Hybrid Media Consumption: How Tweeting During a Televised Political Debate Influences the Vote Decision

Misa Maruyama[1], Scott P. Robertson[2], Sara Douglas[1], Bryan Semaan[2], Heather Faucett[1]
[1]Communication and Information Sciences Program, [2]Information and Computer Sciences Department
University of Hawai'i at Mānoa
Honolulu, HI 96822 USA
{misattm, scott.robertson, sarad, bsemaan, ha8}@hawaii.edu

ABSTRACT

An increasing number of people are using microblogs to broadcast their thoughts in real time as they watch televised political events. Microblogging social network sites (SNSs) such as Twitter generate a parallel stream of information and opinion. It is presumed that the additional content enhances the viewing experience, but our experiment explores the validity of this assumption. We studied how tweeting, or passively observing Twitter during a debate, influenced affect, recall and vote decision. For most measures, participants' average feeling and recall toward the candidates did not depend on Twitter activity, but Twitter activity did matter for vote choice. People who actively tweeted changed their voting choice to reflect the majority sentiment on Twitter. Results are discussed in terms of the possibility that active tweeting leads to greater engagement but that it may also make people more susceptible to social influence.

Author Keywords

Social networking; social media; e-participation; e-citizenship; microblogging; digital democracy

ACM Classification Keywords

K.4.3 [Computers and Society]: Organizational Impacts – Computer-Supported cooperative work

General Terms

Experimentation; Human Factors; Measurement.

INTRODUCTION

Social networking sites (SNSs), like Facebook and Twitter, have become important tools in the realm of political information seeking, deliberation and decision making. A recent survey by the Pew Internet and American Life project [24] found that 25% of respondents said they used SNSs to discuss or debate political issues with others. Thirty percent of respondents said that their friends post occasionally about politics, and 25% said that they have become more involved politically as the result of

information in their social networking feeds.

The study of the use of social media to obtain and discuss political information and candidates has become an active research topic in the HCI and CSCW community. Several types of social media have been studied like political blogs [1, 9, 38, 40], discussion forums [23], online videos [18, 19], and pure SNSs like Facebook and Twitter [14, 25, 26, 27]. Twitter, a popular microblogging SNS, has more recently received particular attention by the research community in the context of political discourse. Tumasjan and colleagues [39] found that sentiment and political topic distribution on Twitter reflect the same patterns as "real-world" sentiment and topic. Designers have begun to discuss how sentiment and content on Twitter might be visualized during live events, like debates, to enhance the experience [7]. In this paper, we wish to understand the cognitive and affective experiences of using Twitter during a live political event from the perspective of the individual.

Media hybridity

In mass media research, watching television has traditionally been viewed as a solitary and passive activity [30]. However, people are now appropriating interactive media during television broadcasts to create a hybrid media environment. The rise of the Internet has allowed more people to become "networked" viewers who capitalize on the web's interactivity and connectivity to augment their television viewing experience. For example, according to a recent Pew Internet and American Life project survey, about 20% of cellphone owners who use the Internet, e-mail or apps have also used their phone to see what other people are saying online about a TV program they are watching [35]. The same percentage used their phone to post their own comments online [35]. Watching a televised event while connecting with others to talk about the event blurs the line between the one-to-many broadcast audience and the many-to-many networked audience [16].

Twitter use during political debates

Twitter is a form of lightweight chat that allows users to send short messages in real time to people subscribed to their streams. Twitter messages—or "tweets"—can only be 140 characters or less, and can be sent or retrieved using a variety of technologies like laptops and mobile phones.

One main feature of Twitter is the use of the '@' symbol, followed by a user name, i.e. @username, within the text of

a message. The use of the '@' symbol is also referred to as an @reply or an @mention, where the message is directed to a specific user or talking about a user. Another major feature of Twitter is the use of the '#' symbol—hashtag—followed by a keyword like #PresidentialDebate. When people compose a tweet that includes a hashtag, the message text is available to the public (barring privacy settings) and will be available to anyone who is searching for any given hashtag and keyword combination. The tweet feeds for hashtag searches are updated while tweets are being posted by Twitter users.

Because Twitter allows people to broadcast their thoughts in real time, people have used it to interact with others during televised political events. This makes watching a debate a social experience and generates a parallel stream of information and opinion that augments the debating candidates' words. Media and campaigns often encourage this activity by suggesting hashtags and mentions that microblogging participants can use. Several studies have examined the role of microblogging during political debates [17, 32, 33, 34]. Mascaro and Goggins [17], for example, found that communities coalesced around hashtags, where tweeters could interact with political officials and journalists. Shamma et al. [32] studied the use of Twitter during President Obama's inauguration and found that tweet activity declined during critical moments, as people were more preoccupied with witnessing the event on television than tweeting about it.

Much of the research on social influence on microblogging platforms has used Social Network Analysis (SNA) to study massive Twitter data sets [2, 4, 11, 13, 21, 29]. Cha, Haddadi, Benevenuto and Gummadi [4] found that the number of retweets and mentions reflect a different kind of influence than the number of followers. Huberman, Romero, and Wu [11] also found that interactional measures may provide a more accurate reflection of a social network than bond-based ties (follower-followee relationships). As further support for these findings, a study by Romero, Galuba, Asur and Huberman [28] found that influence depends not only on the size of an audience, but also on how active or passive the audience members are. As a way to predict influence, Anger and Kittl [2] proposed a quantitative method that not only included the number of followers, but also the frequency of retweets and mentions.

While much of the research on social influence and Twitter has analyzed large data sets, experiments have the potential to show the complex way in which people are influenced on the individual level. Zhu, Huberman and Luon [42] conducted a series of online experiments to gauge social influence in which they asked participants to choose between a pair of images with and without knowledge of others' opinions. They found that social influence caused users to switch their choice, although the effect was nuanced: social influence was stronger for delayed decisions and in situations where users faced a moderate, instead of a large, number of opposing opinions.

In a field experiment, we examined the influence of using Twitter during a political debate on users' affect, cognition and vote choice. This is the first of this type of study, as far as we are aware. The power of an experiment is two-fold. One, through random assignment, we eliminate self-selection bias. Second, we attempt to control the information available during the debate by asking participants to view the debate on a particular channel and, for those who are viewing Twitter, to view a particular hashtag feed. We maintained external validity by asking participants to use Twitter while viewing the debate in their natural environments.

While our exploratory study was not guided by hypotheses, we held a general belief that using Twitter while watching a debate would influence voters' affect, cognition and vote decision. We compared people who could tweet to those who viewed content passively in order to understand how engagement might influence the way users feel, think and vote. Our research questions were:

RQ1: Does exposure to Twitter influence feelings toward political candidates and is this different for people who actively tweet versus those who passively monitor Twitter?

RQ2: Does exposure to Twitter influence what is learned about political candidates and is this different for people who actively tweet versus those who passively monitor Twitter?

RQ3: Does exposure to Twitter influence vote choice and is this different for people who actively tweet versus those who passively monitor Twitter?

METHOD

Participants

Fifty-one Twitter users, 29 women and 22 men, participated in the study. They were recruited through multiple methods and sites. These included our laboratory's Twitter feed, recruitment fliers and in-class recruitment on the University of Hawai`i at Mānoa campus, in-person recruitment at local community meetings for the general public, and an e-mail newsletter announcement distributed via a local online newspaper. Participants were included only if they were in Hawaii, familiar with Twitter (held a Twitter account and had posted at least one tweet), and had agreed to watch a candidate debate on a specific date. Participants were compensated with a $20 gift card for participating on the night of the debate and an additional $10 gift card for participating in a follow-up online questionnaire two weeks later.

The participants represented a diversity of age categories. Eleven participants were 18 to 20 years old, 19 were 21 to 29 years old, 11 were 30 to 39 years old, 6 were 40 to 49 years old, and 4 were 50 or older. Ninety-two percent had completed at least some college, and 46% had completed at

least some graduate school. Twenty-three people, or nearly half of our respondents, reported being Independent or Non-Partisan. Twenty-three participants identified as Democrats and four identified as Republican. The average participant's political interest was moderately high. The means, minimums and maximums for political interest and Twitter use are shown in Table 1.

Most of the participants were daily users of Twitter. About 80 percent looked at Twitter at least once a day, and about half of the participants posted a tweet daily. To get a sense of participants' political activity on Twitter, we asked them to rate how frequently they looked at Twitter to learn about politics. Interestingly, a majority of participants never or rarely went to Twitter for political information, yet 60 percent said they stumbled upon political tweets on a daily basis even when they were not looking for them.

Frequency	Mean	Min	Max
Look at Twitter (1=Never)	4.08	1	5
Tweet (1=Never)	3.49	1	5
Political Interest (1=Very low)	3.48	1	5
Stumble upon political tweets (1=Never)	3.62	1	5
Go to Twitter to learn about politics (1=Never)	2.62	1	5
Tweet about politics (1=Never)	2.02	1	5

Table 1: Participant demographics on 1-5 scales.

Setting

We conducted the study during the November 2012 U.S. election. Rather than expose participants to an election they were unfamiliar with, we wanted to present them with an election they may care about. In some of our previous laboratory studies on social media and political deliberation, we exposed participants to content related to an election in a different part of the United States. But the participants shared in post-experiment interviews that the material was not relevant to them, saying for example:

"… I was just not really caring because we previously said it's not in Hawaii. So, it's just like if it doesn't apply to me, I shouldn't, I don't really care."

To ameliorate this issue, our field experiment approach raised external validity by asking participants to learn about real candidates they would be eligible to vote for – with the hope that it would increase engagement with the study material.

We selected the election to replace Hawaii's U.S. Senator Daniel Akaka, who was retiring after serving 22 years in the Senate. Two locally well-known politicians were vying for his seat: sitting U.S. Congresswoman Mazie Hirono (Democratic party candidate), who also served two terms as Hawaii's lieutenant governor, and former two-term Hawaii

Governor Linda Lingle (Republican party candidate). We chose this race due to its prominence in local politics as well as its visibility on the national stage. The October 16, 2012, debate was the second of five debates between Hirono and Lingle. It took place three weeks prior to the November 6 general election and lasted one hour. A local news station and a local online newspaper hosted the event.

Design

Three independent groups were formed based on their interaction with Twitter:

• Tweeters: Tweeters watched the debate and tweeted whenever and whatever they pleased. In this condition, we asked participants to log in to their personal Twitter accounts and start by searching #kitvdebate. We asked them to actively tweet during the debate using the hashtag #kitvdebate.

• Twitter Observers: Twitter observers watched the debate and monitored the Twitter stream, but did not tweet. In this condition, we asked participants to log in to their personal Twitter accounts and start by searching #kitvdebate. We asked them to monitor Twitter during the debate but refrain from tweeting.

• No Twitter (Control): Participants in this condition watched the debate and did not look at Twitter.

Measures of affect, knowledge, memory, and vote choice (described below) were taken at three different times relative to the debate: pre-debate, immediately after the debate, and two weeks after the debate.

Thus the design for most analyses was a 3x3 mixed design with independent groups defined by their level of interaction with Twitter and repeated measures defined by the times at which measurements were taken.

At all three times (pre-debate, immediately following the debate, and two weeks after the debate), dependent measures were taken on knowledge, affect, and voting preferences. Recall was taken at only the second and third times (as there would have been nothing to recall prior to the debate). Dependent measures were as follows:

• *Knowledge of the race*. Participants were asked to "Please indicate your general knowledge level of the 2012 U.S. Senate election in Hawaii" and given the response choices of 1=No knowledge to 5=Very high knowledge.

• *Knowledge of the candidate*. Participants were asked to "Please indicate your knowledge about each candidate." Next to the candidate names, the following response choices were provided: 1= "No knowledge" to 5= "Very high knowledge."

• *Feeling thermometer*. Subjects were asked to rate the candidates on a "feeling thermometer" with the response choices of 1="Cold" to 7="Warm." Feeling thermometers

are commonly used in research on attitudes toward political candidates [41].

• *Liking the candidate.* Subjects were asked to "Please indicate how much you like/dislike this candidate" using the response scale of 1="Dislike very much" to 5="Like very much." This item was used as a means of comparison against the feeling thermometer indicator, which also assessed how much respondents liked a candidate.

• *Likelihood of voting for candidate.* Subjects were asked to "Please indicate how likely it is that you would vote for this candidate" using the response scale of 1="Definitely would not," to 5="Definitely would."

• *Feeling toward the candidate.* Participants were asked to rate the candidates on five affective terms: "angry," "hopeful," "afraid," "proud," and "anxious". The first four items on the checklist were used in the American National Election Studies, and "anxious" was added per Marcus, Neuman and MacKuen [15]. The checklist questions were in the form "When you think about <candidate name>, how <affect term> does she make you feel?" The response scale ranged from 1="Not <affect term> at all," to 5="Extremely <affect term>."

• *Vote.* Participants were asked, "If the 2012 U.S. Senate election were held today, whom would you vote for?" using the responses 1="Mazie Hirono" and 2="Linda Lingle."

• *Recall.* The dependent measures for recall were asked as the following two free response items:

> • *Debate recall.* Participants were asked to "Please write down as many things as you can remember from the debate. Recall from memory without using any reminders (i.e. open tabs, other windows) or asking anyone who might be with you."

> • *Twitter recall.* In addition to the debate recall, participants who looked at Twitter were also asked to "Please write down as many things as you can remember from Twitter during the debate. Recall from memory without using any reminders (i.e. open tabs, other windows) or asking anyone who might be with you."

For the recall measures, the free response entry field for each question was 20 lines long and 100 characters wide. We asked participants not to worry about spelling, punctuation or complete sentences and recommended they spend about 10 minutes on each free recall response.

• *Tweets.* All of the public tweets that included the #kitvdebate hashtag during the debate were collected using the streaming Twitter API. We also used the streaming API to collect all of the tweets posted by the participants during the debate. We conducted a sentiment analysis of the tweets about each candidate to code for positive, negative or neutral sentiment. We identified tweets that mentioned the candidates by first querying the candidates' names,

including possible misspellings and wildcards to capture non-exact matches (i.e., possessives and hashtags). The tweets were filtered into three groups: tweets that mentioned only Hirono, tweets that mentioned only Lingle and tweets that mentioned both candidates. If the tweet mentioned only one candidate, it was coded as having positive, negative or neutral sentiment about the candidate. If a tweet mentioned both candidates, two codes were assigned, including one for each candidate. We also coded the participants' tweets to determine whether they declared a winner of the debate.

Procedure

Surveymonkey.com was used for the online consent form and all questionnaires. One hour before the debate, participants were e-mailed a link to directions and the pre-exposure questionnaire. After completing the pre-exposure questionnaire, the participants were randomly assigned to one of the three Twitter conditions (without being told about the other conditions) and asked to prepare to watch the debate. All participants were asked to view the debate either on a televised broadcast on a specific local news channel or as a live stream on the station's website. Immediately after the debate, participants were e-mailed a link to an online post-exposure questionnaire, which they were required to complete within one hour. Two weeks after the debate, participants were again e-mailed a request to complete a follow-up online questionnaire. While it is possible that new information in the subsequent two weeks influenced the dependent measures, we concluded that if a major event disrupted the election, any influence would likely be similar across all of the groups.

We asked participants in the Tweet and Twitter Observer groups to start by searching #kitvdebate on Twitter because the hashtag had been popular during a previous political debate that had aired on the same station. We intentionally exposed the Tweet and Twitter Observer groups to the same Twitter content to keep the conditions consistent while manipulating only their level of engagement with Twitter.

Our intent was to insert our participants into the naturally occurring dialogue on Twitter. Use of the widely distributed #kitvdebate hashtag exposed the participants to enough voices to ensure inconspicuousness. Use of this hashtag also increased the study's external validity, exposing participants to the full diversity of voices – candidates' campaigns, news sources, advocacy groups and citizens – available to all Twitter users. In addition, because participants' perception of their audience influences their communicative acts [16], it was important that they knew the audience was real, broad, and consisted of other Twitter users watching the debate (including those outside of our study).

RESULTS

Rating Scales

Rating scale data was analyzed using a 3x3 mixed design Analysis of Variance, with the exception of the recall data which was analyzed using a 3x2 mixed-design Analysis of Variance. Missing data was replaced by the mean of the relevant scale across all participants, regardless of condition. No more than four values were replaced for any scale.

Since we used six rating scale dependent measures in the same 3x3 design (knowledge of the race, knowledge of the candidate, feeling thermometer, liking, and likelihood of voting), we applied the Bonferroni correction for multiple significance tests when considering this data. With this correction, experimentwise significance at $p<.05$ will require $p<.008$ (.05/6). This correction was applied to all ANOVAs discussed below.

The descriptive statistics for the rating scales suggest the data were fairly normally distributed. Skewness ranged from -.895 to 1.83, and kurtosis ranged from -1.30 to 3.12, which are within normal limits. The standard deviation ranged from .85 to 1.85, which is normal given a 5-point or 7-point scale.

Over time, subjects grew more angry about Hirono, $F(2,96)=17.545$, $p<.001$ (means = 1.61, 2.44, and 2.12 for pre-exposure, immediate post-exposure, and two-week post exposure, respectively) and more afraid regarding Hirono, $F(2,96)=7.275$, $p=.001$ (means = 1.65, 2.12, and 1.80 for pre-exposure, immediate post-exposure, and two-week post exposure, respectively), and more anxious, $F(2,96)=5.543$, $p<.01$ (means = 1.81, 2.28, and 2.04 for pre-exposure, immediate post-exposure, and two-week post exposure, respectively).

Other rating scale measures were not significant.

		Democrat	Republican
Tweet			
	Before	11	4
	Immediate	7	8
	Two Weeks	8	7
Observe Twitter			
	Before	6	8
	Immediate	5	9
	Two Weeks	5	9
No Twitter			
	Before	11	7
	Immediate	11	7
	Two Weeks	13	5

Table 2. Frequencies of participants across the three groups reporting their vote for the Democrat or Republican before, immediately after, and two weeks after the debate.

Vote choice

The vote choice dependent measure was a dichotomous category choice (Hirono or Lingle), thus the data was analyzed in terms of frequencies of participants who chose one or the other candidate and tested using the chi-square statistic.

Table 2 shows the raw frequencies of subjects in all three Twitter groups who reported that they would vote for the Democrat (Hirono) or the Republican (Lingle) before, immediately after, and two weeks after the debate (four subjects were removed from this analysis for missing data).
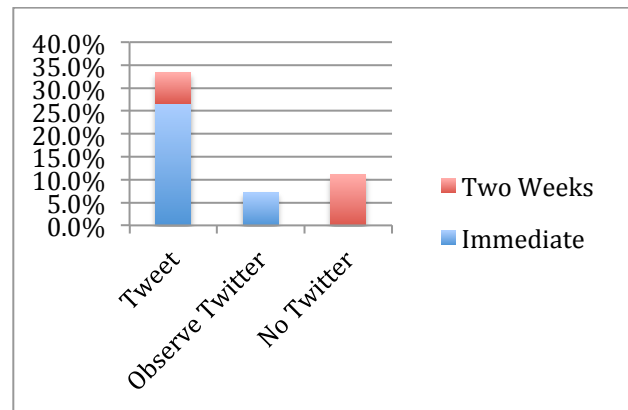

Figure 1: Percentage of participants in the three Twitter groups who changed their vote immediately after and two weeks after the debate.

Figure 1 expresses this data as the percentage of participants in the three Twitter groups who changed their vote (in either direction) immediately after the debate and two weeks later. The Tweet group had far more changes than the other groups, especially immediately but also after a delay. A chi-square test on the change percentages shows a significant contingency between the Twitter group factor and the time measure, $\chi2(2)=27.9$, $p<.0001$. In other words, active Tweeters were the most likely to change their position.
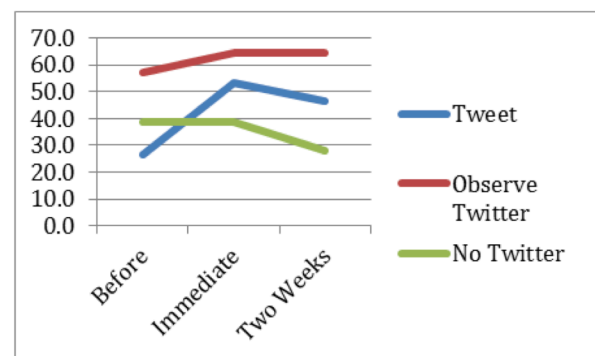

Figure 2. Percentage of subjects in the three Twitter groups reporting that they would vote for the Republican before, immediately after, and two weeks after the debate.

Figure 2 expresses the data as the percentage of Republican votes in each Twitter group at each time period. In Figure 2 we note a strong movement in the active Tweet group toward the Republican candidate immediately after the debate, which was still present two weeks later.

Even though the groups started out in different places (the Twitter Observers had more Republicans to begin with), there was room to see a similar change in the Twitter Observer group.

A chi-square test shows a marginally significant contingency between the Twitter group factor and the time measure for the Republican percentage data, $\chi^2(4)=8.35$, $p<.08$. In other words, active Tweeters were the most likely to change their position toward the Republican candidate immediately, and this change was persistent. We examine reasons for this when we treat the content of the tweets in a subsequent section.

Recall

We analyzed the recall data using a method similar to grounded theory [37], with the goal of identifying what type of information was recalled and whether the types varied across conditions. We define recall as it has been defined in other studies [12]: all ideas presented in the free recall response, including opinions, personal stories and inaccurate memories. Each participant's recall was split into idea units, which were often clauses that contained a subject and predicate. However, sometimes they were as short as a noun or as long as a phrase (if parts of the phrase failed to make sense when standing alone).

A subset of the recall data was independently coded by two authors – not for final coding decisions but to identify situations where the coding scheme did not fit. One additional code was identified through this process and one was eliminated, and through iterative cycles of redefining the codes and reducing under axial coding, we collaboratively identified a set of 13 codes.

The coders independently coded one third of the recall of the debate using Atlas.ti. Inter-rater reliability between the two coders was substantial (κ=.70). When the coders disagreed on coding a recall item, they discussed the discrepancy to come to an agreement. The first author coded the remaining two-thirds of the recall data.

For brevity, we discuss only the seven codes that were used 15 times or more. These included:

• Candidate Tactics: comments about a candidate's political ploys to influence the meaning of an event, issue or their opponent's identity.

• Political Views: comments about a candidate's political behavior, opinions or platform issues.

• Public speaking: comments about a candidate's verbal and non-verbal communications skills.

• Character: judgments of candidate's integrity, personality.

• Outside References: discussion of events that occurred outside of the debate itself including personal stories, current events and discussions that occurred on Twitter.

• Final Impression: sweeping judgment about a candidate that summarizes a participant's opinion on the candidate

• Candidate Competence: judgments about a candidate's experience, knowledge or understanding of a political topic, or ability to serve.

The mean numbers of recall instances for each code category across all participants are shown in Figure 3. Participants most frequently recalled ideas about Candidate Tactics, Political Views and Public Speaking and Character.

We analyzed the recall data using a 3x2 mixed-design Analyses of Variance. The two independent variables included group and time.

There was a main effect of time for Candidate Tactics, $F(1,45)=5.542$, $p<.05$ (means = 3.77 and 2.50 for immediate post-exposure and two-week post exposure, respectively), and Character, $F(1,45)=5.597$, $p<.05$ (means = 1.00 and .542 for immediate post-exposure and two-week post exposure, respectively). Both decreased from the initial recall to the follow-up recall.

The candidate competence recall showed a main effect of group, $F(2,45)=3.20$, $p<.05$ (means = .16, .67, and .12 for Tweet, Twitter Observer and No Twitter groups, respectively). In other words, Twitter Observers recalled more items in this category than the Tweeters and the No Twitter group. We found no significant difference between the groups for the other recall categories.
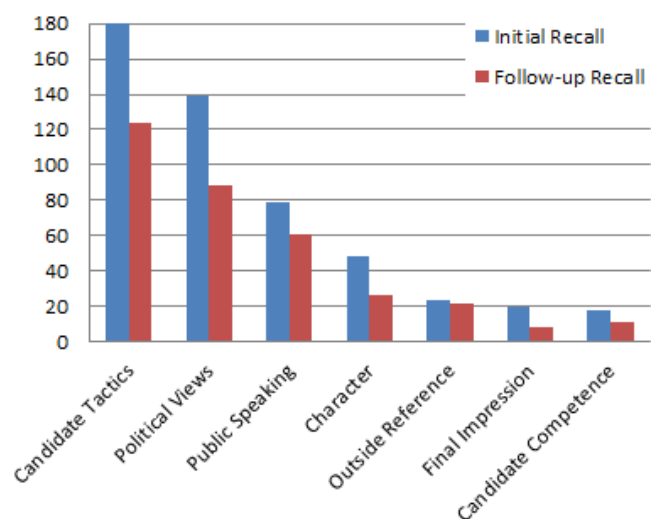

Figure 3: Recall Counts by Code Type

Tweet analysis

We manually coded #kitvdebate tweets that explicitly mentioned the candidates during the debate. We collected 407 tweets that were posted using the #kitvdebate hashtag

during the hour-long debate. Of them, 303 mentioned at least one candidate's name, including 102 tweets that mentioned both candidates. The average number of tweets posted by general Twitter users who used the #kitvdebate hashtag was 5.9 tweets. In comparison, Tweeters in our study posted an average of 9.5 tweets, almost double the number posted by all users who included the hashtag.

The analysis showed that tweets about Lingle were mostly favorable and tweets about Hirono were mostly critical. Tweets that mentioned Lingle were 73 percent positive, 16 percent negative and 11 percent neutral. Tweets that mentioned Hirono were 9 percent positive, 87 percent negative and 4 percent neutral. The percentages include tweets that mentioned one of the candidates or both of the candidates. More than a third of the tweets that mentioned both candidates were positive about Lingle and negative about Hirono.

We also conducted an analysis of tweets posted by participants in the Tweet group to determine what they were tweeting during the debate. Our intention was to analyze not only what they saw (tweets posted using the #kitvdebate hashtag) but also what they said (participant tweets). While presenting a complete analysis is beyond the scope of this paper, we were interested in the extent to which participants were vocalizing their support for the candidates.

We found that three of the four "vote switchers" in the Tweet condition posted a tweet declaring that Lingle won the debate. For example, one vote switcher used a well-known local Pidgin English phrase to say that Lingle was winning the debate: "Linda is small kine killing it #kitvdebate." At about 43 minutes into the debate, another vote switcher in the Tweet group modified a retweet saying, "Me too. RT @username: i am a democrat and am 100% siding with Lingle #whatsithis #kitvdebate #hisen". In the last two minutes of the debate, the third vote switcher wrote: "Hirono needs to get to the point with concise responses! I think Lingle nailed it. #kitvdebate." The fourth vote switcher in the Tweet condition did not declare a debate winner and was the only participant who changed her vote back to Hirono two weeks after the debate. In other words, only the vote switchers who declared a debate winner on Twitter stuck with their new vote decision after two weeks. The three other Tweeters who posted a comment about Lingle winning the debate had voted for her prior to the debate and stuck with their vote two weeks later.

DISCUSSION
Our first research question was: Does exposure to Twitter influence feelings toward political candidates and is this different for people who actively Tweet versus those who passively monitor Twitter? We found no effects of Twitter exposure on our rating scales. All participants experienced a surge in negative feelings toward Hirono and positive feelings toward Lingle immediately after the debate, which

eased with time but did not quite return to pre-debate levels. We therefore found no evidence to answer "yes" to this research question.

The second research question was: Does exposure to Twitter influence what is learned about political candidates and is this different for people who actively Tweet versus those who passively monitor Twitter? We found only a main effect of Twitter group for the number of Candidate Competence recall items. Mondak and Huckfeldt [20] found that cues about candidate competence influence voting decision even when people receive clear signals about a candidate's political party and ideology, which was the case in the debate study. However, recall about candidate competence comprised a small fraction of the total number of recall items, so it may not have wielded influence on feeling toward the candidates or vote decision.

Our most interesting finding is related to our third research question: Does exposure to Twitter influence vote choice and is this different for people who actively Tweet versus those who passively monitor Twitter? We found that participants in the Tweet group were significantly more likely to change their vote decision immediately after the debate and to stick with their new vote decision two weeks later. By contrast, the Twitter Observer and No Twitter groups were less swayed by the debate. Their votes remained mostly unchanged after the debate and after the two-week delay. Thus, we can answer that active Tweeting seems to have more of an influence on vote choice than just exposure to Twitter, which in turn is not much different than no exposure to Twitter.

We propose two possible explanations for the observed vote change in the Tweet group: The first explanation is that there existed increased engagement through the cognitive act of creating the messages–allowing for new viewpoints to form [8]. Eveland [8] notes that the act of participating in discussion presents the opportunity to process the information in new ways as it is re-synthesized to produce conversation with others. The reformulation and cognitive elaboration necessary for conversing can then lead to changes in attitude beyond that of those with solitary motives such as to privately view media [8, 31]. It may be that engagement matters beyond observation. Those who engaged in discussion may have processed the information more elaborately or in new ways compared to the other groups because they were expecting to engage in discussion with others.

The second explanation is related to social influence, or, in this case, the majority sentiment on Twitter [5]. The #kitvdebate tweets were predominantly pro-Lingle (tweets that mentioned Lingle were 73 percent positive) and anti-Hirono (tweets that mentioned Hirono were 87 percent negative). To better understand how prevalent group conformity was amongst people in the Twitter conditions (Tweeters and Twitter Observers), the lead author coded the Twitter recall data for mentions of a debate winner. We

found that 16 of the 26 respondents who completed the initial recall mentioned a Twitter favorite in the debate. They told the same story: "most" or "many" people on Twitter praised Lingle or criticized Hirono. There was a similar consensus in the follow-up recall. This suggests that participants were cognizant of the majority opinion on Twitter. Research suggests that the greater the size of the majority, the more likely the minority will be to conform [3].

Cialdini and Goldstein [5] describe how social influence, or peer pressure, can lead people to change their beliefs or behavior to reflect the group majority. Conformity tends to take two forms: informational and normative [5]. Informational conformity refers to privately changing your beliefs to reflect the opinions of credible others. Normative conformity means changing behavior to fit in to the crowd, even when this change is not accompanied by a private change in opinion [36].

If the Tweeters switched their opinion to reflect the opinions of people who they deemed to be credible, then this would be an example of informational conformity. On the other hand, tweeting about the popular candidate in order to attain social rewards would demonstrate normative conformity. There was some evidence of social pressure in the #kitvdebate tweet stream, which suggests the possibility of normative conformity. In cases where people showed positive sentiment towards Hirono, their posts were met with criticism (a reflection of group majority sentiment). For example, one person in the minority who tweeted "Hirono is schooling Lingle" met backlash: "Hirono isn't schooling anyone." When the Hirono supporter later wrote, "It's a wrap, Hirono's more genuine, logical and people-oriented. Lingle's all tacticsdirty politics," someone else replied "R U on CRACK?"

Interestingly, three of the four vote switchers declared on Twitter that Lingle was the debate winner. The vote switchers might have been expressing what they believe to be accurate information based on the opinions of knowledgeable others (informational conformity) or they may have been conforming to group norms even though it may not have been attached to a private change in opinion (normative conformity) [36]. The two types of conformity are frequently intertwined [6] and often work in tangent to maintain one's self-concept [5].

For instance, the Tweeters may have declared Lingle as the debate winner to gain social acceptance at first (public compliance) but later changed their vote decision (private acceptance) to maintain a consistent self-concept. Twitter Observers could not commit to a new opinion publicly. They did not have the opportunity to conform to the group norm publicly and did need to modify their vote decision to keep their self-concept intact. We speculate that communicating their pro-Lingle sentiment may have furthered the group bond for the Tweeters and made them more likely to switch their vote to reflect the sentiment of the Twitter majority.

Regardless of why engagement may have influenced the vote decision, it seems that being able to participate can lead to real change whereas merely observing may not. While social media has been criticized for having a polarizing effect [1], our study suggests that this is not always true. People's vote decisions can be swayed, meaning confirmation bias does not always lead people to interpret information in a way that supports pre-existing beliefs [22]. While our study suggests that the majority may be able to sway the opinion of the minority, the reverse is not necessarily true. No one in the Tweet or Twitter Observer groups switched their vote from the Republican favorite on Twitter to the Democratic candidate.

Implications for Theory

Most of the research that has focused on microblogging during live political events has analyzed data on a macro scale [7, 29, 32, 33]. Few studies have studied the cognitive and affective effects of participation during political events on the microbloggers themselves. Although we did not find evidence that engaging on Twitter had any influence on what people knew or how they felt about the candidates, our study did suggest that Twitter engagement mattered when it came to vote choice. This indicates that the relationship between participation and action may be different from the relationship between participation and affect and intention.

We found an interesting pattern in which vote switchers frequently declared a debate winner – a form of public commitment that may have been related to their decision-making process. We suggest future studies include post-experiment interviews to further examine the potential relationship between expressing a belief publicly and the decision-making process.

Our study design allowed us to explore the private beliefs and feelings behind public communication on social media. While big data collection, SNA, and automated content analysis are effective in studying the nature of interaction on social media (behavior), experimental studies can show *how* the interaction influences the thoughts and feelings of users (cognition and affect). Moreover, while research methods such as SNA can study the content created by contributors, they cannot tell the stories of the countless people who do not post. Our study explores the experience of silent social media users by including the Twitter Observer group. Future work should investigate microblogging in different political contexts, including debates in which the political sentiment on Twitter is more even-handed, as well as in national debates.

Implications for Practice

Our study suggests that having the power to tweet about a political event may influence the tweeter's vote decision in ways that reading tweets may not.

We advanced two possible explanations for this: active participation heightened cognitive processing of the debate or it made the minority more susceptible to the influence of the majority opinion. These are not mutually exclusive. If the capacity to tweet led to more elaborate memories (which we did not find in this study), then tweeting could be viewed as a means to developing a more thoughtful electorate willing to listen to the political opposition. On the other hand, listening to the majority may increase cognitive processing of the debate, especially if the majority opinion is different from one's own opinion. But this may not always be good for democracy. If active participation increases group conformity, then tweeting could be viewed as culprit in increased homogenization of political thought. It would also imply that at times decisions can be based on social rewards rather than accurate information, which threatens rational discourse in the public sphere [10]. We cannot determine which explanation is true based on our data. The stakes for democracy are high, which is why we believe more research must be conducted on how social media participation influences the way people think, act and vote.

LIMITATIONS
Our participants were more highly educated, more politically interested, more liberal and had more Twitter experience than the average U.S. citizen; thus, any generalizations are limited to this demographic. However, our study was designed to examine how the use of Twitter influenced a politically interested and technologically experienced group of Twitter users.

Also, importantly, we have several limitations related to our experimental design that should be addressed in future work. Firstly, in our study, we randomly assigned participants to one of the three Twitter conditions. Despite using random assignment, the groups were different prior to the debate on at least one dimension that we cared about: the ratio of people who voted for the Democrat (Mazie Hirono) was higher in the Tweet and the No Twitter groups than in the Twitter Observer group. This may be an indication that our groups were different in a systematic way, which would mean that group differences rather than the treatment may have influenced the results. We attribute the randomization problem to our relatively small sample size. We feel it would be beneficial in future studies to use a larger sample size, which would likely yield groups that were more similar prior to the debate, or to use stratified randomization to control and balance the influence of covariates. Other potential covariates that we did not measure or account for include participants' propensity to change their minds and their susceptibility to social influence. To address these issues, future studies should use existing, or develop new, measures for how willing people are to change their minds or be influenced by others.

Secondly, this was an 'in situ' study that looked at how users behave and feel in their natural tweeting environment,

increasing ecological validity. However, as is the nature of this type of approach, the evidence of internal validity is harder to judge since we did not control participants' behavior. We only controlled what we *invited* participants to do. For example, we do not know whether the participants watched and discussed the debate with others physically present with them, nor whether they viewed social media content beyond the assigned hashtag feed. Thus, it is possible that people beyond the #kitvdebate hashtag stream influenced feelings, knowledge and votes. We recommend that future field experiments monitor participants' online and offline behavior more carefully. Future studies may also be conducted in a controlled laboratory setting, where participant behavior can be manipulated, albeit at the expense of ecological validity.

Thirdly, the design of the study allowed Twitter Observers to view tweets posted by the Tweeters. Therefore, it is possible that the Twitter Observers were influenced by both their treatment and the treatment of the Tweet group. To control for this, interaction between groups can be prevented by artificially constructing simulations of identical but *separate* Twitter feeds for each group to view. Because the feeds would be experimental sandboxes, they would not be live nor would they be plugged into real world discourse. Again, ecological validity may decrease with this more controlled approach.

Lastly, our participants' political leanings may have contributed to the way in which our experiment unfolded. Twenty-three participants reported being Independent or Non-Partisan, 23 identified as Democrats and 4 identified as Republican. Because almost half of our respondents identified as Independent or Non-Partisan, our participants may have been more open to changing their vote decisions.

CONCLUSION
An increasing number of people are using social media to learn about and discuss politics, including microbloggers who provide live commentary about televised political debates. The result is a hybrid media environment in which anyone with an Internet connection can watch the conversation unfold or join the discussion. Our study suggests that people who intend to actively post may experience the debate differently than people who do not.

In this experiment, we found that users who could contribute to the conversation on Twitter were more likely to switch their vote to reflect the majority. Additionally, it so happened that vote switching seemed to be related to a public declaration on Twitter of a debate victor. When people post, it is assumed that they have the capacity to influence others. But this study suggests that social media contributors may also have the power to influence themselves.

REFERENCES

1. Adamic, L., & Glance, N. (2005). The political blogosphere and the 2004 U.S. election: Divided they blog. *Proc. of the International Workshop on Link Discovery*.

2. Anger, I., & Kittl, C. (2011). Measuring influence on Twitter. In *I-KNOW*, 31.

3. Asch, S.E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs*, *70*(9).

4. Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, K. (2010). Measuring user influence in Twitter: The million follower fallacy. *Proc. of the Int'l AAAI Conference on Weblogs and Social Media*.

5. Cialdini, R.B., & Goldstein, N.J., (2004). Social Influence: Compliance and Conformity. *Annual Review of Psychology, 55*, 591-621.

6. David, B., & Turner, J. C. (2001). Majority and minority influence: A single process self-categorization analysis. In C. De Dreu; N. De Vries (Eds.), *Group consensus and minority influence: Implications for innovation* (pp. 91-121). Malden: Blackwell Publishing.

7. Diakopoulos, N. A., & Shamma, D. A. (2010). Characterizing Debate Performance via Aggregated Twitter Sentiment. *Proc. of CHI '10*. New York, NY:ACM.

8. Eveland, W. P. (2004). The effect of political discussion in producing informed citizens: The roles of information, motivation, and elaboration. *Political Communication, 21*(2), 177-193.

9. Farrell, H., & Drezner, D.W. (2008). The power and politics of blogs. *Public Choice*, 134, 15- 30.

10. Habermas, J. (1989). *The structural transformation of the public sphere*. Cambridge: MIT Press.

11. Huberman, B., Romero, D. M., and Wu, F. (2009). Social networks that matter: Twitter under the microscope. *First Monday,* 14. 1-5. http://firstmonday.org/ojs/index.php/fm/article/view/2317/2063

12. Hyman, I.E. (1994). Conversational remembering: Story recall with a peer versus for an experimenter. *Applied Cognitive Psychology*, 8, 49-66.

13. Kivran-Swaine, F., & Naaman, M. (2011). Network properties and social sharing of emotions in social awareness streams. *Proc. of the conference on Computer supported cooperative work*, 379-382. ACM.

14. Lampe, C., Ellison, N., & Steinfield, C. (2006). A face (book) in the crowd: Social searching vs. social browsing. *Proc. of the conference on Computer Supported Cooperative Work*.

15. Marcus, G.E., W.R. Neuman, & M. MacKuen. (2000) *Affective Intelligence and Political Judgment*. Chicago: University of Chicago Press.

16. Marwick, A., & boyd, d. (2011). I Tweet Honestly, I Tweet Passionately: Twitter Users, Context Collapse, and the Imagined Audience. *New Media & Society*, 1-20.

17. Mascaro, C., & Goggins, S.P. (2012). Twitter as Virtual Town Square: Citizen Engagement During a Nationally Televised Republican Primary Debate. APSA 2012 Annual Meeting Paper.

18. Milliken, M., & O'Donnell, S. (2008). User-generated online video: The next public sphere? *IEEE Int'l Symposium on Technology and Society*, 1-3.

19. Milliken, M., Gibson, K., & O'Donnell, S. (2008). User-generated video and the online public sphere: Will YouTube facilitate digital freedom of expression in Atlantic Canada? *American Communication Journal*, 10, 3.

20. Mondak, J.J., & Huckfeldt, R. (2006). The accessibility and utility of candidate character in electoral decision making. *Electoral Studies, 25*(1), 20-34.

21. Naaman, M., Boase, J., & Lai, C. H. (2010). Is it really about me?: message content in social awareness streams. *Proc. of the conference on Computer supported cooperative work*.

22. Nickerson, R. S.(1998) Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*(2), 175-220.

23. Poor, N. (2005). Mechanisms of an online public sphere: The website Slashdot. *Journal of Computer-Mediated Communication*, 10, 2.

24. Raine, L., & Smith, A. (2012). Politics on social networking sites. Pew Internet and American Life Project.

25. Robertson, S.P., Douglas, S., Maruyama, M., & Chen, L. (2012). Political dialog evolution in a social network, In *Proc. of the International Digital Government Research Conference*. New York: ACM, 40-48.

26. Robertson, S.P., Douglas, S., Maruyama, M., & Semaan, B. (2013). Political discourse on social networking sites: Sentiment, in-group/out-group orientation and rationality. *Information Polity 18(2)*, 107-126.

27. Robertson, S.P., Vatrapu, R.K., & Medina, R. (2009). The social life of social networks: Facebook linkage

patterns in the 2008 U.S. presidential election, *Proc. of the International Digital Government Research Conference*. New York: ACM, 6-15.

28. Romero, D.M., Galuba, W., Asur, S., and Huberman, B.A. (2011). Influence and Passivity in Social Media. *Proc. of the European Conference on Machine Learning and Knowledge Discovery in Databases.*

29. Romero, D.M., Meeder, B., and Kleinberg, J. (2011). Differences in the Mechanics of Information Diffusion Across Topics: Idioms, Political Hashtags, and Complex Contagion on Twitter. *Proc. of the International World Wide Web Conference.*

30. Ruggiero, T. (2000) Uses and Gratifications Theory in the 21st Century. *Mass Communication and Society, 3*(1), 3-37.

31. Scheufele, D. A. (2002). Examining differential gains from mass media and their implications for participatory behavior. *Communication Research*, *29*(1), 46-65.

32. Shamma, D. A., Kennedy, L., & Churchill, E. (2010a). Tweetgeist: Can the twitter timeline reveal the structure of broadcast events. *CSCW Horizons*.

33. Shamma, D. A., Kennedy, L., & Churchill, E. F. (2009). Tweet the debates: understanding community annotation of uncollected sources. *Proc. of the SIGMM Workshop on Social Media.*

34. Shamma, D. A., Kennedy, L., & Churchill, E. F. (2010b). Conversational Shadows: Describing Live Media Events Using Short Messages. *Proc. of ICWSM.*

35. Smith, A., & Boyles, J. (2012). The rise of the 'connected viewer.' *Pew Internet and American Life Project.*

36. Stangor, C. (2004). *Social groups in action and interaction*. Psychology Press.

37. Strauss, A. L. & Corbin, J. M. (1998). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. California: Sage.

38. Tremayne, M. (2006). *Blogging, citizenship, and the future of media*. Routledge, New York.

39. Tumasjan, A., Sprenger, T., Sandner, P., & Welpe, I. (2010) Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. *Proc. of International AAAI Conference on Weblogs and Social Media,* 178-185.

40. Vatrapu, R., S.P. Robertson, and W. Dissanayake. (2008). Are Political Weblogs Public Spheres or Partisan Spheres? *International Reports on Socio-Informatics, 5*(1), 7-26.

41. Weisberg, H.F., and J.G. Rusk. (1970). Dimensions of Candidate Evaluation. *American Political Science Review, 64*, 1167-1185.

42. Zhu, H., Huberman, B.A., Luon, Y. (2012). To switch or not to switch: Understanding influence in online choices. *Proc. of CHI 2012*, 2257-2266.