increasingly dominates how individuals consume television. While several studies have investigated the content of concurrent online discussions, it is much more challenging to capture effects of accompanying viewership on perceptions. This study makes an attempt at filling this gap. We analyze data from a panel study that was collected on the occasion of a televised candidate debate during the 2013 German federal election. The debate is arguably the most important event on the campaign trail. It is thus crucial to understand the factors that drive perceptions of candidate performance. We consider whether and how performance evaluations on the second screen shape perceptions of concurrent social media users. The results indicate a strong effect on candidate performance evaluations.

Keywords: *Second screen; Candidate debates; Candidate evaluations; Twitter; Germany*

*Goethe-Universität Frankfurt
†Freie Universität Berlin

1 Introduction

The second screen is an increasingly prominent phenomenon in the way that individuals consume television. The prevalence of smartphones and tablet computers allows viewers to join online communities and exchange views on programs. An interactive viewership is at times even explicitly encouraged when dedicated platforms are created or social media accounts are set up to frame shows. Academic analysts have been quick to consider the ramifications of this phenomenon for the viewer experience (Gottfried et al., 2017; Han and Lee, 2014; Mukherjee and Jansen, 2014; Shim et al., 2015; van Cauwenberge, Schaap and van Roy, 2014). While a number of observers have investigated the content of concurrent online discussions dedicated to particular programs (Doughty, Rowland and Lawson, 2011; Kalsnes, Krumsvik and Storsul, 2014; Shamma, Kennedy and Churchill, 2009; Trilling, 2015), it is much more challenging to capture the effects of accompanying viewership.

The present study makes an attempt at filling this gap. In an analysis of the German televised candidate debate, we consider to what extent performance evaluations on the second screen shape performance evaluations of concurrent social media users. This is to say that we are interested in the effects of passive usage – users as consumers of opinion. This focus sets the present research effort apart from several previous studies which have considered the effects of active usage on perceptions – users as producers of opinion (Houston, Hawthorne, Spialek, Greenwood and McKinney, 2013; Houston, McKinney, Hawthorne and Spialek, 2013).

Televised candidate debates are the most prominent events during contemporary electoral campaigns (McKinney and Carlin, 2004; Reinemann and Wilke, 2007). They have been the subject of a broad scientific literature (Faas and Maier, 2011*a*,*b*; Maier and Faas, 2011; Maier, Faas and Maier, 2013; Maurer and Reinemann, 2007), which has focused extensively on perceptions of candidate performance and the effects of debate performance on electoral outcomes (Maier et al., 2006; Maier and Faas, 2005; Maier, Faas and Maier, 2014; Maier, 2007*b*). Given the elevated status of debates in modern electoral campaigns and given their potentially sizable effects on electoral choices, it is crucial to understand which factors shape these perceptions. To assess the effects of opinions on the second screen on perceived candidate performance evaluations, we rely on evidence from a two-wave panel survey with a sample of German Twitter users. In addition to the survey evidence we reconstruct survey participants' home feeds to assess the general tendency of candidate evaluations by users' Twitter friends. Applying supervised text classification to the content of the home feeds allows us to link candidate performance perceptions with user-specific performance evaluations on the second screen. The results indicate that the Twitter content exhibits a strong effect on performance evaluations.

2 Candidate perceptions and the second screen

Media effects are among the most fundamental research concerns of communication science (Bryant and Oliver, 2009; Scheufele, 2000; Scheufele and Tewksbury, 2007; Sparks, 2006), particularly with regard to the evaluation of political actors (Banducci and Karp, 2003; Miller and Krosnick, 1996; Pan and Kosicki, 1997). Not least in light of an extensive body of research that has provided evidence for biases in media coverage (Chiang and Knight, 2011; D'Alessio and Allen, 2000; Gentzkow and Shapiro, 2010) and their effects on vote choices (DellaVigna and Kaplan, 2007; Druckman and Parkin, 2005; Kahn and Kenney, 2002) is it crucial to assess how media shape perceptions of and preferences for political actors.

Arguably the single most important events in contemporary election campaigns are televised candidate debates. These formats, brought about by the imperatives of a heavily mediated campaign landscape (Bennett and Entman, 2001; Corner and Pels, 2003; Strömbäck, 2008), tend to accentuate individual political actors – party leaders and top candidates – who come to represent parties in their campaign efforts. In short, televised debates are the most prominent expression of a personalized campaign style (Holtz-Bacha, Langer and Merkle, 2014; Karvonen, 2010; van Zoonen and Holtz-Bacha, 2000).

First and foremost, televised debates command a substantial amount of media attention as they reach a sizable share of the electorate, such that a bad performance could seriously harm the electoral prospects of the perceived loser. Moreover, by virtue of their accessibility and dramatic style, televised candidate debates have the potential to speak to an audience that is somewhat remote from politics. The promise of shaping preferences of a politically less sophisticated and, by extension, less politically committed viewership highlights the perceived make-or-break status of televised debates. Finally, candidate debates structure the salience of the campaign issues well beyond the specific broadcast (Benoit, Hansen and Verser, 2003; Swanson and Swanson, 1978). To be sure, televised debates influence the national discourse precisely because of the attention that is devoted to them, such that effects and attention are mutually reinforcing.

Given the elevated status of televised candidate debates, it is not surprising that scholars have been highly attentive to potential effects of the debates, for example on issue knowledge (Bishop, 1978; Chaffee, 1978; Gottfried et al., 2014; Holbrook, 1999), candidate perception (Benoit and Hansen, 2004; Benoit, Hansen and Verser, 2003; Maier, 2007*a*) and eventual vote choice (Geer, 1988; Maurer and Reinemann, 2007). However, although numerous studies have commendably explored the many ways in which televised candidate debates affect different elements of the voter perception, the conventional research set-up considers an immediate effect of content on political perceptions. Increasingly, this model of the viewer experience diverges from the reality of television consumption.

The growing pervasiveness of smartphones and tablet computers has brought about the phenomenon of the second screen where viewers simultaneously engage with the television and another device (van Cauwenberge, Schaap and van Roy, 2014). While the second screen label might refer to interactions with any type of content, the present contribution is interested in the second screen phenomenon where viewers access content that is linked to the content of the main screen. More specifically then, we could conceive of this phenomenon as social television where viewers join some form of online community to either discuss the broadcast or be exposed to evaluations thereof. Turning to the specific subject of media effects, the present contribution asks how a thus conceived social television experience shapes perceptions of the program's content. More to the point and regarding the televised candidate debate, we aim to assess whether second screen content determines candidate performance evaluations on the main screen.

Although the phenomenon of the second screen is reasonably recent, there is some scholarly research on second screen usage during US presidential debates. Several studies have investigated the content of online discussions surrounding the debates (Freelon and

Karpf, 2015; Hawthorne, Houston and McKinney, 2013; Klemm and Michel, 2014; McKinney, Houston and Hawthorne, 2014). Conversely, only few contributions have considered the effects of the second screen on candidate perceptions and performance evaluations. In two studies on Twitter usage during US presidential debates, Houston and colleagues find that debate learning and debate attention is related to more active tweeting (Houston, Hawthorne, Spialek, Greenwood and McKinney, 2013; Houston, McKinney, Hawthorne and Spialek, 2013). Moreover, using survey evidence (Gottfried et al., 2017) provide some indication that engagement with Twitter is related to issue learning. The present study differs from previous contributions on the subject by considering the passive engagement with Twitter – users as consumers of opinion – while employing an explicit measure of the content that users might have been exposed to, not a mere measure of the engagement with the platform.

The specific debate under consideration took place on September 1, 2013 between 8:30 and 10:00 pm, three weeks prior to the German federal election. It was the only such event scheduled during the 2013 electoral campaign and, consequently, a considerable amount of public attention was devoted to the debate. The debate pitted incumbent chancellor Angela Merkel (CDU) against Peer Steinbrück (SPD). It was simultaneously broadcast on two public and two private channels – ARD, ZDF, RTL, and ProSieben – and was moderated by one representative from each of the four stations. The debate had a combined audience of more than 17 million viewers. Considering an electorate of approximately 62 million eligible voters, it is clear that the debate reached a sizeable chunk of the electorate, highlighting the unique status of the debate.

3 Data and methods

To assess the research interest outlined in the previous section, a two-wave panel survey was conducted with a sample of German Twitter users. In order to uncover the effects of Twitter usage on candidate performance evaluations, we collected the content of the respondents' Twitter feeds. This section begins by outlining the panel component of the study (section 3.1). Subsequently, we elaborate the collection of tweets (section 3.2), as well as the techniques that are applied to estimate candidate performance evaluations in Twitter messages (section 3.3). The section closes with discussion of the model set-up (section 3.4).

3.1 The panel component

The first wave of the panel study was fielded on August 30, 2013. The field period ended immediately prior to the beginning of the broadcast at 8:30 pm on September 1, 2013. Users were invited to provide their e-mail address in order to be re-contacted for the panel component of the study. Survey participants were also invited to input the names of their Twitter accounts in the first wave in order to link candidate perceptions with the content of their Twitter feeds. The second wave was fielded on September 3, 2013 and lasted for two weeks until September 15. 2013. Approximately 85% of users that participated in the second wave of the panel study did so within the first two days of the field period. We restrict the sample to these respondents in order not to dilute the substantive conclusions by a substantial time lag between the debate and the survey.

The participants were recruited using a dedicated Twitter account that created awareness for the study. It is important to note that this convenience sample precludes inferences on the general Twitter public in a strict sense, as highly politically interested users are more likely to self-select into the sample. Moreover, as there is no readily available baseline for the sociodemographic composition of German Twitter users, we cannot hope to correct for the imbalances among survey participants with a weighting scheme. Yet, finding that the sample is composed of highly politically interested participants is not in and of itself evidence for the degree to which the sample is biased as the Twitter user base is not representative of the general public to begin with (Bekafigo and McBride, 2013; Vaccari et al., 2013). As such, studying Twitter is a veritable exercise in elite research and as the aim of the present study is not and cannot be a generalization beyond the confines of Twitter, a sample of well-educated and politically knowledgeable respondents might be less of a cause for alarm. Nevertheless, this limitation needs to be kept in mind when interpreting the results of the analysis.

One beneficial upshot of the recruitment strategy is that we do not require respondents to operate in a setting which might be perceived as artificial and potentially uncomfortable by respondents. Consider as an example the work by Houston and colleagues who study the effects of live-tweeting during the presidential candidate debates in 2012 (Houston, Hawthorne, Spialek, Greenwood and McKinney, 2013; Houston, McKinney, Hawthorne and Spialek, 2013). Besides having respondents watch the debate in a laboratory environment,

> "[p]articipants who tweeted during the debate [...] were informed they must have a public Twitter account already established so their tweets about the debate could be captured for separate analysis" (Houston, Hawthorne, Spialek, Greenwood and McKinney, 2013, 304).

While their setting has the merit of a more controlled research environment, the set-up of the present study captures Twitter effects in a more ordinary context, not least by studying users who are well versed in the medium and who would naturally engage with Twitter during the debate.

We were able to realize a sample of 98 participants who responded to both waves of the survey questionnaire. We discarded participants who did not provide the names of their Twitter accounts or where the accounts' privacy settings did not allow accessing the list of friends, which is necessary for reconstructing the users' home feeds. We also dropped users that responded not to have used Twitter during the debate in the post-debate survey.[1] Finally, we discarded users whose home feeds did not contain at least twenty evaluative tweets per candidate in order to not unduly influence the evaluation estimates by random noise (see section 3.3).

3.2 Reconstructing the home feeds

The first task in linking the content of tweets with candidate performance perceptions in the survey is to reconstruct users' home feeds. Home feeds are composed of all messages that were posted by friends during the debate. Importantly, Twitter differentiates between friends and followers, which is to say that in contrast to other social network sites, relationships on Twitter are not mutual by design. In their home feeds, users are shown all tweets that were

[1] Twitter usage during the televised debate was queried using the following question: "Have you, at least occasionally, used Twitter during the televised debate between Angela Merkel and Peer Steinbrück?" ["Haben Sie während des Fernsehduells zwischen Angela Merkel und Peer Steinbrück Twitter zumindest vorübergehend verfolgt und/oder genutzt?"]

published by friends, sorted by time stamp. If a tweet addresses another user – tweets starting with @username –, users are only shown the tweet if they also follow the user that is being addressed. At the time of the data collection Twitter did not algorithmically alter the home feeds. Therefore, knowing the set of users' friends is sufficient for reconstructing the home feeds that users were potentially exposed to.

After collecting the friends of the survey participants, we assembled a database of all tweets that were posted by friends' accounts during the debate.[2] This database allows reconstructing the home feeds and thus to assess the prevalence of the candidate debate in the feeds and how candidate performances were evaluated.[3]

3.3 Estimating candidate evaluations in tweets

Despite a moderately-sized user sample, the number of tweets that were featured in the feeds is quite extensive and comprises roughly 112,000 messages (approx. 28,000 unique tweets). Therefore, in order to capture the general tendency in the feeds' candidate evaluations, we rely on supervised sentiment classification. Four trained coders labeled a sample of tweets regarding the sentiment expressed toward the candidates. Coders employed a trifold classification scheme with tweets being labeled as expressing a positive, negative, or neutral sentiment toward the actor addressed in the tweet.[4] The training data consist of approximately 3,200 tweets addressing Angela Merkel (CDU) and about 2,800 tweets addressing Peer Steinbrück (SPD).

The resulting set of classified tweets is used as training data for several supervised classification algorithms which estimate the sentiment expressed toward the candidates in the entire corpus.[5] Two distinct training corpuses are created – one for Angela Merkel and one for Peer Steinbrück – as negativity might be expressed using terms that are specific to the candidates. For example, Angela Merkel was criticized in the months leading up to the

[2]The data was collected using the twitteR package in R.

[3]There are a few missings in the dataset when friends' accounts were set to private, such that tweets are only visible if users make them explicitly accessible to another user upon request. However, the number of missing accounts is fairly small, as accounts are public by default.

[4]We are grateful to Will Lowe for providing the infrastructure for manually coding the tweets.

[5]The exposition of the supervised classification needs to remain somewhat cursory. A more detailed elaboration of this aspect is provided elsewhere (Nyhuis and Faas, Forthcoming).

election for a statement in which she labeled the web as *Neuland* (*virgin soil*). In light of news regarding the large-scale US spying programs that broke around the same time, this quote was taken as evidence for an unwillingness or inability of the Merkel government to address the issue of online security – an issue that evokes particularly strong and negative sentiment among tech-savvy Twitter users (Nyhuis and Friederich, 2017). For the same reason, we discarded tweets from the training data that make reference to both Merkel and Steinbrück.

The idea of supervised classification is to employ the evidence from the labeled data (training data) to make inferences on the categorical membership of the non-labeled data (test data). More to the point, similarities in the features of the labeled data to those in the non-labeled data are used to classify the virgin elements into the given categories. In the specific case at hand, term occurrences indicate that a tweet belongs to one of the three sentiment categories.

There are numerous algorithms for performing supervised classification. For the present study, we apply *random forests* (Breiman, 2001), *support vector machines* (D'Orazio et al., 2014; Steinwart and Christmann, 2008), and *maximum entropy* (Berger, Della Pietra and Della Pietra, 1996). The sentiment label is assigned to the texts on the basis of the most frequently estimated label as ensemble classification tends to outperform individual classification algorithms (Seni and Elder, 2010).

For the practical application, we created two distinct corpuses for Merkel and Steinbrück, discarding all tweets that do not mention either or both of the candidates. The latter is done in order to ensure that the sentiment label is assigned to the correct candidate, not least since the procedure disregards syntax: A term-document matrix is created from the training and the test data where each tweet is represented in one column and each term is represented in one row. The cell entries contain the frequency with which a given term appears in a tweet.

It is likely that several tweets are mislabeled by the algorithms. However, as we have shown elsewhere, the automatically assigned labels exhibit a fair degree of overlap with manually labeled texts (Nyhuis and Faas, Forthcoming). Prior to estimating the categorical membership of the virgin tweets, we perform several common pre-processing operations.

Specifically, we remove numbers, punctuation, frequent terms (*stopwords*), and URLs, convert the tweets to lower case and reduce the terms to their stems.[6]

After the classification step, we aggregate the sentiment scores to two user- and candidate-specific evaluation measures. Specifically, we take the sum of the evaluative messages (positive = 1, neutral = 0, negative = -1) and divide it by the total number of evaluative tweets for each candidate. This operation yields a measure of the general tendency that is expressed toward the candidates, ranging from -1 (completely negative) to 1 (completely positive).[7] Section 4 provide some summary statistics on the overall and user-specific sentiment that is expressed toward the candidates.

3.4 Variables

In order to assess the effects of the candidate evaluations in the users' home feeds, we run several linear regression models – separated by candidate – with the *perceived candidate performance* as the main dependent variable. The perceived performance was collected in the post-debate wave on a five-point scale, ranging from *Very good* (1) to *Very poor* (5). The independent variable of interest is the estimated *candidate performance evaluation* in the users' home feeds. The most important control variable is the *expected candidate performance* that was collected in the pre-debate wave on a five-point scale, ranging from *Very good* (1) to *Very poor* (5). The pre-debate expectation provides a summary control for a number of factors related to partisan misperceptions (Taber, Lodge and Glathar, 2001), which should affect both pre-debate performance expectations as well as post-debate performance perceptions. In addition to this summary control, we also explicitly control for *candidate sympathy scores*, which were collected on an eleven point scale, ranging from -5 to +5. Finally, we control for

[6]Some of the content that is removed in these operations might arguably carry information on the sentiment that is expressed in the tweets. Consider as an example – *Merkel :-(* – where the punctuation does convey a sentiment. In much the same way, the supervised classification algorithms might pick up on the fact that certain hyperlinks do reference texts that are favorable or critical of either of the two candidates. This information is discarded by removing URLs. In practice, the result of the sentiment classification is not greatly affected by any of these preprocessing operations. Moreover, since we are only interested in the general sentiment tendency that is expressed in the users' home feeds, individual misclassifications do not shape the estimates by a large margin.

[7]We discard users with fewer than 20 evaluative messages per candidate in order to not unduly influence the estimates by random noise.

the *usage frequency of different media sources*.[8] Table 5 in the Appendix provides summary statistics for the independent variables.

4 Candidate evaluations in home feeds

TABLE 1: Prevalence of candidate debate in user feeds

	tvduell	merkel	steinbrück	merkel OR steinbrück
Mean	0.67	0.33	0.23	0.48
SD	0.09	0.08	0.08	0.12
Min	0.29	0.09	0.10	0.17
Max	0.86	0.48	0.49	0.72

The Table displays the prevalence of the televised debate in users' home feeds during the broadcast. Column 1 displays the percentage of tweets that contained the term tvduell, columns 2 and 3 show the percentages pertaining to the two candidates. Column 4 provides the relative frequency of tweets that contained either of the two candidates.

Before turning to the substantive analysis, this section provides some detail about the estimated candidate evaluations in the users' home feeds. As a preliminary step, Table 1 provides some summary statistics on the prevalence of the candidate debate in the home feeds of the survey respondents. Column 1 displays the share of tweets that contain the term tvduell.[9] On average, about two out of three tweets contain the term, highlighting the enormous attention that was paid to the debate. Although this figure is subject to some user-specific variation, even at the lower bound close to one in three tweets referenced the debate. Turning to the candidates more specifically (columns 2–4), we find that, on average, about 50 percent of tweets referenced either of the two candidates, but no less than 15 percent.

Table 2 provides some summary statistics on the estimated candidate evaluations in the users' home feeds. Overall, the performance of Merkel was rated reasonably poorly. Nevertheless, there is some pronounced variation across the feeds and, indeed, at the positive end her performance is rated rather favorably – more favorably even than the comparable

[8]The indicators for media usage frequency explicitly queried the usage frequency of political information: "How frequently have you sought out information about political parties and the federal election on the following channels during the past week?" ["An wie vielen Tagen haben Sie sich in der vergangenen Woche in den folgenden Kanälen über Parteien und die Bundestagswahl informiert?"].

[9]Debate-related tweets were marked with the hashtag #tvduell.

TABLE 2: Estimated evaluation in feeds

	Merkel	Steinbrück
Mean	-0.71	-0.03
SD	0.21	0.23
Min	-0.92	-0.79
Max	0.57	0.41

The Table provides summary statistics on the evaluation of the candidate performance in the users' home feeds. Column 1 provides the estimated evaluations for Merkel, column 2 provides the figures for Steinbrück.

figure for Steinbrück. The figures for Merkel stand in stark contrast to the estimates for Steinbrück who receives a mixed set of messages, resulting in an average score close to 0. However, in much the same way as before, this figure varies considerably across users, ranging from a strongly negative to a reasonably positive evaluation of the candidate. Importantly, the gap between the overall evaluations of Merkel and Steinbrück suggests that the figures are not a mere artifact of the procedure that is applied for estimating the sentiment in the feeds, but rather reflect a true difference in the performance evaluations of Merkel and Steinbrück.

5 Twitter effects on perceived candidate performance

Turning to the main interest of the present paper, Table 3 presents the results from a model of the perceived performance of Angela Merkel. In all three model variants there is a strong and systematic negative effect of the candidate performance evaluation on Twitter on the perceived performance of the incumbent chancellor.[10] The effect remains significant when controlling for the expected candidate performance prior to the debate and the candidate sympathy rating in model 2. Both of these indicators are plausibly signed in that a more negative expectation is associated with a less favorable performance perception, while a positive candidate sympathy rating is related to a favorable perception of the candidate performance. The effects are reasonably consistent when additionally controlling for the indicators of political media usage (model 3). There is little in the way of a systematic relationship between any of these measures and the perceived candidate performance. The only notable exception

[10] Recall that the perceived candidate performance is collected on a five-point scale, ranging from 1 (*Very good*) to 5 (*Very poor*), such that negative coefficients reflect a more favorable candidate evaluation.

is the indicator for tabloid usage, which suggests that tabloid readers had a somewhat less favorable view of the Merkel performance. Overall, the model exhibits a fair fit with the empirical data, which is able to explain roughly 50 percent of the variance. Notably, including the media usage indicators improves the model fit only marginally.

TABLE 3: Candidate performance evaluation: Angela Merkel

	Model 1	Model 2	Model 3
(Intercept)	1.67^{***}	1.91^{***}	2.11^{***}
	(0.27)	(0.34)	(0.37)
Peformance evalution Twitter	-2.38^{***}	-1.26^{**}	-1.13^{**}
	(0.37)	(0.41)	(0.41)
Expected performance		$0.15^{\dagger}$	0.20^{*}
		(0.08)	(0.08)
Sympathy score: Merkel		-0.10^{***}	-0.12^{***}
		(0.03)	(0.03)
Media usage: Local newspapers			$-0.05^{\dagger}$
			(0.03)
Media usage: National newspapers			-0.03
			(0.03)
Media usage: Tabloid			0.12^{*}
			(0.06)
Media usage: Public television			-0.02
			(0.03)
Media usage: Private television			-0.06
			(0.04)
N	91	91	88
R^2	0.32	0.46	0.53
adj. R^2	0.32	0.44	0.49
Resid. sd	0.73	0.66	0.63

Standard errors in parentheses

† significant at $p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Table 4 presents the results from an identical model for Peer Steinbrück. By and large, the results are comparable to the previous observations. The candidate performance evaluation on Twitter is negatively related to the perceived candidate performance, albeit only marginally statistically significant in the full model. Both the anticipated performance and the candidate sympathy score are systematically related to the perceived performance in the expected direction. Notably, the expected performance provides a better predictor for the perceived performance – both substantively as well as statistically – for Steinbrück relative to Merkel. As before, the media usage indicators are barely related to the perceived candidate performance, with the exception that readers of national newspapers are slightly more favorable of the Steinbrück performance. Interestingly, the coefficient for the tabloid usage is once more positive, although not statistically significant. This observation might reflect the notion that tabloid news consumers have a generally more cynical view of politics (Jebril, Albæk and de Vreese, 2013) which expresses itself in a more negative performance evaluation of both candidates. As before, the model explains approximately 50 percent of the variance, with the indicators of media usage barely improving the model fit.

TABLE 4: Candidate performance evaluation: Peer Steinbrück

	Model 4	Model 5	Model 6
(Intercept)	1.95^{***}	1.28^{***}	1.37^{***}
	(0.08)	(0.25)	(0.29)
Peformance evalution Twitter	-1.70^{***}	-0.72^{*}	$-0.65^{\dagger}$
	(0.33)	(0.36)	(0.36)
Expected performance		0.33^{***}	0.36^{***}
		(0.09)	(0.09)
Sympathy score: Steinbrück		-0.07^{*}	$-0.07^{\dagger}$
		(0.03)	(0.04)
Media usage: Local newspapers			−0.00
			(0.03)
Media usage: National newspapers			-0.06^{*}
			(0.03)
Media usage: Tabloid			0.07
			(0.05)
Media usage: Public television			0.00
			(0.03)
Media usage: Private television			−0.02
			(0.04)
N	90	89	87
R^2	0.23	0.48	0.54
adj. R^2	0.22	0.46	0.49
Resid. sd	0.72	0.60	0.58

Standard errors in parentheses

† significant at $p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

6 Conclusion

Televised candidate debates are a staple as much as a focal point of contemporary election campaigns. They are viewed as make-or-break moments for political careers by pundits and scholars alike. Given their perceived status and their effects on electoral outcomes, it is crucial to understand the factors that shape perceptions of candidate performance.

The phenomenon of the second screen has prominently shaped the way that viewers consume television in recent years. The present contribution has taken this development as occasion to assess the link between candidate evaluations on the second screen and perceptions of candidate performance. Employing evidence from a two-panel survey study of a sample of German Twitter users, we were able to relate a systematic measure of the candidate-specific sentiment with perceived candidate performances. The analysis has provided evidence that evaluations on the second screen are indeed strongly related to perceptions of candidate performance. Moreover, these effects are stable when controlling for a number of factors that might affect candidate evaluations.

The study is subject to the limitation that in order to ensure a good coverage of German Twitter users, we employed a convenience sample. Although the research is explicitly interested in Twitter users that are well-versed in the medium and therefore not representative of the general public, it is ultimately impossible to assess whether the sample is even representative of German Twitter users. While it would have been technically feasible to recruit a sample that is more representative of the general public by relying on a general social survey, there are severe practical restrictions for such a recruitment strategy. First and foremost, the number of respondents would have to be prohibitively large in order to ensure a sufficient number of Twitter users that are willing to share their account details. This limitation is particularly severe in a country like Germany where Twitter is a veritable elite phenomenon and where, consequently, fewer respondents would have a Twitter account to begin with.

Previous research on the second screen phenomenon has assessed the effects of active Twitter usage on candidate perceptions, while the present study has considered effects of users as consumers of opinion. Future research might fruitfully take an integrated perspective that links both aspects of the interaction with Twitter and candidate evaluations. In fact, the text analysis techniques employed for the present study are well applicable to the content of the tweets that users send during the broadcast.

Students of political preferences have emphasized the crucial status of inter-personal exchange for political perceptions and vote choices (Beck et al., 2002; Cho, 2005; Schmitt-Beck, 2003). The present study indicates that joining an online community and exchanging political views might be subject to similar systematics, such that a less immediate inter-personal exchange leads to substantial effects on political preferences all the same. In a political environment that is increasingly characterized by added layers of online discourse, this observation might be considered positive inasmuch as it limits the potentially detrimental effects of a discourse monopoly by traditional media. At the same time, the present study needs to end on the cautions note that online communication in a bubble of likeminded individuals (Barberá, 2015) can result in systematic biases in political perceptions.

Appendix

TABLE 5: Summary statistics

Variable	Mean	SD	Min	Max
Perceived performance: Merkel (W2)	3.37	0.88	1.00	5.00
Perceived performance: Steinbrück (W2)	1.99	0.81	1.00	5.00
Expected performance: Merkel (W1)	2.81	0.90	1.00	5.00
Expected performance: Steinbrück (W1)	2.35	0.92	1.00	5.00
Sympathy score: Merkel (W1)	-1.30	3.15	-5.00	5.00
Sympathy score: Steinbrück (W1)	1.18	2.82	-5.00	5.00
Media usage: Local newspapers (W1)	2.03	2.53	0.00	7.00
Media usage: National newspapers (W1)	2.34	2.64	0.00	7.00
Media usage: Tabloid (W1)	0.42	1.42	0.00	7.00
Media usage: Public television (W1)	4.16	2.21	0.00	7.00
Media usage: Private television (W1)	0.99	1.78	0.00	7.00

References

Banducci, Susan A. and Jeffrey A. Karp. 2003. "How elections change the way citizens view the political system: Campaigns, media effects and electoral outcomes in comparative perspective." *British Journal of Political Science* 33(3):443–67.

Barberá, Pablo. 2015. "Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data." *Political Analysis* 23(1):76–91.

Beck, Paul Allen, Russell J. Dalton, Steven Greene and Robert Huckfeldt. 2002. "The social calculus of voting: Interpersonal, media, and organizational influences on Presidential choices." *American Political Science Review* 96(1):57–73.

Bekafigo, Marija Anna and Allan McBride. 2013. "Who tweets about politics? Political participation of Twitter users during the 2011 gubernatorial elections." *Social Science Computer Review* 31(5):625–43.

Bennett, W. Lance and Robert M. Entman. 2001. *Mediated politics: Communication in the future of democracy.* Cambridge: Cambridge University Press.

Benoit, William L. and Glenn J. Hansen. 2004. "Presidential debate watching, issue knowledge, character evaluation, and vote choice." *Human Communications Research* 30(1):121–44.

Benoit, William L., Glenn J. Hansen and Rebecca M. Verser. 2003. "A meta-analysis of the effects of viewing U.S. Presidential debates." *Communication Monographs* 70(4):335–50.

Berger, Adam L., Stephen A. Della Pietra and Vincent J. Della Pietra. 1996. "A maximum entropy approach to natural language processing." *Computational Linguistics* 22(1):39–71.

Bishop, George F. 1978. "Debate watching and the acquisition of political knowledge." *Journal of Communication* 28(4):99–113.

Breiman, Leo. 2001. "Random forests." *Machine Learning* 45(1):5–32.

Bryant, Jennings and Mary Beth Oliver. 2009. *Media effects: Advances in theory and research.* New York: Routledge.

Chaffee, Steven H. 1978. "Presidential debates: Are they helpful to voters?" *Communication Monographs* 45(4):330–46.

Chiang, Chun-Fang and Brian Knight. 2011. "Media bias and influence: Evidence from newspaper endorsements." *Review of Economic Studies* 78(3):795–820.

Cho, Jaeho. 2005. "Media, interpersonal discussion, and electoral choice." *Communication Research* 32(3):295–322.

Corner, John and Dick Pels. 2003. *Media and the restyling of politics: Consumerism, celebrity and cynicism.* London: Sage.

D'Alessio, Dave and Mike Allen. 2000. "Media bias in presidential elections: A meta-analysis." *Journal of Communication* 50(4):133–56.

DellaVigna, Stefano and Ethan Kaplan. 2007. "The Fox News effect: Media bias and voting." *Quarterly Journal of Economics* 122(3):1187–234.

D'Orazio, Vito, Steven Landis, Glenn Palmer and Philip Schrodt. 2014. "Separating the wheat from the chaff: Applications of automated document classification using support vector machines." *Political Analysis* 22(2):224–42.

Doughty, Mark, Duncan Rowland and Shaun Lawson. 2011. "Co-viewing live TV with digital backchannel streams." Proceedings of the International Interactive Conference on Interactive Television.

Druckman, James N. and Michael Parkin. 2005. "The impact of media bias: How editorial slant affects voters." *Journal of Politics* 67(4):1030–49.

Faas, Thorsten and Jürgen Maier. 2011*a*. Das TV-Duell. In *Zwischen Langeweile und Extremen: Die Bundestagswahl 2009*, ed. Hans Rattinger, Sigrid Roßteutscher, Rüdiger Schmitt-Beck and Bernhard Weßels. Wiesbaden: VS Verlag für Sozialwissenschaften pp. 223–36.

Faas, Thorsten and Jürgen Maier. 2011*b*. Medienwahlkampf: Sind TV-Duelle nur Show und damit nutzlos? In *Der unbekannte Wähler? Mythen und Fakten über das Wahlverhalten der Deutschen*, ed. Evelyn Bytzek and Sigrid Roßteutscher. Frankfurt am Main: Campus pp. 99–114.

Freelon, Deen and David Karpf. 2015. "Of big birds and bayonets: Hybrid Twitter interactivity in the 2012 presidential debates." *Information, Communication and Society* 18(4):390–406.

Geer, John G. 1988. "The effects of presidential debates on the electorate's preferences for candidates." *American Politics Research* 16(4):486–501.

Gentzkow, Matthew and Jesse M. Shapiro. 2010. "What drives media slant? Evidence from U.S. daily newspapers." *Econometrica* 78(1):35–71.

Gottfried, Jeffrey A., Bruce W. Hardy, Kenneth M. Winneg and Kathleen Hall Jamieson. 2014. "All knowledge is not created equal: Knowledge effects and the 2012 presidential debates." *Presidential Studies Quarterly* 44(3):389–409.

Gottfried, Jeffrey A., Bruce W. Hardy, R. Lance Holbert, Kenneth M. Winneg and Kathleen Hall Jamieson. 2017. "The changing nature of political debate consumption: Social media, multitasking, and knowledge acquisition." *Political Communication* 34(2):172–99.

Han, Eunyoung and Sang-Woo Lee. 2014. "Motivations for the complementary use of text-based media during linear TV viewing: An exploratory study." *Computers in Human Behavior* 32:235–43.

Hawthorne, Joshua, J. Brian Houston and Mitchell S. McKinney. 2013. "Live-tweeting a presidential primary debate: Exploring new political conversations." *Social Science Computer Review* 31(5):552–62.

Holbrook, Thomas M. 1999. "Political learning from presidential debates." *Political Behavior* 21(1):67–89.

Holtz-Bacha, Christina, Ana Ines Langer and Susanne Merkle. 2014. "The personalization of politics in comparative perspective: Campaign coverage in Germany and the United Kingdom." *European Journal of Communication* 29(2):153–70.

Houston, J. Brian, Joshua Hawthorne, Matthew L. Spialek, Molly Greenwood and Mitchell S. McKinney. 2013. "Tweeting during presidential debates: Effect on candidate evaluations and debate attitudes." *Argumentation and Advocacy* 49:301–11.

Houston, J. Brian, Mitchell S. McKinney, Joshua Hawthorne and Matthew L. Spialek. 2013. "Frequency of tweeting during presidential debates: Effect on debate attitudes and knowledge." *Communication Studies* 64(5):548–60.

Jebril, Nael, Erik Albæk and Claes H. de Vreese. 2013. "Infotainment, cynicism and democracy: The effects of privatization vs. personalization in the news." *European Journal of Communication* 28(2):105–21.

Kahn, Kim Fridkin and Patrick J. Kenney. 2002. "The slant of news: How editorial endorsements influence campaign coverage and citizens' views of candidates." *American Political Science Review* 95(1):49–69.

Kalsnes, Bente, Arne H. Krumsvik and Tanja Storsul. 2014. "Social media as a political backchannel: Twitter use during televised election debates in Norway." *Aslib Journal of Information Management* 66(3):313–28.

Karvonen, Lauri. 2010. *The personalization of politics: A study of parliamentary democracies*. Colchester: ECPR Press.

Klemm, Michael and Sascha Michel. 2014. "Social TV und Politikaneignung: Wie Zuschauer die Inhalte politischer Diskussionssendungen via Twitter kommentieren." *Zeitschrift für Angewandte Linguistik* 60(1):3–35.

Maier, Jürgen. 2007*a*. Erfolgreiche Überzeugungsarbeit: Urteile über den Debattensieger und die Veränderung der Kanzlerpräferenz. In *Schröder gegen Merkel: Wahrnehmung und Wirkung des TV-Duells 2005 im Ost-West-Vergleich*, ed. Marcus Maurer, Carsten Reinemann, Jürgen Maier and Michaela Maier. Wiesbaden: VS Verlag für Sozialwissenschaften pp. 91–109.

Maier, Jürgen, Marcus Maurer, Carsten Reinemann and Thorsten Faas. 2006. "Reliability and validity of real-time response measurement: A comparison of two studies of a televised debate in Germany." *International Journal of Public Opinion Research* 19(1):53–73.

Maier, Jürgen and Thorsten Faas. 2005. Schröder gegen Stoiber: Wahrnehmung, Verarbeitung und Wirkung der Fernsehdebatten im Bundestagswahlkampf 2002. In *wahlen und wähler: Analysen aus Anlass der Bundestagswahl 2002*, ed. Jürgen W. Falter, Oscar W. Gabriel and Bernhard Weßels. Wiesbaden: VS Verlag für Sozialwissenschaften pp. 77–101.

Maier, Jürgen and Thorsten Faas. 2011. ""Miniature campaigns" in comparison: The German televised debates, 2002-09." *German Politics* 20(1):75–91.

Maier, Jürgen, Thorsten Faas and Michaela Maier. 2013. Mobilisierung durch Fernsehdebatten: Zum Einfluss des TV-Duells 2009 auf die politische Involvierung und die Partizipationsbereitschaft. In *Wahlen und Wähler: Analysen aus Anlass der Bundestagswahl 2009*, ed. Bernhard Weßels, Harald Schoen and Oscar W. Gabriel. Wiesbaden: Springer VS pp. 79–96.

Maier, Jürgen, Thorsten Faas and Michaela Maier. 2014. "Aufgeholt, aber nicht aufgeschlossen: Wahrnehmungen und Wirkungen von TV-Duellen am Beispiel von Angela Merkel und Peer Steinbrück 2013." *Zeitschrift für Parlamentsfragen* 45(1):38–54.

Maier, Michaela. 2007*b*. Verstärkung, Mobilisierung, Konversion: Wirkung des TV-Duells auf die Wahlabsicht. In *Schröder gegen Merkel: Wahrnehmung und Wirkung des TV-Duells 2005 im Ost-West-Vergleich*, ed. Marcus Maurer, Carsten Reinemann, Jürgen Maier and Michaela Maier. Wiesbaden: VS Verlag für Sozialwissenschaften pp. 145–65.

Maurer, Marcus and Carsten Reinemann. 2007. Warum TV-Duelle Wahlen entscheiden können: Befunde und Konsequenzen der TV-Duell-Studie 2005. In *Schröder gegen Merkel: Wahrnehmung und Wirkung des TV-Duells 2005 im Ost-West-Vergleich*, ed. Marcus Maurer, Carsten Reinemann, Jürgen Maier and Michaela Maier. Wiesbaden: VS Verlag für Sozialwissenschaften pp. 229–46.

McKinney, Mitchell S. and Diana B. Carlin. 2004. Political campaign debates. In *Handbook of political communication research*, ed. Lynda Lee Kaid. Mahwah: Lawrence Erlbaum Associates pp. 203–34.

McKinney, Mitchell S., J. Brian Houston and Joshua Hawthorne. 2014. "Social watching a 2012 Republican presidential primary debate." *American Behavioral Scientist* 58(4):556–73.

Miller, Joanne M. and Jon A. Krosnick. 1996. News media impact on the ingredients of presidential evaluations: A program of research on the priming hypothesis. In *Political Persuasion and Attitude Change*, ed. Diana C. Mutz, Paul M. Sniderman and Richard A. Brody. Ann Arbor: University of Michigan Press pp. 79–99.

Mukherjee, Partha and Bernard J. Jansen. 2014. Social TV and the social soundtrack: Significance of second screen interaction during television viewing. In *Social computing,* *behavioral-cultural modeling and prediction*, ed. William G. Kennedy, Nitin Agarwal and Shanchieh Jay Yang. Heidelberg: Springer pp. 317–24.

Nyhuis, Dominic and Johannes Friederich. 2017. Begleitung des TV-Duells auf Twitter. In *Merkel gegen Steinbrück: Analysen zum TV-Duell vor der Bundestagswahl 2013*, ed. Thorsten Faas, Jürgen Maier and Michaela Maier. Wiesbaden: Springer VS pp. 157–72.

Nyhuis, Dominic and Thorsten Faas. Forthcoming. Twitter als Spiegel öffentlicher Meinung? Die Schätzung politischer Bewertungen auf Twitter mittels halbautomatischer Textklassifizierung. In *Computational Social Science: Die Analyse von Big Data*, ed. Joachim Behnke, Andreas Blätte, Kai-Uwe Schnapp and Claudius Wagemann. Baden-Baden: Nomos.

Pan, Zhongdang and Gerald M. Kosicki. 1997. "Priming and media impact on the evaluations of the President's performance." *Communication Research* 24(1):3–30.

Reinemann, Carsten and Jürgen Wilke. 2007. "It's the debate, stupid! How the introduction of televised debates changed the portrayal of chancellor candidates in the German press, 1949-2005." *International Journal of Press/Politics* 12(4):92–111.

Scheufele, Dietram A. 2000. "Agenda-setting, priming, and framing revisited: Another look at cognitive effects of political communication." *Mass Communication & Society* 3(2-3):297–316.

Scheufele, Dietram A. and David Tewksbury. 2007. "Framing, agenda-setting, and priming: The evolution of three media effects models." *Journal of Communication* 57(1):9–20.

Schmitt-Beck, Rüdiger. 2003. "Mass communication, personal communication and vote choice: The filter hypothesis of media influence in comparative perspective." *British Journal of Political Science* 33(2):233–59.

Seni, Giovanni and John Elder. 2010. "Ensemble methods in data mining: Improving accuracy through combining predictions." *Synthesis Lectures on Data Mining and Knowledge Discovery* 2(1):1–126.

Shamma, David A., Lyndon Kennedy and Elizabeth F. Churchill. 2009. "Tweet the debates: Understanding community annotation of uncollected sources." Proceedings of the First SIGMM Workshop on Social Media.

Shim, Hongjin, Poong Oh, Hyunjin Song and Yeonkyung Lee. 2015. "An exploration of motivations for two screen viewing, social interaction behaviors, and factors that influence viewing intentions." *Cyberpsychology, Behavior, and Social Networking* 18(3):158–64.

Sparks, Glenn G. 2006. *Media effects research: A basic overview.* Belmont: Thomson Wadsworth.

Steinwart, Ingo and Andreas Christmann. 2008. *Support vector machines.* New York: Springer.

Strömbäck, Jesper. 2008. "Four phases of mediatization: An analysis of the mediatization of politics." *International Journal of Press/Politics* 13(3):228–46.

Swanson, Linda L. and David L. Swanson. 1978. "The agenda-setting function of the first Ford-Carter debate." *Communication Monographs* 45(4):347–53.

Taber, Charles S., Milton Lodge and Jill Glathar. 2001. The motivated construction of political judgments. In *Citizens and politics: Perspectives from political psychology*, ed. James H. Kuklinski. Cambridge: Cambridge University Press pp. 198–216.

Trilling, Damian. 2015. "Two different debates? Investigating the relationship between a political debate on TV and simultaneous comments on Twitter." *Social Science Computer Review* 33(3):259–76.

Vaccari, Christian, Augusto Valeriani, Pablo Barberá, Richard Bonneau, John T. Jost, Jonathan Nagler and Joshua Tucker. 2013. "Social media and political communication: A survey of

Twitter users during the 2013 Italian general election." *Rivista Italiana di Scienza Politica* 43(3):381–410.
van Cauwenberge, Anna, Gabi Schaap and Rob van Roy. 2014. ""TV no longer commands our full attention": Effects of second-screen viewing and task relevance on cognitive load and learning from news." *Computers in Human Behavior* 38:100–09.
van Zoonen, Liesbet and Christina Holtz-Bacha. 2000. "Personalisation in Dutch and German politics: The case of talk show." *Javnost – The Public* 7(2):45–56.