Is News Sharing on Twitter Ideologically Biased?

Jonathan Scott Morgan
School of Journalism
Michigan State University
404 Wilson Road
East Lansing, MI 48824
morga125@msu.edu

M. Zubair Shafiq
CSE Department
College of Engineering
Michigan State University
East Lansing, MI 48824
shafiqmu@msu.edu

Cliff Lampe
School of Information
University of Michigan
105 S. State St
Ann Arbor, MI 48109
cacl@umich.edu

ABSTRACT
In this paper we explore effects of perceived ideology of news outlets on consumption and sharing of news in Twitter. Selective exposure theory suggests that when given access to a broad range of information, people will tend to consume and share news that confirms their existing beliefs and biases. We find that users share news in similar ways regardless of outlet or perceived ideology of outlet, and that as a user shares more news content, they tend to quickly include outlets with opposing viewpoints. This suggests that while perceived ideology does not inspire most Twitter users to treat liberal or conservative news outlets differently, it is a factor in their news consumption and sharing. Specifically, users in our sample who sent multiple tweets tended to increase the ideological diversity in news they shared within two or three tweets, and users' information diversity increased as their number of tweets sent increased.

Author Keywords
Twitter; online discussion; information flow; political discourse; journalism.

ACM Classification Keywords
H.5.3 [Group and Organization Interfaces]: Web-based interaction.

General Terms
Social Computing and Social Navigation; Computer-Mediated Communication; Social Networking Site Design and Use

INTRODUCTION
As social media sites become increasingly popular, more and more everyday communication is migrating from established communication mediums like email and the telephone to the social web. This personal information is joining the increasing amounts of other types of information that government and other organizations are making available online, creating an immense storehouse of information that people are interacting with, linking to and referencing across many different information channels, both mediated and unmediated.

This wealth of available information is theoretically invaluable for citizens in a democracy. Those with an Internet connection have access to an unprecedented diversity of thought, fact, and opinion. Access and use are not the same, however. Bounded rationality is defined by Simon as "rational choice that takes into account the cognitive limitations of the decision-maker - limitations of both knowledge and computational capacity." [38] In the context of a democratic public sifting through this wealth of information to inform civic and electoral decisions, bounded rationality suggests that the public has limited time and cognitive resources to devote to its civic duties, and so will use simplifying strategies when informing civic decisions.

One such simplifying strategy that is theorized to affect news consumption is selective exposure [35,41]. Selective exposure suggests that the Internet's overwhelming news options and ability to choose content without editorial intervention make online news consumers more likely to consume and share news that confirms their preformed biases and avoid news that does not, in order to ease the cognitive burden of analyzing and internalizing discordant information. Some scholars argue that the ease of personalizing content choices on the Internet may increase the likelihood of selective exposure [40]. In addition, examination of links between conservative and liberal blogs showed relatively few references between blogs of different ideologies, potentially indicating that some type of selective exposure was occurring [1].

Online social network sites may help increase exposure of users to news of different types. Most people interact in these online communities with friends or colleagues who have different political ideologies [7,22,32]. The lower switching costs of changing from one online news source to another might also make people more likely to seek out diverse viewpoints online than they would be in other contexts where switching is more costly.

In this study, we capture Tweets that reference 12 popular news outlets across a spectrum of perceived ideology, from liberal to conservative. We examine traits of tweeters, message cascades, and tweets to look for differences in news sharing based on outlets or perceived ideology. We then examine diversity of perceived source ideology in

news tweets shared by Twitter users to see if news sharing choices provide evidence of underlying selective exposure.

LITERATURE REVIEW

There have been a number of studies recently that deal with how people interact and share information using the Internet and social media sites [1,4,18,19,31]. Within Twitter, for example, people have recently studied how hashtags evolve [13], the spread of information in general [25,27,44] and political information in particular [33,37], and the collaborative potential of conversation [21]. In this study, we examine the interactions between multiple channels (journalism and Twitter), the cognitive burden that results as more diverse information is placed online and integrated across channels, and ways that participants in these channels deal with this cognitive burden.

Multiple Interconnected Channels

In the broader sense of mediated communication using the Internet, each of these social network sites and collaborative technologies is one of many interconnected channels for information consumption and sharing on the Internet. Studies of these sites, however, with some notable exceptions [23,24,26,42], tend to limit themselves to a single channel's content and interactions, even though it is clear that these sites are more and more being used to share information across multiple channels, both mediated and unmediated.

A substantial portion of the "user-generated content" on social media sites is content shared from another channel – images aggregated in Pinterest, for example, likes and recommendations of articles in Facebook, or URLs, videos, and images shared in Twitter. The increasing number of mechanisms that make it easier to share content between these sites is creating richer online interactions that span multiple channels. As information channels become more interconnected, there is a need to better understand motivations for and patterns of multi-channel sharing, to learn ways we can design around both adaptive and maladaptive sharing.

Twitter and News

Twitter is a micro-blogging service that allows users to broadcast 140-character messages, called tweets. The Twitter application has a number of affordances that facilitate its use in both news gathering and discussion. Users can follow each other to subscribe to each other's tweets. Any message can be easily re-tweeted, causing it to be resent by a user to all of their followers [44]. Hashtags, categories assigned to tweets preceded by a pound sign ("#"), can be entered as part of any tweet, providing a flexible way to target tweets for a variety of purposes [13,23,33]. URLs included in tweets are automatically shortened, allowing for links to news to be easily shared within the tweet character limit [5]. And at-mentions, explicit references to a user's Twitter username preceded by an at sign ("@"), allow tweets to both target and reference users [21].

Twitter makes sharing information easy, and this helps make the sharing and discussion of news between news outlets and Twitter users a genuine bi-directional exchange. News outlets quote tweets and reporters use twitter in their news-gathering, while many twitter users share and discuss news articles [3,6,39]. This interaction has become an important part of coverage of and reaction to news events, including politics, and elections [9,37] and natural disasters [39,43].

Information Overload and Bounded Rationality

There is also a need to examine how users of increasingly interconnected systems make sense of the rich network of information these linked channels create. Simon's bounded rationality adapts the economic theory of subjective expected utility (SEU) to more realistically account for the cognitive limits of the human brain. SEU conceptualizes people as understanding all possible outcomes and their probabilities, then deciding among them based on maximizing a utility function. Simon believes this entails too heavy a computational burden on decision-makers, and looks for "procedures for choosing that are computationally simpler, and that can account for observed inconsistencies in human choice patterns." [38] These procedures include: making do with information found most easily (satisficing); relying on sources they have used in the past; and when seeking diverse information, using relatively simple heuristics to choose new information sources that simplify decision-making, but could also substantially limit actual source diversity.

Perceived Ideology of News

When examining the ideological diversity of outlets included in news sharing, one must first decide how to categorize news outlets' ideological position. In American journalism, there is often a substantial difference between the perception of slant or bias in a given news outlet [15,29] and the concrete instances of bias in the news they produce [2,14,16,36]. One bad article about a popular topic or person could pre-dispose whole classes of people to assuming an outlet is biased, even when that outlet is otherwise mostly neutral. This perception effect is also exacerbated by the trend of political parties and political partisans methodically asserting that the press is politicized and partisan, regardless of the truth of stories, when journalists make news that the parties do not like [12]. We understand the difference between perception of ideological slant and concrete analysis of content. In this study, we categorize news outlets based on the public's perception of their bias, since perceptions of the similarity or differences with news outlets play a central role in use of selective exposure to choose news content.

Selective Exposure and Ideological Homogeneity

Selective exposure is one of many heuristics that people may use to deal with cognitive overload. Selective exposure is the tendency to limit information exposure to messages that are compatible with already held attitudes or

interests, and to avoid information that might cause cognitive dissonance [8,35].

Some researchers are concerned that as more and more of the sources of information for political discourse (and the venues in which political discourse occur) are accessed via the Internet, the public will respond to increasing amounts of information by seeking out and perpetuating ideological homogeneity - selectively exposing themselves only to like-minded individuals and information sources, leading to an increasingly ill-informed and polarized body politic [40]. Adamic and Glance found the political blogosphere to be polarized in this way in 2004, with homogeneous groups of liberal and conservative bloggers linking to and reading each other, but rarely linking to posts or blogs of the opposite ideology [1].

This fear is not universally accepted, however [10,17,35], and research indicates that most people interact with others in their networks who are of different political ideologies [7,22,32]. In fact, people who may not expose themselves to seemingly contrary news in the mass media channels (e.g. a liberal might avoid watching Fox News on cable) may be exposed to that content by maintaining a diverse social network in online systems.

As people share information in Twitter, they are often choosing to share links to news stories. These news stories may come with a perceived or actual bias that confirms the beliefs of the person sharing that link. Consequently, if selective exposure is at play as a heuristic by which people decide to share information, we would expect them to only share news from sources that confirm those biases.

RESEARCH QUESTIONS
RQ1: How do people share news on Twitter?

RQ2: Is news from news outlets that are perceived to be ideologically similar shared similarly?

RQ3: Are Twitter users more likely to share multiple news items from news outlets that have similar perceived ideology or from outlets with diverse ideologies?

METHODS
To examine the ways that tweets from news outlets across the spectrum of political ideologies in America are included in Twitter conversations, we first selected 12 news outlets broken into three ideological categories: liberal, conservative, and a third category of outlets that are either neutral or are not American, and so are a foreign control. For each outlet, we gathered a week's worth of tweets that contained either a link to news content published on their web site or a formal at-mention of the outlet's Twitter username. We carried out a series of outlet-to-outlet comparisons to create a baseline of variance across outlets, and to make sure that there were differences between the tweets of the different outlets. We also grouped the outlets by ideology (our three categories from above) and a simple measure that compared number of uses to number of tweets in the sample (to assess diversity of those tweeting), then implemented the same comparisons among these groupings, to see how much of the variance between the individual outlets could be explained by ideology or homogeneity of those who tweeted about each outlet. Finally, we shifted focus to the users in our data set, examining the range of outlets each user shared content from to see if sharing behaviors provided evidence of users' ideological or political biases manifesting in selective exposure.

News Outlet Selection and Categorization
For this study, we used Twitter's spritzer feed (a 1% random sample of tweets matching any filter criteria passed to the API) to collect data. Given this limited sample of tweets, we chose national news outlets that would have a larger volume of tweets.

We used the 2010 biennial news consumption survey conducted by The Pew Research Center for the People & the Press [46] as a first source for popular news sources for our three ideological categories: liberal, conservative, and neutral/control. One method the Pew survey uses to examine ideology in news is to ask those it surveys their ideological perspective: liberal, moderate, or conservative. It then asks these same people about their news consumption habits, including use of major popular news outlets, and breaks out consumption based on the proportions of overall readership for each outlet that self-reports liberal, conservative, or moderate/other.

To categorize outlets based on their perceived ideology, we took all outlets with a roughly 2-to-1 liberal-to-conservative audience ratio and added them to our "liberal" category (MSNBC, The New York Times, NPR, and CNN), and added those with the same ratio of conservative to liberal to our "conservative" category (Fox News, Wall Street Journal). We ignored the percentage of consumers that self-reported moderate, since that classification is ambiguous. We also selected one outlet with roughly equal liberal and conservative consumption (USA Today) to be the first member of our "moderate/neutral/control" group.

We then chose a few non-traditional outlets that are relatively new to fill out the ideological categories, adding the Drudge Report to our conservative group and the Huffington Post to the liberal group, and added a few other large outlets to the control/neutral group: BBC News, CBS News, and the Washington Post.

Data Collection
To examine the differences in sharing between tweets from our twelve news outlets, we gathered 1,378,336 tweets from the Twitter spritzer feed (a 1% random sample of tweets matching any filter criteria passed to the API) from February 28, 2012 to March 8, 2012. For each outlet, we ran a separate tweet collector implemented with the tweetstream python module [20] against Twitter's streaming API. We conceptualize exposure to a news outlet broadly as including both exposure to news by an outlet and

News Outlet	Domain Name	URL Shortener	Twitter User	Tweet Count
BBC	bbc.co.uk/news/	bbc.in	@bbcnews	21,658
CBS News	cbsnews.com	t.co	@cbsnews	35,856
CNN	cnn.com	on.cnn.com	@cnn	470,988
Drudge Report	drudgereport.com	drudge.tw	@drudge_report	9,895
Fox News	foxnews.com	fxn.ws	@foxnews	69,879
Huffington Post	huffingtonpost.com	huff.to	@huffingtonpost	186,325
MSNBC	msnbc.msn.com	on.msnbc.com	@msnbc	152,196
New York Times	nytimes.com	nytm.es	@nytimes	131,329
NPR	npr.org	n.pr	@nprnews	16,843
USA Today	usatoday.com	usat.ly	@usatoday	34,118
Wall Street Journal	online.wsj.com	on.wsj.com	@wsj	189,126
Washington Post	washingtonpost.com	wpo.st	@washingtonpost	60,123

Table 1: News outlet selection details

exposure to and awareness of information about an outlet itself, so each collector filtered the overall tweet stream based on the presence of either a URL referencing content on that outlet's web site (indicated by URLs that contained either the organization's web address or their URL shortening domain) or the Twitter username of the outlet's highest level news Twitter account. Only 805 tweets in the sample were from outlet usernames, which is a relatively small number given the size of the sample. The list of outlets and the filter strings we used for each is in Table 1. Because we chose outlets based on their perceived ideological bias and did not take size or popularity into account, some of the outlets in our study are very different in terms of newsgathering resources and popularity. This caused the number of tweets in a week for each outlet to vary, and to vary widely in some cases.

Analysis

Operationalizing Selective Exposure
In this study, we define exposure broadly as any exposure to either a news outlet or the news created by that outlet, independent of attitude or valence toward the outlet or type of information shared. To investigate selective exposure, we look at news sharing behaviors to see if they provide evidence of underlying attitudes affecting news exposure. Our data doesn't provide explicit information on attitudes, beliefs, or biases of the twitter users we are studying, but information that is shared must have been consumed at some level, and so homogeneity or diversity of ideology in

news tweets shared does at least provide information on a subset of each user's news consumption choices, and so has the potential to contain evidence of selective exposure.

Traits of Tweets
We computed 7 different traits per tweet for use in our analysis: retweet count, follower count and count of those the user follows, and user mention count, hashtag count, url count, and tweet text length. We chose these traits to capture at a high level and include in our analysis three concepts: an idea of how far the tweet spread (retweet count); structural traits of the social context and network of the tweeter (follower count – those whom the tweeter influences - and count of those the user follows – those who influence the tweeter); and manifest traits of the content of the tweet itself (user mention count, hashtag count, url count, and tweet text length). See Table 2 for descriptive statistics of these computed traits.

We examined differences in the distributions of these traits across news outlets and categories of news outlets. We also used these traits in logistic regressions to predict the outlet or category to which a given tweet belonged, and to investigate if spread of a tweet, network of the tweeter, or traits of tweet text could help to predict which tweets belonged to which outlet or category.

Categorizing Tweets By Heterogeneity
In addition to our categorization based on perceived ideology of news outlets, we also generated a basic statistic to assess the overall heterogeneity of the people who shared information. The heterogeneity score for each outlet was the ratio of the number of unique users who tweeted that outlet's content to the total number of tweets related to that outlet. This score ranges from 0 to 1, and outlets were clustered into three non-overlapping categories: low (between 0.46 and 0.52: CNN, Fox News, MSNBC, Wall Street Journal), medium (between 0.58 and 0.60: BBC, Drudge, Huffington Post), and high (between 0.66 and 0.70: CBS, NYTimes, USA Today, Washington Post; and NPR with 0.78). A score approaching 1 indicates that more users are responsible for sharing of the outlet's news, and so each user shares fewer items, where a score of 1 would indicate that each tweet for the outlet is sent by a unique user. A score approaching zero indicates that fewer people are responsible for the distribution that outlet receives in Twitter. We call this a heterogeneity score - a reflection of whether those who tweet an outlet's content are a small group sharing many items each (similar to what Adamic and Glance [1] found liberal and conservative bloggers to be, sharing with each other) or a larger group of people sharing fewer items each.
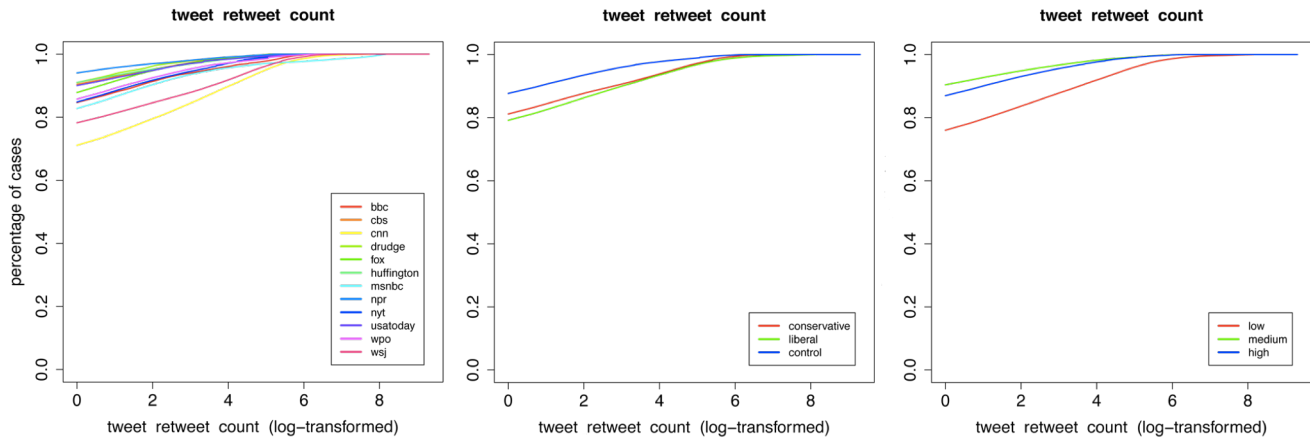
Figure 1: Normalized Cumulative Distribution Function (CDF) plots of retweet count broken out by organization, ideology, and heterogeneity (in that order, from left to right).

CDF Plots and the Two-Sample K-S Test

To assess differences between distributions for news outlets and categories of news outlets, we compared the distributions of the aforementioned tweet traits using normalized Cumulative Distribution Function (CDF) plots and the two-sample Kolmogorov-Smirnov test (K-S test) of similarity of distributions. These methods and statistics are not particularly well-known in this field, and so we'll describe them briefly now before we outline our analysis.

A CDF plot is an alternate way of visually presenting a distribution of data. It takes the counts in a histogram and cumulatively sums them as the function moves from left to right over the X-axis, then plots the cumulative sum as the Y-value for each X. CDF plots make visual comparison of distributions easier, but when differences are subtle (as in our data – see Figure 1), a more precise measure of difference between distributions is needed.

The K-S test is a non-parametric test for the equality of distributions that uses the points from a CDF plot to approximate and compare underlying distributions. It is sensitive to differences in locations and shapes of distributions. The K-S statistic that results runs from 0 to 1, where 0 indicates that points from two CDFs reflect an identical underlying distribution and 1 indicates no overlap between the underlying distributions represented by the two samples, indicating that they are entirely distinct [45].

Comparing Outlets and Outlet Categories

To assess presence of variance in tweets across outlets, we generated CDFs of the aforementioned traits of tweets across individual outlets, outlet categories based on their perceived ideology, and outlet categories based on the heterogeneity or diversity of people referencing their name and content. Figure 1 shows CDF plots for retweet count broken out by news outlet, ideology, and heterogeneity.

We then used these CDF plots to compute K-S statistics for all possible pairs of the news outlets and pairs of categories.

Predicting Outlet and Category Membership

To look at the variance between news outlets and between categories that each of our traits explained, we also implemented logistic regression to see how well our 7 trait variables could predict a given tweet being from a given news outlet or category. We implemented this using SPSS Multinomial Logistic Regression (MLR), which performs a series of binomial logistic regressions across all combinations of a categorical variable using a single sample from the overall data set, allowing per-unit assessment, but also accruing information to create overall estimates of the model, including an estimated R-squared value.

Examining Sharing Behaviors of Users

Since selective exposure is typically identified in individuals, we concluded our analysis by aggregating traits of tweets from each user in our sample to look for patterns of sharing behaviors. Our goal was to find evidence of selective exposure in users' sharing behaviors, i.e. whether users select news outlets they include in tweets based on the perceived ideology of the outlet.

RESULTS

Traits of Tweets

In addition to our conceptual groupings, the traits we computed for tweets fall into two categories: those that are inherently range-limited by a given tweet being at most 140 characters (tweet text length, URL count, hashtag mention count, and user mention count), and those that are not (retweet count, user's follower count and count of people the user follows).

None of the traits were normally distributed, and most appeared to have a power law distribution, so they were log-normalized when used in regression.

	Retweet Count	Follower Count	Following Count	User Mentions	Hashtag Count	URL Count	Text Length
Minimum	0	0	0	0	0	0	3
Maximum	11,032	9,744,612	554,667	17	28	6	353
Mean	18.63	5,166.54	278.54	1.00	.62	.83	107.02
- SE of Mean	*.109*	*91.754*	*4.290*	*.001*	*.001*	*.000*	*.026*
Median	.00	225.00	5.00	1.00	.00	1.00	113.00
Mode	0	0	0	1	0	1	140
Std. Dev.	127.944	107,721.506	5,036.580	1.093	1.480	.537	30.480
Variance	16,369.645	11,603,922,894.819	25,367,141.753	1.194	2.191	.288	929.000
Skewness	17.416	46.915	84.436	2.761	4.120	.157	-.713
- SE of Skewness	*.002*	*.002*	*.002*	*.002*	*.002*	*.002*	*.002*
Kurtosis	472.850	2,551.441	8,673.320	14.588	22.606	2.058	-.380
- SE of Kurtosis	*.004*	*.004*	*.004*	*.004*	*.004*	*.004*	*.004*

Table 2: Descriptive statistics for data calculated per tweet. Standard Error (SE) of Mean, Skewness, and Kurtosis are in italics in a separate row beneath each statistic.

Measures of centrality of all of the traits indicated variance both in the data set as a whole (see Table 2) and across categories. In addition, the ranges of a given trait sometimes differed substantially within categories, e.g. the mean of the retweet count ranges from 1.61 to 44.84 for different outlets and the maximum ranges from 161 to 11,032. Therefore, it was difficult to compare the variance across groupings using measures of centrality alone. To better assess the differences between the distributions of traits for different outlets and groupings of outlets, we sought more sophisticated ways of visualizing and analyzing differences in distributions.

Differences Between Outlets and Outlet Categories

Distribution Differences: CDF plots and K-S Statistics
The visual inspection of normalized CDF plots across outlets showed that while the shapes of the plot lines were similar, suggesting a similar underlying distribution, the slopes and magnitudes varied across news outlets, suggesting minor differences in distributions. The

Tweet Trait	Max K-S	Mean K-S	Median K-S
Retweet Count	0.254	0.080	0.066
Follower Count	0.158	0.067	0.064
Following Count	0.209	0.072	0.062
User Mentions	0.419	0.162	0.134
Hashtag Count	0.292	0.111	0.100
URL Count	0.398	0.172	0.180
Text Length	0.190	0.095	0.089

Table 3: K-S statistic summary by tweet trait.

distributions were too close to visually assess the precise magnitude of these differences.

A comparison of K-S statistic values indicated that while some of the individual news outlets have substantially different distributions for certain traits, the average K-S statistic value for a given trait is never above 0.2, and only 2 of 7 make it above 0.1 (see Table 3 for more details). Some traits have maximum K-S values of around 0.3 or higher (user mention count, K-S = 0.419; URL count, K-S = 0.398; hashtag count, K-S = 0.292), indicating that there are some substantially different distributions among the outlets for these tweet traits. The modest mean and median K-S statistics across all traits suggest that these substantial differences are exceptions and that overall these distributions are similar.

K-S statistic values across outlets also indicate that while there are substantial differences in certain traits between outlets, the distributions of the outlets' tweet traits are mostly similar. All outlets have a maximum K-S statistic values, including all traits, of between 0.279 and 0.419 – some larger than others, but all in the range of K-S statistic values that indicate only moderate difference. However, they also all have mean and median K-S statistic values that are around 0.1 (see Table 4 for more details). While there are some substantial differences in distributions of traits between outlets, they are spread evenly among the outlets, and no outlet has a mean K-S statistic value is above 0.139.

Categorization of outlets based on ideology and heterogeneity of tweeters show even more modest amounts of differences between the underlying distributions of the traits. Across outlet categories based on ideology, all K-S

News Outlet	Max K-S	Mean K-S	Median K-S
BBC	0.398	0.139	0.112
CBS News	0.338	0.098	0.085
CNN	0.311	0.111	0.095
Drudge Report	0.355	0.119	0.098
Fox News	0.279	0.100	0.088
Huffington Post	0.373	0.106	0.084
MSNBC	0.336	0.091	0.084
New York Times	0.419	0.127	0.100
NPR	0.384	0.089	0.072
USA Today	0.398	0.111	0.088
Wall Street Journal	0.351	0.085	0.066
Washington Post	0.419	0.125	0.108

Table 4: K-S statistic summary by outlet.

statistic values are below 0.1, so clustering based on ideology actually seems to mask the differences in traits between outlets. Therefore, it is not useful in explaining the between-outlet distribution differences. Across outlet categories based on heterogeneity, we observe more variance in distributions of traits, with some K-S statistic values around 0.12, and a few values above 0.2. Overall, the K-S statistic values of around 0.2 are still modest.

Logistic Regression
To further explore the variance between news outlets and the amount of that variance our categorizations of news outlets explained, we also implemented Multinomial Logistic Regression to investigate how well these 7 traits could predict a given tweet being from a given news outlet or outlet category.

All three logistic regression models had chi-squared values that indicated statistically significantly fit, $p < .001$, likely because of the size of our data set. Low Pearson correlations among predictors (highest Pearson's $r = 0.281$, then -0.236, then 0.199, and the rest were less than absolute value of 0.072), relatively high tolerance values (all above 0.867), and Variance Inflation Factor values (VIF ranges from 0 to positive infinity, where large values indicate inflation of standard error due to collinearity) between 1 and 1.154 all indicate no violation of MLR's collinearity assumption.

Given the size of our data set, we also assessed model fit using a test that is more robust to large sample sizes than chi-squared. Chance criteria for assessing model fit, adapted from discriminant analysis [34], compares a model's accuracy to that expected by chance to determine the quality of the model. It is not affected by sample size. In discriminant analysis, the specific test used depends on how equally divided the cases are among the different groupings [28]. In multivariate logistic regression, it is

accepted to use the Proportionate Chance criteria regardless of comparative cell sizes. Using this criteria, a model is a good fit when its *Actual Accuracy* is greater than its *Chance Accuracy + 25%*. The individual news outlet model passed the proportionate test, with an actual accuracy of 0.635, a full ten percent greater than the goal of 0.535. The other models were close, but did not meet this criteria (Table 5).

Though the ideology and heterogeneity models are poor fits, the Cox and Snell [11] and Nagelkerke [30] pseudo R-squared values (see Table 6) are in line with the trends observed in K-S statistics. Grouping tweets by outlet

Grouping	Actual Accuracy	Chance Accuracy + 25%
Outlet	***0.635***	0.535
Ideology	0.826	***0.876***
Heterogeneity	0.756	***0.759***

Table 5: Proportionate Chance criteria – MLR model Actual Accuracy compared to Chance Accuracy + 25%.

Grouping	Cox and Snell	Nagelkerke
Outlet	0.228	0.243
Ideology	0.08	0.117
Heterogeneity	0.114	0.152

Table 6: Cox and Snell and Nagelkerke Pseudo R-Squared values for logistic regression models by outlet, ideology,

explains a similar amount of variance as an R-Squared of between 0.23 and 0.24 would represent in a linear regression model, a respectable amount given the sparse sample and the basic nature of traits being analyzed. Grouping by ideology and heterogeneity explained proportionately much less, and as with distribution comparisons, while heterogeneity is the better explainer of variance in tweet traits of the two, it is not better by much.

For the outlet model, SPSS "Likelihood Ratio Tests" indicated that all predictors were significant parts of the model, $p < .001$. None of the predictors had standard errors over 2, which suggests that there were no numerical anomalies with the model.

News Selection at the User Level
In our sample, 81% of users shared tweets from only one outlet. Of the 19% who shared more than one tweet, 44% shared 2, and so 56% of users who shared news from more than one source shared from 3 or more.

Liberal news is more prevalent in this sample than news from conservative or control outlets in terms of numbers of both tweets and users. 77% of users included outlets we categorized as liberal in the tweets they forwarded, either alone or in combination with other ideologies. This fits with the numerical prevalence of tweets from liberal outlets, in particular CNN (69% of tweets were from liberal outlets, 34% overall were from CNN).

It is interesting to note that as people share more tweets, however (see Table 7), their sharing quickly becomes more diverse, and the proportions of people who only share from one ideology drop quickly: to 63% of people who share 2 or more, to 40% when 5 or more tweets are sent, and approaching 30% as number of tweets increases further. Even given the prevalence of liberal tweets, this highlights that as people share more tweets, they are more likely to include news outlets with diverse ideologies, and it isn't just the heaviest tweeters who share from liberal, conservative, and our control – 23% of users who share 5 or more tweets share from all three of our groups.

This trend becomes evident starting when users share 2-3 tweets, and could indicate that while the limited data we've created might not predict our ideological categories well, the categorizations have some merit. Our control group seems not to be very salient – most of the diversity comes from sharing both liberal and conservative tweets, not from including the outlets in our control group – and people tend to include news from both liberal and conservative outlets pretty quickly as they share more.

DISCUSSION
First, this study has limitations. Our data set included a limited number of news outlets and was only a 1% sample of tweets. Though it spanned multiple days, it did not include data from multiple points in time to support longitudinal analysis. Our method of sampling by perceived ideology of outlet resulted in inclusion of outlets with substantially different magnitudes of Twitter sharing traffic, which complicated analysis. We did not differentiate between links to a news outlet's articles and mentions of an outlet's name, and these two types of references to an outlet's content could be part of very different types of conversation, though we assert that they are both evidence of at least some exposure to the news outlet. We only captured the most basic of information on the tweets we gathered. We didn't analyze the Twitter users to assess different user types or goals. Perceived ideology of a news

outlet is complicated, also, and our method of assigning ideology did not have much explanatory power in our data.

Even with these limits, we did learn something about how Twitter users share news. Both our distribution analysis and the logistic regression results suggest that there are a few reasonably substantial differences in tweets across outlets. Overall, however, we observed that people seem to share news in similar ways, regardless of outlet, and the traits we examined did not indicate a substantial difference between outlets. Though our ideology-based categorization might not have been perfect, our sample did include news from a range of ideological perspectives, and the similarity of sharing across these outlets suggests that selective exposure is not a strong actor at the outlet level.

The lack of difference across our categorizations also supports the idea that selective exposure lacks power in influencing news outlet choice when sharing tweets. Grouping tweets by perceived ideology of outlets and our basic measure of heterogeneity both resulted in K-S statistic values smaller than those when outlets were compared, and MLR models for the categories were poor fits and were rejected outright. Therefore, these categorizations were not much help in explaining differences in our tweet data.

Information on users' sharing behaviors also provided indirect refutation of selective exposure, and suggested that even though our ideological categorization did not explain differences in sharing, it does have at least some validity. People who share more than one tweet seem pretty quickly to include news from different ideologies in the news they share, the opposite of what selective exposure would predict. A substantial percentage shared tweets from 3 or more outlets, especially given the limited snapshot present in a 1% sample. Moreover, as number of tweets increased, the percentage of users who included multiple ideologies quickly increased (half of all users who shared 3 tweets included outlets with different ideologies). This could indicate that switching costs are low enough in consuming

Min Tweets	Users	Only Control	Only Lib.	Only Cons.	Only 1	Lib. + Cons.	Cons. + Control	Lib. + Control	Only 2	1 or 2	All 3
1	597,536	0.094	0.663	0.132	0.889	0.049	0.006	0.036	0.091	0.98	0.02
2	178,810	0.036	0.513	0.08	0.628	0.162	0.021	0.12	0.304	0.933	0.067
3	91,232	0.022	0.42	0.061	0.503	0.203	0.018	0.144	0.365	0.868	0.132
4	58,445	0.018	0.365	0.053	0.436	0.215	0.014	0.148	0.376	0.813	0.187
5	41,960	0.016	0.329	0.05	0.395	0.216	0.012	0.146	0.374	0.769	0.231
6	32,240	0.016	0.303	0.049	0.367	0.213	0.01	0.143	0.366	0.734	0.266
7	25,982	0.015	0.286	0.049	0.35	0.208	0.009	0.137	0.354	0.704	0.296
8	21,599	0.015	0.272	0.049	0.336	0.203	0.008	0.133	0.344	0.68	0.32
9	18,314	0.015	0.263	0.05	0.328	0.197	0.008	0.127	0.331	0.659	0.341
15	8,694	0.016	0.23	0.061	0.307	0.169	0.006	0.105	0.279	0.586	0.414

Table 7: Distribution of tweets across ideological spectrum as number of tweets shared increases.

and sharing online news that people are more likely to explore and share different points of view online, even if they are prone to selective exposure in other venues. It could also indicate that people are less affected by selective exposure in general than was previously thought, and more representative data in services like Twitter, even though it is not perfectly representative of the population as a whole, is detailed enough that it is helping us get a more accurate representation of the public.

CONCLUSION

Many researchers have studied Twitter's role in politics and political discussion. In this study, we focus on the role news articles and organizations play in online discussion, how users share news articles, and what those sharing behaviors suggest about the character of the conversation on Twitter, the role news plays in online discussion, and how users filter information. In this study we have shown that Twitter users tend to share news without bias toward or against the perceived ideology of news outlets. While selective exposure might be a strong influence at some level, in the sample of tweets we examined, it does not seem to strongly influence choice of news outlet. Twitter users even seem to actively include news from outlets of different perceived ideology as they share more news, counter to what selective exposure would predict.

There are a number of ways to move this research forward: methods that better take social context and network patterns of sharing into account, for example; better assessing and including traits of users; or including more nuanced analysis of content of tweets and content of things linked by tweets, to better detect the tone of tweets, tweets that are about the same topic, similar topics, etc., and so differentiate between types of information.

Even so, this study indicates that a simplistic model of Twitter news consumers being inchoate and guided by selective exposure is incorrect, and the more nuanced and open-minded information consumer our work implies bears substantial further study.

ACKNOWLEDGEMENTS

REFERENCES

1. Adamic, L.A. and Glance, N. The Political Blogosphere and the 2004 U.S. Election: Divided They Blog. *ACM LinkKDD-2005*, ACM Press (2005).

2. Alterman, E. *What Liberal Media?: The Truth about Bias and the News*. Basic Books, 2004.

3. An, J., Cha, M., Gummadi, K.P., Crowcroft, J., and Quercia, D. Visualizing Media Bias through Twitter. *Sixth International AAAI Conference on Weblogs and Social Media*, (2012).

4. Antin, J. and Cheshire, C. Readers are not free-riders: reading as a form of participation on wikipedia. *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, ACM (2010), 127–130.

5. Antoniades, D., Polakis, I., Kontaxis, G., et al. we.b: the web of short urls. *Proceedings of the 20th international conference on World wide web*, ACM (2011), 715–724.

6. Armstrong, C.L. and Gao, F. Now Tweet This How News Organizations Use Twitter. *Electronic News 4*, 4 (2010), 218–235.

7. Bakshy, E., Rosenn, I., Marlow, C., and Adamic, L. The Role of Social Networks in Information Diffusion. *arXiv:1201.4145*, (2012).

8. Baran, S.J. and Davis, D.K. *Mass Communication Theory: Foundations, Ferment, and Future*. Wadsworth Publishing, 2008.

9. Bruns, A. and Burgess, J. #Ausvotes: How twitter covered the 2010 Australian federal election. *Communication, Politics & Culture 44*, 2 (2011), 37.

10. Chaffee, S.H., Saphir, M.N., Graf, J., Sandvig, C., and Hahn, K.S. Attention to Counter-Attitudinal Messages in a State Election Campaign. *Political Communication 18*, 3 (2001), 247–272.

11. Cox, D.R. and Snell, E.J. *Analysis of Binary Data, Second Edition*. CRC Press, 1989.

12. Crawford, C. *Attack the Messenger: How Politicians Turn You Against the Media*. Rowman & Littlefield, 2007.

13. Cunha, E., Magno, G., Comarela, G., Almeida, V., Gonçalves, M.A., and Benevenuto, F. Analyzing the dynamic evolution of hashtags on Twitter: a language-based approach. *Proceedings of the Workshop on Languages in Social Media (LSM 2011)*, Association for Computational Linguistics (2011), 58–65.

14. D'Alessio, D. and Allen, M. Media bias in presidential elections: a meta-analysis. *Journal of Communication 50*, 4 (2000), 133–156.

15. Eveland, W.P. and Shah, D.V. The Impact of Individual and Interpersonal Factors on Perceived News Media Bias. *Political Psychology 24*, 1 (2003), 101–117.

16. Fico, F., Richardson, J.D., and Edwards, S.M. Influence of Story and Structure on Perceived Story Bias and News Organization Credibility. *Mass Communication & Society 7*, 3 (2004), 301–318.

17. Garrett, R.K. Echo chambers online?: Politically motivated selective exposure among Internet news users. *Journal of Computer-Mediated Communication 14*, 2 (2009), 265–285.

18. Gruhl, D., Guha, R., Liben-Nowell, D., and Tomkins, A. Information diffusion through blogspace. *Proceedings of the 13th international conference on World Wide Web*, ACM (2004), 491–501.

19. Halvey, M.J. and Keane, M.T. Exploring social dynamics in online media sharing. *Proceedings of the 16th international conference on World Wide Web*, ACM (2007), 1273–1274.

20. Halvorsen, R. tweetstream - simple twitter streaming API. https://bitbucket.org/runeh/tweetstream/src.

21. Honey, C. and Herring, S.C. Beyond Microblogging: Conversation and Collaboration via Twitter. *42nd Hawaii International Conference on System Sciences, 2009. HICSS '09*, IEEE (2009), 1–10.

22. Horrigan, J.B., Garrett, K., and Resnick, P. *The Internet and Democratic Debate*. Pew Internet & American Life Project, 2004.

23. Huang, J., Thornton, K.M., and Efthimiadis, E.N. Conversational tagging in twitter. *Proceedings of the 21st ACM conference on Hypertext and hypermedia*, ACM (2010), 173–178.

24. Kittur, A. and Kraut, R.E. Beyond Wikipedia: coordination and conflict in online production groups. *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, ACM (2010), 215–224.

25. Kwak, H., Lee, C., Park, H., and Moon, S. What is Twitter, a social network or a news media? *Proceedings of the 19th international conference on World wide web*, ACM (2010), 591–600.

26. Lerman, K. and Ghosh, R. Information Contagion: An Empirical Study of the Spread of News on Digg and Twitter Social Networks. *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, AAAI Press (2010), 90–97.

27. Van Liere, D. How far does a tweet travel?: Information brokers in the twitterverse. *Proceedings of the International Workshop on Modeling Social Media*, ACM (2010), 6:1–6:4.

28. Marcoulides, G.A. and Hershberger, S.L. *Multivariate Statistical Methods: A First Course*. Psychology Press, 1997.

29. Morris, J.S. Slanted Objectivity? Perceived Media Bias, Cable News Exposure, and Political Attitudes. *Social Science Quarterly 88*, 3 (2007), 707–728.

30. Nagelkerke, N.J.D. A Note on a General Definition of the Coefficient of Determination. *Biometrika 78*, 3 (1991), 691–692.

31. Preece, J. and Shneiderman, B. The Reader-to-Leader Framework: Motivating Technology-Mediated Social Participation. *AIS Transactions on Human-Computer Interaction 1*, 1 (2009), 13–32.

32. Rainie, L. and Smith, A. *Pew Internet Social networking sites and politics*. Pew Research Center's Internet & American Life Project, 2012.

33. Romero, D.M., Meeder, B., and Kleinberg, J. Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. *Proceedings of the 20th international conference on World wide web*, ACM (2011), 695–704.

34. Schwab, A.J. Multinomial Logistic Regression: A Problem in Personnel Classification. http://www.utexas.edu/courses/schwab/sw388r7/Tutorials/RulonMulitNomialLogisticRegressionProblem_doc_html/026_The_Classification_Matrices_as_a_Measure_of_Model_Accuracy.html.

35. Sears, D.O. and Preedman, J.L. Selective Exposure to Information: A Critical Review. *Public Opinion Quarterly 31*, 2 (1967), 194–213.

36. Shah, D.V., Watts, M.D., Domke, D., Fan, D.P., and Fibison, M. News Coverage, Economic Cues, and the Public's Presidential Preferences, 1984-1996. *Journal of Politics 61*, 4 (1999), 914.

37. Shamma, D.A., Kennedy, L., and Churchill, E.F. Tweet the debates: understanding community annotation of uncollected sources. *Proceedings of the first SIGMM workshop on Social media*, ACM (2009), 3–10.

38. Simon, H.A. Bounded Rationality. In J. Eatwell, M. Milgate and P. Newman, eds., *Utility and Probability*. W. W. Norton & Company, 1990, 15–18.

39. Starbird, K., Palen, L., Hughes, A.L., and Vieweg, S. Chatter on the red: what hazards threat reveals about the social life of microblogged information. *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, ACM (2010), 241–250.

40. Sunstein, C.R. *Republic.com 2.0*. Princeton University Press, 2009.

41. Sweeny, K., Melnyk, D., Miller, W., and Shepperd, J.A. Information avoidance: Who, what, when, and why. *Review of General Psychology 14*, 4 (2010), 340–353.

42. Tsagkias, M., De Rijke, M., and Weerkamp, W. Linking online news and social media. *Proceedings of the fourth ACM international conference on Web search and data mining*, ACM (2011), 565–574.

43. Vieweg, S., Hughes, A.L., Starbird, K., and Palen, L. Microblogging during two natural hazards events: what twitter may contribute to situational awareness. *Proceedings of the 28th international conference on Human factors in computing systems*, ACM (2010), 1079–1088.

44. Yang, Z., Guo, J., Cai, K., et al. Understanding retweeting behaviors in social networks. *Proceedings of the 19th ACM international conference on Information and knowledge management*, ACM (2010), 1633–1636.

45. Young, I.T. Proof without prejudice: use of the Kolmogorov-Smirnov test for the analysis of histograms from flow systems and other sources. *Journal of Histochemistry & Cytochemistry 25*, 7 (1977), 935–941.

46. *Ideological News Sources: Who Watches and Why - Americans Spending More Time Following the News.* The Pew Research Center for the People & the Press, 2010.