Title

Communicating with Algorithms: A Transfer Entropy Analysis of Emotions-based Escapes from Online Echo Chambers

Journal

Communication Methods and Measures, 12(4)

ISSN

1931-2458

Authors

Hilbert, Martin
Ahmed, Saifuddin
Cho, Jaeho
et al.

Publication Date

2018-10-02

DOI

10.1080/19312458.2018.1479843

Peer reviewed

Communicating with algorithms: A transfer entropy analysis of emotions-based escapes from online echo chambers

June 2018

Martin Hilbert[*], Saifuddin Ahmed[*], Jaeho Cho[*], Billy Liu[+], Jonathan Luu[+]

[*] Department of Communication, University of California, Davis
[+] Computer Science, University of California, Davis

Author's version, published in (and please cite as):

Hilbert, M., Ahmed, S., Cho, J., Liu, B., & Luu, J. (2018). Communicating with Algorithms: A Transfer Entropy Analysis of Emotions-based Escapes from Online Echo Chambers. *Communication Methods and Measures*, 0(0), 1–16. https://doi.org/10.1080/19312458.2018.1479843

ABSTRACT:

Online algorithms have received much blame for polarizing emotions during the 2016 U.S. Presidential election. We use transfer entropy to measure directed information flows from human emotions to YouTube's video recommendation engine, and back, from recommended videos to users' emotions. We find that algorithmic recommendations communicate a statistically significant amount of positive and negative affect to humans. Joy is prevalent in emotional polarization, while sadness and fear play significant roles in emotional convergence. These findings can help to design more socially responsible algorithms by starting to focus on the emotional content of algorithmic recommendations. Employing a computational-experimental mixed method approach, the study serves as a demonstration of how the mathematical theory of communication can be used both to quantify human-machine communication, and to test hypotheses in the social sciences.

Communicating with algorithms: A transfer entropy analysis of emotions-based escapes from online echo chambers

Algorithms intermediate almost all of the roughly 3.5 hours per day that each American communicates online (Center for the Digital Future, 2016). Algorithms decide how fast information is transferred and what information is presented (Hannak et al., 2013; Lazer, 2015). The proactive role of algorithms in communication processes can be conceptualized as an active dialogue between users and algorithms. Users communicate preferences (advertently or not), which are interpreted by algorithms (e.g., by recommender engines or bots), which then send responses back to users, who receive, interpret, and react to the algorithmic reply. In this study, we quantify the flow of information in the conceptual communication channel between users and socially responsive algorithms.

Particularly, we first demonstrate that human emotions (joy, fear, sadness, etc.) are communicated to the selection mechanism of videos recommended by YouTube (via an individuals' choice of search terms), and second that the video's emotions influence how the viewer feels after being exposed to the videos. We study emotions related to Presidential candidates from the 2016 U.S. election. Rather than making assumptions about how recommender systems work, we systematically manipulate YouTube's real-world recommender system to create stimuli in a mixed-method computational-experimental approach.

Algorithmic Filter Bubbles and Echo Chambers

In the two-step communication between algorithms and humans it is common (but by no means obligatory) that the algorithm takes on the role of a confirmatory communication partner that reassures and often reinforces received information through positive feedback. This leads to the notorious "filter bubbles" (Pariser, 2011). The separation of users from contradictory information also gathers likeminded people in similar communication spaces, which then creates reinforcing echo chambers (Jamieson & Cappella, 2008; Sunstein, 2001). The resulting dynamic is nowadays widespread in online networks (Colleoni, Rozza, & Arvidsson, 2014; Garrett, 2009a), and has resulted in detectable tendencies of opinion extremism and political polarization (Bakshy, Messing, & Adamic, 2015; Bessi et al., 2016).

The prevalence of this reinforcing communication tendency of algorithms is mainly due to two reasons. On the one hand, people tend to prefer confirmatory and harmonious communication over confrontational and critical exchanges (Iyengar & Hahn, 2009; Mutz, 2006). Exploiting this keeps users engaged with the offered service and therefore maximizes the resulting economic profit. For example, YouTube explicitly works with what it calls a watch time optimization algorithm (YouTube, 2016b). On the other hand, self-reinforcing responses are much easier to program than any other kind of dialectical or critical response. This is simply because there are relatively few ways to agree with somebody, but infinite ways to disagree. Mathematically, while

there are relatively few ways to reinforce the direction of an identified vector in a multidimensional vector space, there are innumerable ways of not doing so. This makes this issue currently a hot topic for the technology and social media industry (Dreyfuss, 2016; Jigsaw, 2016).

The Role of Emotions

This article contributes to this growing literature by investigating *which kind of algorithmic responses can help citizens burst the filter bubbles and escape online echo chambers*? We already know that confronting users with diametrically opposed extremes also lead to opinion polarization (Keegan, 2016, Sherif & Hovland, 1961, Wong, Levin, & Solon, 2016). This is unfortunate, because they are mathematically relatively easy to identify. At the same time, we also know that people do not necessarily avoid all and every kind of challenging viewpoints when online (Garrett, 2009b, 2009a; Garrett & Stroud, 2014; Gentzkow & Shapiro, 2011). The big question becomes how to choose opinions that are neither reinforcing nor diametrically opposed, but still both engaging and challenging.

In this article, we propose that the research for socially responsible algorithmic responses can be enlightened by paying attention to how emotions flow to and from algorithms. Emotions have long been linked to political preferences (Glaser & Salovey, 1998; Marcus, 2000), political conversation (Cho, 2013), and political participation (Valentino, Brader, Groenendyk, Gregorowicz, & Hutchings, 2011).

It has not only been shown that online communication is laden with emotions (Derks, Fischer, & Bos, 2008; Holyst, 2017; Vincent & Fortunati, 2009), but that emotions can be transferred through algorithmic choices (Kramer, Guillory, & Hancock, 2014). Furthermore, research has shown a strong link between emotions and the strengthening or attenuating of opinions and partisanship, and therefore polarization (Nabi, 2003). For example valence (positive and negative affect) has shown to be a strong indicator of homophilic clustering (Himelboim et al., 2016). Anger results in extremist responses (Abelson et al., 1982), in pro-attitudinal partisan and ideological beliefs (Hasell & Weeks, 2016; Weeks, 2015), and in unwillingness to engage in contrasting dialogues (Valenzuela & Bachmann, 2015).

Our methods aim at quantifying the contagious effect of emotions from users to personalized online content, and then from algorithmically recommended online content back to users. Our analysis demonstrates the reinforcing effect of human-algorithm convergence or polarization of emotional states.

Method

We use a combination of experimental and computational methods to obtain the required variables to flesh out the information flows in our human-algorithm communication channel. We started by using a deep neural network for semantic analysis to code the general emotions of recommended videos on YouTube on the day of our experiment. In parallel we asked participants about their political search preferences, which we used to bias the recommendations of a virgin YouTube account per participant. Participants were then invited into the lab to complete a questionnaire about the emotions they felt about each of the two major-party candidates. Next, they watched the top-5 recommended videos from the account that was biased with their political preferences. We also did a semantic analysis on the emotions of those videos. Finally, participants repeated the candidate-emotion questionnaire to capture any resulting change. This gave us four distinct evaluations of emotional states: videos pre-intervention (baseline), user pre-intervention, videos post-intervention (recommendations), users post-intervention. We quantify the involved information flow of emotions from users to recommendations and from recommendations to users with transfer entropy (Schreiber, 2000). It is a directed measure of influence from the toolbox that grew out of Shannon's "mathematical theory of communication" (Shannon, 1948, p. 379). We feel that this is framework provides the natural choice to model information flows in communication channels such as this one. Our analysis focuses on differences in the transfer of emotional information among users of different partisanship, candidate preferences, ideology, individual and social influences, and in the polarization or convergence of their emotions.

Participants

Participants in this study included a total of 73 upper-level undergraduate students from a large university in the western United States. All participants in this study were volunteers and were awarded extra credit for their participation (average 3rd year of a bachelor degree, 22 years old). 67% were female, 23% only white, and 51% only Asian.

Materials

Each participant took a pre-experiment online survey that contained a list of nine campaign statements copied from each of the two candidates' official campaign websites (Hillary Clinton and Donald Trump). These were statements like "Making college debt free and reducing student debt" and "Build a wall against illegal immigration at the Mexican border". We added four related search terms for each candidate identified by Google Trend, such as "Trump lies" and "Lock her up". This gave us a list of 13 statements per candidate, which we mixed randomly in each survey. In this pre-survey, we asked each participant to rank this list of statements twice, according to two different criteria.

First, we asked everybody: “If you would search online for specific topics of the 2016 Presidential election that interest you personally, what would these topics be? Please rank 10 (*out of 26*) of the following topics according to your personal interest.” We categorize these responses as “individual” interests.

We then repeated the same list of 26 statements and asked: “When online (at social media, email, etc.), which of the following topics are you likely to see posted or recommended by your friends and online circles? Concentrate on the top 10 (*out of 26*) statements that could come from one of your online contacts”. We categorize these responses as “social” interests.

We then chose to experimentally contrast both interests because surveys have shown that users obtain their online input from both individual search results and recommendations from their (more or less algorithmically mediated) online friends (Bakshy et al., 2015; Gottfried & Shearer, 2016). In other words, people click on and consume information that shows up as the result of their own search, as well as information that is presented to them as the result of the interests of friends.

We used these pre-survey responses to train a YouTube recommender system (detailed in the section: Pre-Experiment Algorithmic Biasing) and then exposed participants in two different experimental conditions to the two-different sets (“individual” and “social”) of YouTube biased videos (detailed in the section: Design and Procedure).

Pre-Experiment Algorithmic Biasing

We chose YouTube’s video recommender engine as our subject of study. In the words of Google engineers “YouTube represents one of the largest scale and most sophisticated industrial recommendation systems in existence” (Covington, Adams, & Sargin, 2016, p. 191). During the time of this study in 2016, this video-sharing website was the 2nd most visited page on the Internet (Alexa, 2017). It is used by almost every third internet user and every seventh person on Earth (over 1 billion users). YouTube reaches more 18-49 year-olds than any cable network in the U.S. (YouTube, 2016a). One in five YouTube users get news from it, which makes YouTube the second largest social media news provider in the U.S. (after Facebook) (Gottfried & Shearer, 2016). Polarization has been shown to result from YouTube content (Bessi et al., 2016).

We used the selected campaign statements from the pre-survey to bias YouTube’s recommendation engine without the knowledge of the participants. It is important to work with the actual online algorithm that creates personalized results, not with proxy. Today’s online algorithms are essentially black boxes. While they are still deterministic, their complexity and collective dynamics make their outcomes difficult to predict (Lazer, 2015). Some functionality may be exerted intentionally, while other aspects might be incidental (Diakopoulos, 2015).

Despite their obfuscatory inner workings, per definition, any algorithm must always have an input and output, which offers the two instances in which we evaluate their communicative role.

We start by collecting the top-five trending videos on the general YouTube site (youtube.com/feed/trending). These videos are recommended based on general trends across YouTube. We will use this as a control variable.

For the biasing of accounts, we used a combinatorial logic to combine the top seven selected statements from the pre-survey of each participant into 70 search terms (top seven search terms + 21 pairwise combinations + 35 randomly selected three-part combinations + top seven search terms in reverse order). Our goal was to work with the strongest opinions of the subject, but also with a long enough list of search term to affect YouTube's recommender engine (we ran tests for its adaptive sensitivity).

We wrote a Python script in PyCharm that worked with a Chrome browser extension to bias one virgin YouTube accounts per search term list. This allowed us to bias YouTube accounts with users' preferences without the explicit knowledge of the users. Once started, the script logged into a clean YouTube account, then took the first term of the list of 70 and searched YouTube. The script opened the first recommended video and watched it for seven seconds in order for it to get adopted in YouTube's watch history. It justifies to click on the first recommended video as it has been shown that the highest ranked search results are exponentially more likely to be clicked than lower ranked links (Bakshy et al., 2015). It is also important to watch the video, as we found that YouTube's recommendation algorithm works on the basis of the watch history, not on basis of the search history. We speculate that the reason for this is once again the fact that the final consumption of online content is a mix of own search results and input from their online friends (Gottfried & Shearer, 2016). The fact that YouTube seems to consider the possibility of both influences is another justification to investigate both sources. The script then scraped the title, description, and the (often automatically created) transcript of the video (which is available for more than two-thirds of the videos).

After doing the same for the rest of the 70 search terms, the script scraped the recommended videos presented at youtube.com/feed/recommended. From this list, we selected the top-five recommended videos. In case there were extremely unpolitical videos within those, we maximally skipped two unpolitical videos. This gave us a list of five YouTube videos per participants, based either on their own preferences ("individual") or based on what they would expect to see from their social environment ("social").

Design and Procedure

The experiment was carried out the week after the 2016 U.S. Presidential election, during November 14 – 18. Participants were assigned to two experimental conditions: to a group that watched YouTube videos biased through their "individual" list of interests ($N = 35$) and a second group that watched YouTube videos biased through their "social" list of interests ($N = 38$).

At the first step, irrespective of the experimental condition, the participants filled out a pre-survey with a set of questions that asked participants about their demographics and their feelings toward each Presidential candidate. It asked them to rate positiveness and negativeness (valence) on a feeling thermometer scale between 0 and 100, and presented them with a five-point scale to evaluate the frequency they felt anger, fear, disgust, joy and sadness toward each candidate: never, some of the time, about half the time, most of the time, always.

At the next step, based on their respective experimental condition ("individual" vs "social"), we had the participants watch five algorithmically biased YouTube recommended videos for a total of 20 minutes, advising them to spend some 3-5 minutes on each video. These videos were linked to a new YouTube account we created for each user. At the final step, we asked them to fill out a post-survey with a total of 21 questions that included the same questions about their feelings toward the Presidential candidates.

After the surveys were completed and turned in, the researchers asked the participants if they had any questions and they were thanked for their participation.

Analytical Strategies and Measures

Over the second half of the last century, Shannon's (1948, p. 379) "mathematical theory of communication" has developed into a comprehensive theory known as information theory, which not only consists of hundreds of theorems and proofs (Cover & Thomas, 2006; MacKay, 2003), but also provides the theoretical basis of today's omnipresent communication networks (Gleick, 2011). While social scientists have struggled with finding value in its application to social systems (with notable exceptions, Attneave, 1959; Ellis & Fisher, 1975; Hawes & Foley, 1973; Krippendorff, 2008), recent studies have shown the usefulness of applying information theoretic measures like transfer entropy to social science and social media data (Baek, Jung, Kwon, & Moon, 2005; Borge-Holthoefer et al., 2016; Ver Steeg & Galstyan, 2012). In the following sub-sections we discuss how the "mathematical theory of communication" guides our analytical strategies in answering the research question.

The human-machine communication channel. Figure 1 uses the traditional representation of information theory (Cover & Thomas, 2006, Chapter 7) to outline the complete communication channel, consisting of three sub-channels. Each rectangular box represents a random variable that codifies an emotion into different categorical values (i.e. low vs. high). The

transitions between them (gray lines) represent mutual dependence, expressed through the joint distribution of the variables on the left and right of the connection. Information theory quantifies the information flow through such channels in terms of the intertwindness of the joint distributions among these variables.

We start our analysis with the benchmark of unbiased trending videos presented by YouTube's trending algorithm ($algo_{pre}$). This reflects the general mood prevalent on YouTube at the time. In our analysis, we will use this as a baseline control variable. We then measure the user's opinion before the intervention ($user_{pre}$). This is collected through the pre-survey on a five-point scale that evaluates the frequency students felt anger, fear, disgust, joy and sadness toward each candidate. The first sub-channel conceptualizes the relation between the general mood of trending videos on YouTube, and the feelings of the human user about each candidate. Based on the human search term preferences, the recommendation algorithm presents selected videos with certain emotional content ($algo_{post}$). The relation between the participants' emotions and the emotional content of the videos represents the second sub-channel. Through the consumption of the video, the posterior feelings of the human might be affected ($user_{post}$, again collected through the post survey on a five-point scale). The third sub-channel represents the information flowing from the emotions of the video to the emotions of the consumer. Once the overall channel is represented in such fashion, it is straightforward to analyze it with the well-established information theoretic tools for noisy channels.

Figure 1: Schematic representation of the communication channel with its three sub-channels

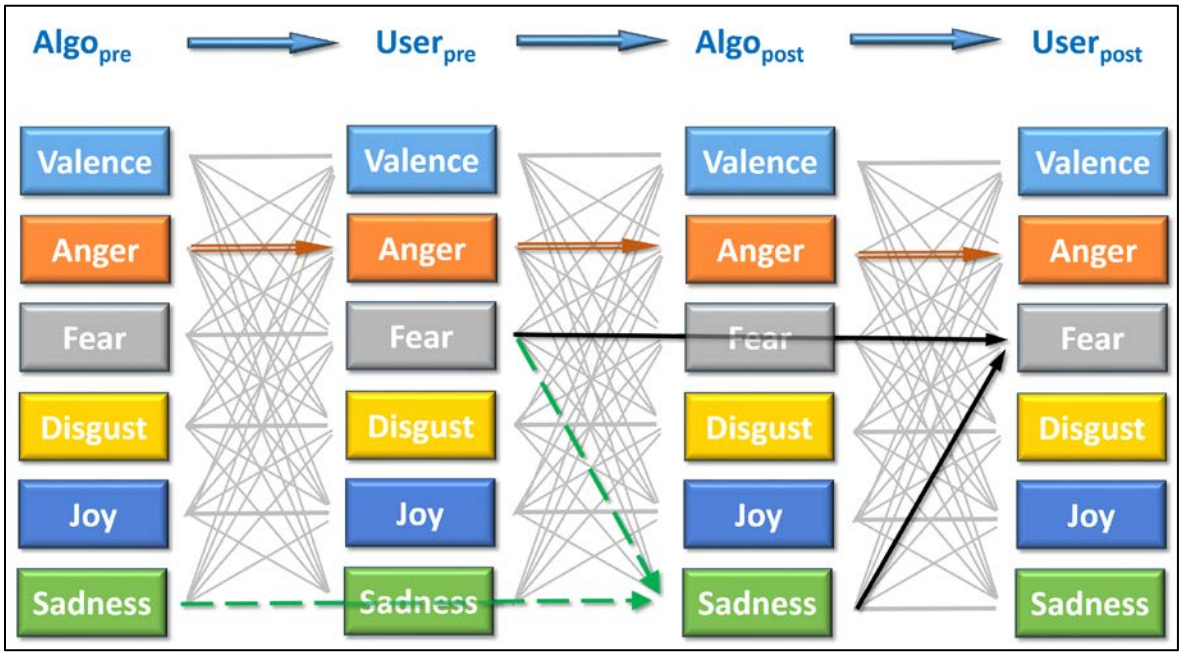


Figure 1 also depicts three illustrative toy examples that help to expose the nature of the involved information flows. The first one is illustrated with double-lined arrows and suggests that information flows from a general feeling of anger in society (resulting in angry trending videos on YouTube), to a particular angry user, which leads to video recommendation with angry content, and an angry user after consuming the video. This is the classic case of noiseless communication in a communication channel (in this case, over three sub-channels). It is also the standard assumption of how emotions flow in political communication, as it is for example both assumed and found that "positive moods induce more positive judgments and negative moods induce more negative judgements" (Marcus, 2000, p. 230).

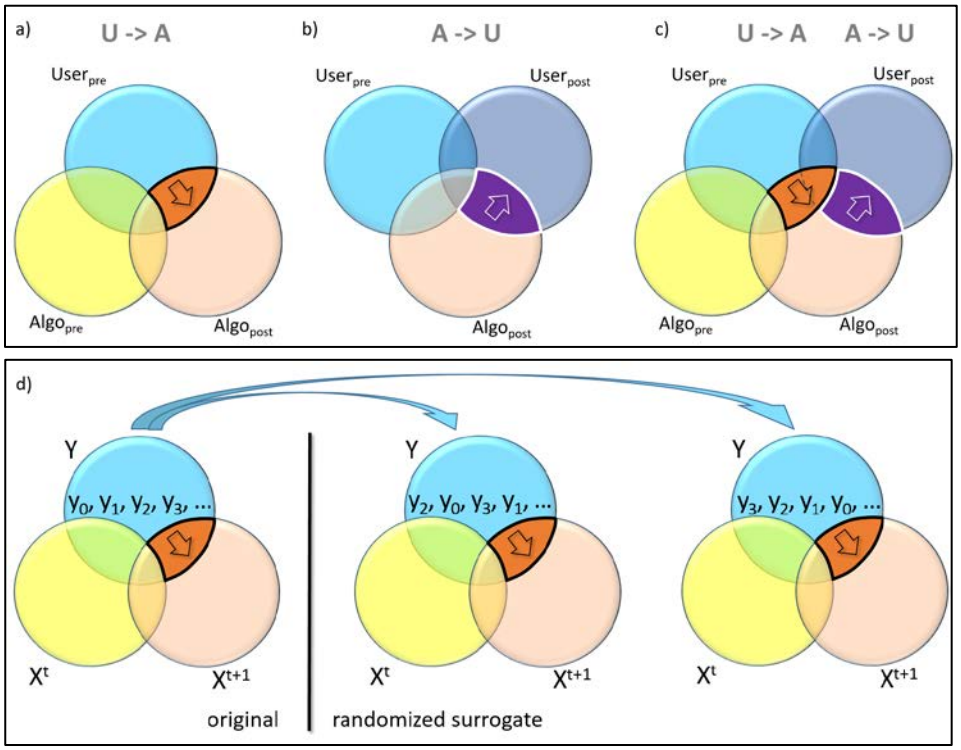
However, in a 'noisy channel' (this is the technical term for a channel that has not only 'straight transitions', but also 'cross-overs'), information also flows by crossing among different variables. The second case (illustrated with dashed arrows) also detects a noiseless transition from sad general videos to sad personalized video recommendations, but, for some reason, fear in a user also leads the recommender system to suggest sad videos. It would therefore be helpful to separate how much information flows through the video's autocorrelation (from sad trend to sad recommendation) and how much of the emotional information contained in the recommended videos is caused exclusively by the user's emotions. The same logic applies to the autocorrelation between pre- and post-opinions of the users, also shown in the third case in Figure 1. This aims at isolating the information transfer from algorithm to the user, while controlling for the fact that initially fearful users will continue to be fearful. Transfer entropy was developed to evaluate such causal effects in communication channels.

Transfer entropy. Transfer entropy is a conditional measure of dependence in information theory. Information theory defines information in terms of uncertainty and frames uncertainty in terms of probability theory (Cover & Thomas, 2006; MacKay, 2003). The less uncertainty, the more information and vice versa. The two basic measures of information theory, entropy and mutual information, have an analogous relation to the more well-known measures of variance and covariance (e.g., Garner & McGill, 1956). Variance and entropy both measure diversity, while covariance and mutual information both measure association (Li, 1990). Mutual information measures the uncertainty of one variable contained in another, or, in other words, the uncertainty resolved about a second variable when knowing the first.

Just like a covariance or correlation, mutual information does not imply causality. What one variable reveals about another is mutual, which makes it a symmetrical measure. Transfer entropy is a conditional mutual information (Schreiber, 2000). The conditional control for a time-delay among three variables introduces Markovian shielding and therefore directionality through temporal delay. This is often depicted with the help of Venn-diagrams (James, Ellison, & Crutchfield, 2011; Yeung, 1991), where the circles represent entropies and their intersection their mutual information (Figure 2). For example, Figure 2a highlights the transfer entropy that flows from user preferences U_{pre} into the algorithmic video recommendations A_{post}, while

controlling for the influence of the general mood on the platform, A_{pre}. The variables A_{pre} measures the emotional content of recommended videos before personalization (i.e. trending videos), and A_{post} the emotional content of the recommended videos after personalization. A_{pre} surely has an innate effect on the resulting A_{post}, which would confound our interest in measuring the information flow from user to algorithm. Transfer entropy controls for the autocorrelation between both time-delayed measures, and isolates the mutual information between the user, U_{pre}, and A_{post} (quantified by the respective light-orange overlaps in the diagram).

Figure 2: information diagram representation of both involved transfer entropies: (a) U => A, (b) A => U, (c) U => A and A => U, (d) significance testing of the transfer entropy measurement by generating a number of surrogate source datasets drawn from the original cause.



Transfer entropy can be viewed as a non-parametric equivalent for the more well-known Granger causality (Amblard & Michel, 2011), with the difference that it naturally also works for nonlinear categorical variables. Loosely speaking, what Granger causality is to correlations, transfer entropy is to mutual information. Like Granger causality, if the future values of a variable A contain information about the past of another variable U, that were not contained in past observations of A, then it is said that information is transferred from U to A. In principle, the mutual information between both is symmetric (undirected), but the experimentally introduced time delay allows to establish directionality.

In practice, one convenient way of calculating transfer entropy is to take the difference between conditional entropies, H. In particular, the conditional entropy of A_{post}, conditioned on both A_{pre} and U_{pre}, and the entropy of A_{post} conditioned on A_{pre}:

$$T_{U_{pre} \to A_{post}} = H(A_{post} \mid A_{pre}) - H(A_{post} \mid A_{pre}, U_{pre})$$
$$= \sum p(a_{post}, a_{pre}, u_{pre}) * log_2 \frac{p(a_{post} \mid a_{pre}, u_{pre})}{p(a_{post} \mid a_{pre})}$$

where the base of the logarithm defines the informational unit, in this case, bits. The formula applies equivalently for the information transfer from the recommended video A_{post} to the post-intervention feelings of the user U_{post}: $T_{A_{post} \to U_{post}}$ (Figure 2b). Here we control for the autocorrelation effect of the human emotions pre-intervention, isolating the information flow from the algorithmic recommendation to the user's emotions post intervention. Taken together, Figure 2c visualizes the information transfer from user preferences to algorithmic recommendations and from there back to the user.

Semantic analysis. Categorical emotion analysis typically yields half a dozen of different basic emotions (Ortony, Clore, & Collins, 1990), usually including the big five emotions of anger, fear, disgust, joy, and sadness (Ekman, Sorenson, & Friesen, 1969; Philippot, 1993). The dimensional view of emotions takes a complementary view (Bucy, 2000), usually including emotional valence (positive and negative affect) and arousal (Lang, 1988). We separately test for the big five emotions and additionally for emotional valence.

We used AlchemyLanguage from the IBM Watson Developer Cloud (now called "Watson Natural Language Understanding") to execute a sentiment analysis that evaluated the feelings attached to the videos based on the scraped title, description, and transcript. AlchemyLanguage is a collection of APIs that offer text analysis through natural language processing. Before it was acquired by IBM in 2015, it was known as AlchemyAPI, a deep learning machine learning tool for natural language processing (specifically, semantic text analysis, including sentiment analysis). It evaluates positiveness and negativeness (valence) on a scale from -1 to +1, and assigns values between 0 and 1 to the presence of anger, fear, disgust, joy, sadness (both to the third digit).

Before working with this tool, we followed the advice to validate the automatic content analysis methods (Grimmer & Stewart, 2013), and asked 91 students (that did not participate in our experimental subjects) to watch eight trending YouTube videos and to classify the emotions contained in the video on a scale from 0 to 100 for extra credit of a course. Both the correlation coefficients and the Mean Absolute Error (MAE) place the results on the cutting edge of semantic emotions detection: valence (*R*: 0.87, *MAE*: 0.36); anger (*R*: 0.40, *MAE*: 0.18); disgust (*R*: 0.55, *MAE*: 0.16); fear (*R*: 0.23, *MAE*: 0.21); joy (*R*: 0.49, *MAE*: 0.22); sadness (*R*: -0.51,

MAE: 0.20) (compare with Paltoglou, Theunis, Kappas, & Thelwall, 2013; Thelwall, Buckley, Paltoglou, Cai, & Kappas, 2010).[1]

It turned out that the total of 365 videos watched by our 73 participants consisted of 121 different videos. 68% of these contained a written transcript. We evaluate the feelings separately for the title, the description and the transcript (if available). After some preliminary testing, we decided to create the simple average of these values to obtain the final estimation for each video (either with or without a transcript). This mainly aims at balancing the uneven weight of each (the text of transcripts would dominate if evaluated jointly) and to give more visible roles of the title and description. The total 365 videos had the following average emotional scores: postiveness/negativeness (*M* = 0.007; *SD* = 0.352), anger (*M* = 0.266; *SD* = 0.150), fear (*M* = 0.151; *SD* = 0.094), disgust (*M* = 0.177; SD = 0.096), joy (*M* = 0.215, *SD* = 0.160), sadness (*M* = 0.266, *SD* = 0.112). For participants' emotions, we created averages of the values of the five videos watched by each participant, giving us one score per feeling per participant.

Dichotomization. To calculate the information theoretic measures, we need to convert our data into normalized probability distributions of a categorical random variable. Our measures of emotions are already normalized (survey on a five-point scale, and the computational semantic analysis between -1 and +1, and 0 and 1). We convert all our emotion measures into a binary variable. We assign 0s to values below the variable's arithmetic mean (low on this emotion), and 1s to values above it (high on this emotion), and then count the frequency of their appearance. The reason for this high level of coarse-graining is the inherent trade-off between measurement detail and sample size. Our small sample size forces us to work with a simple binary distinction.

Different conditions. In our analysis, we condition the binary emotional variables on different conditions (i.e. pre- vs. post-intervention; individual vs. social recommendations; Clinton vs. Trump; Liberal vs. Conservatives). These are straightforward subgroups of our samples. Our conditioning variable of emotional polarization requires some additional elaboration. We calculate it for each feeling separately based on each pre- and post-survey evaluation of each participant. We identify polarization if emotional strength moves away from its pre-intervention mean ('toward the poles'), and convergence if it moves closer toward it ('away from the poles toward the mean'). While there are several variables that satisfy this definition, the simplest one consists of taking the absolute value of the pre-post mean difference:

$$\text{Polarization score: } |\Delta POST| - |\Delta PRE|, \text{ with}$$

$$|\Delta POST| = \text{ABS}\left(score_{post} - \text{E}\left[score_{pre}\right]\right)$$

$$|\Delta PRE| = \text{ABS}\left(score_{pre} - \text{E}\left[score_{pre}\right]\right)$$

[1] The negativity of sadness is not worrisome for our purposes, since the strength of information flow is measured based on symmetric distributions.

where the expected value $E[...]$ is taken over all pre-intervention scores of this particular feeling score. This implies polarization away from the mean if the score is positive; convergence toward the mean if it is negative; and no change in emotions if it is zero.

Statistical Test

We need to make sure that our results are not mere artifacts that would arise in any case due to random fluctuations. Since no parametric distribution of errors is known for the nonlinear measure of transfer entropy, suitable surrogate data is needed to test the null hypothesis of independent time series and therefore an absence of causality. This suggests creating statistical ensembles of a randomization of the same data (i.e., by drawing randomly from the original distribution), such that the causal dependency of interest is destroyed, but trivial dependencies of no interest are preserved (see Figure 2d). This means that if causality is suggested such that Y => X^{t+1}, then only Y is randomized. The rationale is to destroy the causal structure of Y on future values of X^{t+1}, under the null hypothesis H_0 that the changes X^t => X^{t+1} have no temporal dependence on the potentially causing source Y. This procedure can be seen as a bootstrap under the null hypothesis because any dependence is eliminated. This method was originally extended from a similar logic applied to Granger causality (Chávez, Martinerie, & Le Van Quyen, 2003; Verdes, 2005) and we adopt it from its successful application in neuroscience research (Lizier, Heinzle, Horstmann, Haynes, & Prokopenko, 2011; Vicente, Wibral, Lindner, & Pipa, 2011).

We can then determine a one-sided p-value of the probability of observing a transfer entropy value greater than the ones expected assuming H_0. This can be done by simply counting the proportion of observed information flows that produce larger transfer entropy than the randomized one. In other words, a 'significant information flow' means that it is likely that the observed amount of transferred information is larger than the information that could be expected to flow due to random chance. For each case, we calculate the transfer entropy of 500 surrogate distributions, and given the stringency of our test, distinguish among three significance levels: $*p < .1$; $**p < .05$; $***p < .01$.

Note that given that we create the surrogate data for our null hypothesis with as little of a disturbance of the original distribution as possible to create the independence condition, large information flows must not automatically be statistically significantly larger than randomly expected. This depends on the skewness of both involved distributions and the remaining degrees of freedom to create a joint distribution. We will find this subtle point in several of our results (i.e., it is clearly shown in the case of convergence for conservative subjects in Figure 5).

Results

We calculate the transfer entropies for each of our six emotions for both sub-channels: U => A and A => U. We illustrate the different amounts of information in a way that adds up the informational bits transmitted by different emotions. This is a choice of representation (as we calculate the information transfer and its significance for each emotion separately), but psychologically justified by the currently dominant multiple-channel theory of emotions, which presumes that affective reactions derive from multiple parallel evaluative processes in multiple dimensions (Marcus, 2000).

From Humans to Algorithms and Back

We start by evaluating the general emotional flow from the human preferences to the recommended videos, and from there, back to the final feelings of the consumer. For this, we add up the feeling scores of the pre- and post-surveys toward both candidates (Clinton and Trump) and then create a single binary distribution for each of the emotions. We calculate the resulting transfer entropy based on the combined sample of all 73 users.

Figure 3 shows that both the transfer entropy from user to algorithm (U => A) and from algorithm to user (A => U) has components that are significantly different than what random fluctuations would suggest. This suggests that users' emotions had a measureable influence on the algorithm and that the recommender algorithm had a measurable influence on the users' emotions. Figure 3 also suggests that the flow of information from user preferences to recommended videos seems smaller than the flow of information from recommended videos to user emotions. In contrast to the skewed emotional transfer from human to algorithms, the emotional information transmitted in the sub-channel from algorithms to human is also more equally distributed. Together both seems to suggest that humans are susceptible to more and more diverse emotions than algorithms. An alternative explanation might be that U => A is a more indirect channel, since it is mediated by the search topics (campaign promises). For the information flow of U => A, only joy and sadness are statistically significantly larger than expected (total of some .08 bits), while in the flow of A => U only valence causes a statistically significant amount of emotional information (some .05 bits).

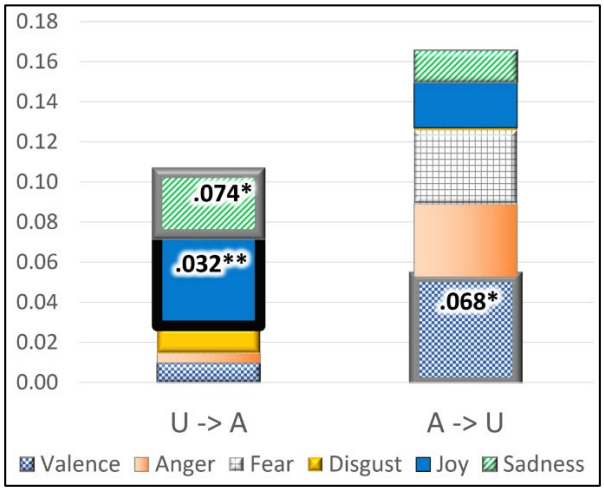


Figure 3: Transfer entropy of emotional information from human to algorithmically recommended videos, and back, from the algorithmic result to the human user, in bits. Numerical labels display significance test results. ** $p < .5$, * $p < .1$, no label $p > .1$.

Individual and Social Recommendations

We then separated both experimental groups, and distinguished between the recommendations that stem from one's individual preferences and the recommendations that result from preferences of the perceived social environment. We randomly chose a group of subjects for which we biased accounts with the individual list of interests, and another one with the hints of social influence. Figure 4a shows that recommendations produced by one's individual search terms leads to some significant information flows, while perceived social preferences do not. Despite all homophily, this is to be expected, as one's own direct preferences are more likely to trigger personally relevant emotions in videos than the indirect preferences of one's friend. For one's individual preferences, we again find joy and sadness to be significantly larger as expected, as well as anger.[2] Valence is again significant for the flow from algorithmic recommendations to users (joy gets close with $p = .108$).

Figure 4: Transfer entropy of emotions over both sub-channels, in bits (a) distinguishing recommendations based on individual preferences and on social preferences of friends, (b) additionally distinguishing emotions toward candidates. ** p < .5, * p < .1, no label p > .1.

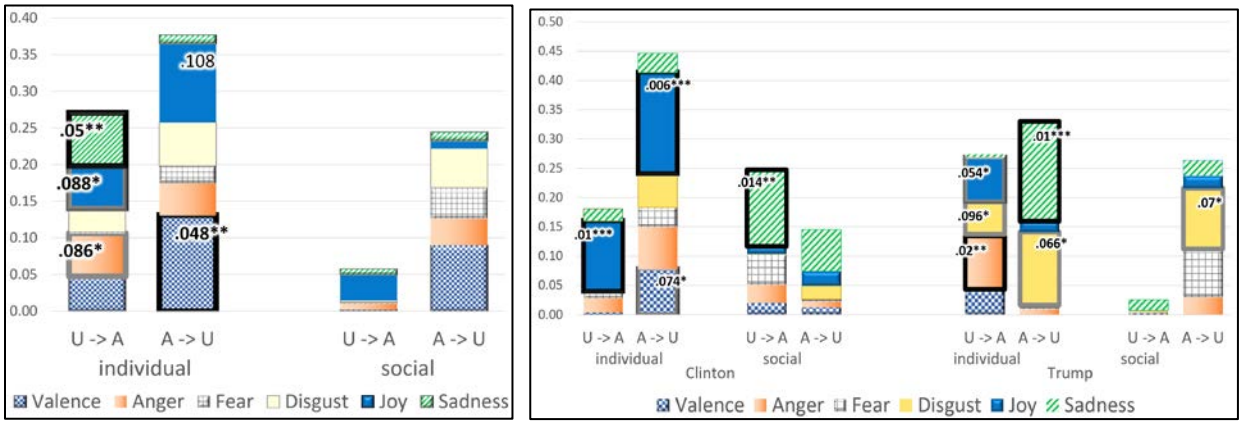


Figure 4b distinguishes among feelings toward each candidate individually, which provides more significant results, also for socially perceived input. This is likely to be a reflection of the fact that feelings from different people are more pronounced toward different candidates (while aggregation creates moderating effects). For the case of individual recommendations for Clinton, a highly significant amount of joy is flowing, both from users to algorithms and from algorithms to human users (the largest blocks in the two left bars). We observe a significant amount of sadness in the recommended videos from the perceived social environment of the candidate who lost the election (see third bar from left in Figure 4b). For the

[2] Comparing this with Figure 3 reminds us that informational measures do not decompose linearly among their constituents, since they are based on joint probability distributions that often exhibit nonlinear relations.

case of the election winner Trump, we find many significant emotional information transfers, including anger, disgust, and joy from humans to algorithms, and disgust and sadness from algorithms to human users.

Figure 4 also complements an established finding in the literature. Anger has often been linked to individual considerations, and fear to societal factors (Goodall, Slater, & Myers, 2013; Nabi, 2003). Both Figure 4a (for the case of 'individual'). Figure 4b (for the case of 'individual+Trump') reconfirm this finding as we find a significant amount of anger information flowing from the human to one's own recommendations, but not into social recommendations from friends. In both representations, we also find larger flows of fear in socially embedded friends' recommendations. But these are not significant in any of the cases.

Comparing Figures 4a and 4b also shows that despite all homophily with the social environment, emotional causations on basis of one's own individual preferences are different from the social preferences of perceived friends, especially when distinguishing between candidates. In the case of Clinton, joy plays a much more prevalent role in one's individual flows, while sadness dominates the social flow. For the case of Trump, the amount of information that flows from the perceived preference of friends to recommendations is quite small, especially when compared to the significant flows from one's own preferences to algorithmic recommendations.

The Emotions of Polarization

Based on our polarization score (see above), we now analyze the emotions that flow conditioned on the cases of emotional polarization and convergence.[3] We calculate the polarization score for each emotion separately, which means that the same person can be part of the polarization group for one emotion, but part of the convergence group for another emotion. We jointly analyze both the group biased with individual-, and the one biased with social influence (resulting in mixed effects on average, much like in real online environments), but distinguish between both candidates. Additionally, we also distinguish among users that lean liberal (extremely liberal, liberal, and slightly liberal, $N = 42$) and those that lean rather conservative (including moderate, slightly conservative, conservative, extremely conservative, $N = 25$), excluding ideological agnostics (6 of our 73 subjects). It justifies joining moderates with conservative-leaning subjects, since for one, the conservative candidate was originally an independent, and because in California there is a much higher barrier of categorizing oneself conservative than liberal.

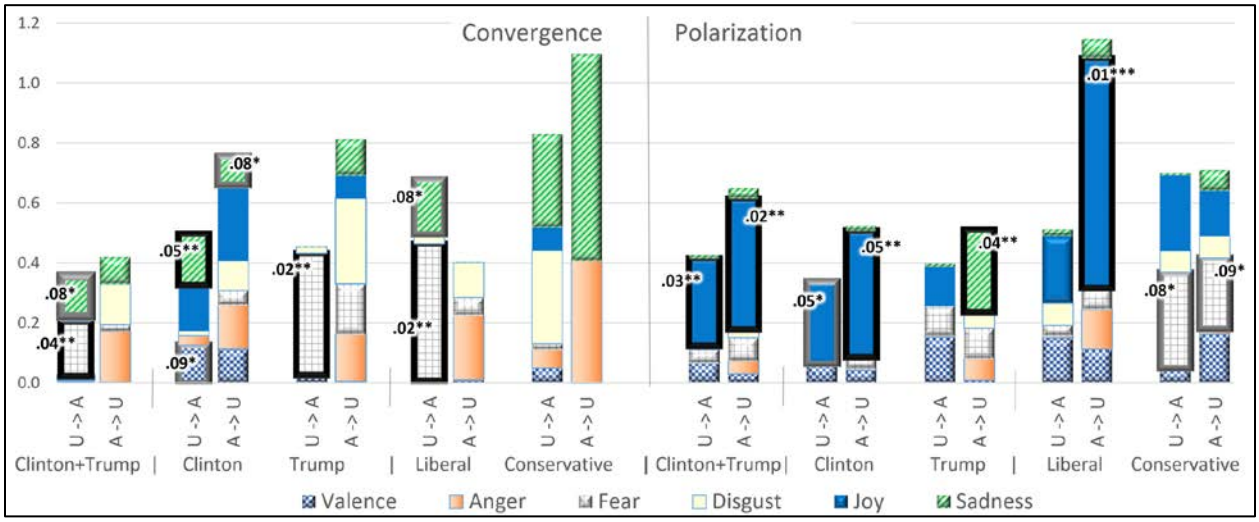
In the case of emotional convergence (left side in Figure 5), we find that a significantly large amount of both fear and sadness flow through our channels (four significant sadness and three significant fear values). In specific, a significant amount of fear flows from the preferences

[3] We do not have to consider the case in which there is no change in emotions (our polarization score is 0, since variables U_{pre} and U_{post} are identical). In this case, no information transfer is present.

of the human to cause changes in the emotional constellation of the algorithmic recommender system (see the squared bars U=>A). For the case of emotional polarization, we detect a significantly large flow of joy throughout both channels (five framed blue bars). Note that this can be the presence or absence of joy (see below). For the cases of conservative ideology and its candidate Trump, we detect a significant amount of sadness and fear (right bars in Figure 5).

While the literature has often detected the role of anger in polarization (Abelson et al., 1982; Hasell & Weeks, 2016; Valenzuela & Bachmann, 2015; Weeks, 2015), we do not find any significant flow of anger in the case of polarization, especially not when compared to the role of joy in polarization. Actually, we find proportionally larger amounts of angry emotions flowing in the case of convergence, but they are not statistically significant.

Figure 5: Transfer entropy of emotional information over both sub-channels for the case of emotional convergence and polarization, for the general case, conditioning on each candidate, and on liberal and conservative ideological leaning, in bits. ** p < .5, * p < .1, no label p > .1.



At this point, it is important to emphasize an often-confused fact about information measures. Measures from information theory are based on a probability distribution of categorical variables, while the different realizations of the involved random variables are not assigned any semantic meaning. This provides information theoretic measures its nonlinear and nonparametric strengths, but also leads to the fact that they do not differentiate between the content. For example, a large amount of information transfer related to disgust does actually not tell us if it is the presence or absence of disgust that contains this information. The magnitude of the resulting information flow does not depend on the decision if we encode the above average presence of disgust with 1 or with 0. The aggregate measure of the probability distribution of the random variable does not discriminate among the content the information stems from.

Discussion

Based on a mixed methods approach that includes both experimental and computational components, we quantified emotional information flows between humans and responsive algorithms. We explored how emotional information flows between algorithms and human based on various conditions, including partisanship, candidate preferences, ideology, individual and social influences, and emotional polarization. One of our findings was that content charged with joyful emotions leads to emotional polarization, while sadness and fear results in emotional convergence. This is a useful insight in the search for the design of socially more responsible recommender algorithms that aim at mitigating the polarization effects of today's filter bubbles.

An interesting question refers to the possibility to burst the filter bubble by supplementing one's individual preferences with preferences of one's friends. Some general indications of this relationship can be seen when comparing Figures 4 and 5. On the one hand, it shows that joy is a more prevalent cause of both polarization and one's individual preferences. On the other hand, we detect sadness in the cases of both social biasing and convergence for Clinton and in the cases of individual biasing and polarization for Trump. This suggests that there are some relations, but that any attempt of fortifying one's individual preferences with emotional preferences of friends might require conditional distinction (i.e. between Clinton and Trump supporters, especially the week after the election).

Three considerations arise from this finding. First, it might be that improved ways of categorization of one's social environment would help to find a clearer pattern. In our case, the preferences of the social environment stem from the subjects' imagination. This has the potential to capture characteristics of the social environment that are particularly relevant for the individual, but comes with selection bias. It would be worthwhile to explore the benefits and drawbacks further. Second, it might be that the well-established distinction between individual and social influences in online polarization (Bakshy et al., 2015) is not the right way out of the echo chamber. It might be that other approaches, such as our proposal to focus on the involved emotions, will turn out to be the more promising approach, independent if the source of influence is one's own past or one's social environment. Another provocative suggestion that arises from our results is the hypothesis that algorithms communicate different emotions to and from people of different political spectra. We found that algorithms pick up and recommend more extreme joy and valence (affect) among liberals and Clinton supporters, and more harsh emotions among conservatives and Trump supporters (disgust, anger, fear). Our results do not allow for a clear picture in this regard, but our presented methodology allows looking deeper into this issue.

There are at least four limitations of our study. First, we work with a limited distinction among individual differences. While we distinguished among political ideology and candidate preference, this study has for example not analyzed differences in psychological characteristics of users. Research in the field of message framing has shown that people's motivations are

subject to individual variations that will cause them to experience emotionally evocative message as a function of those individual differences (Yan, Dillard, & Shen, 2012). Second, while we controlled for the initial condition of videos on YouTube, we did include a separate control group in this analysis. Third, while candidate preferences tend to be fairly stable over the course of an election, emotions toward candidate were not measured at the same time that the statements used to generate search terms were selected. This assumes that the affective responses toward the candidates did not change between online survey and when people came into the lab. And last but not least, we derived the video emotions from scraped titles, descriptions, and transcripts, which facilitates computational semantic analysis. This allowed us to focus on our other methodological explorations, but might not be accurate. We recommend that studies that are more substantial use multimedia processing techniques or human coding to reduce the measurement error.

Finally, the study serves as a demonstration of how the mathematical theory of communication can be used both to quantify human-machine communication, and to test hypotheses in the social sciences. While we could have done a similar analysis with variance-based correlations and structural equations, we find the interpretation of our nonlinear entropy measure particularly intuitive. We started by conceptualizing the proactive role of algorithms in today's digital landscape as a communication channel between human a machine. The mathematical theory of communication provides a natural framework to measure information flows in communication channels with one single measure: bits. Of course, successful communication between A and B also creates a correlation between A and B, which is why traditional correlation based analysis would also work in our case. However, the information theoretic approach is more natural for the modeling of communication channels. While this approach of measuring communication channels was traditionally reserved for technological systems, its adoption for socio-technology systems is not only justified by the theoretical appropriateness, but also by the practical fast-paced human-machine merger that characterizes out modern communication landscape. As humans increasingly communicate with algorithms, it makes sense to conceptualize both with a common methodological framework.

References

Abelson, R., Kinder, D., Peters, M., & Fiske, S. (1982). Affective and Semantic Components in Political Person Perception. *Journal of Personality and Social Psychology*, *42*(4), 619–630. https://doi.org/10.1037//0022-3514.42.4.619

Alexa. (2017). Alexa Top 500 Global Sites. Retrieved February 28, 2017, from http://www.alexa.com/topsites

Amblard, P.-O., & Michel, O. J. J. (2011). On directed information theory and Granger causality graphs. *Journal of Computational Neuroscience*, *30*(1), 7–16. https://doi.org/10.1007/s10827-010-0231-x

Attneave, F. (1959). *Applications of information theory to psychology: A summary of basic concepts, methods, and results* (Vol. vii). Oxford, England: Henry Holt.

Baek, S. K., Jung, W.-S., Kwon, O., & Moon, H.-T. (2005). Transfer Entropy Analysis of the Stock Market. *ArXiv:Physics/0509014*. Retrieved from http://arxiv.org/abs/physics/0509014

Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, *348*(6239), 1130–1132. https://doi.org/10.1126/science.aaa1160

Bessi, A., Zollo, F., Vicario, M. D., Puliga, M., Scala, A., Caldarelli, G., … Quattrociocchi, W. (2016). Users Polarization on Facebook and Youtube. *PLOS ONE*, *11*(8), e0159641. https://doi.org/10.1371/journal.pone.0159641

Borge-Holthoefer, J., Perra, N., Gonçalves, B., González-Bailón, S., Arenas, A., Moreno, Y., & Vespignani, A. (2016). The dynamics of information-driven coordination phenomena: A transfer entropy analysis. *Science Advances*, *2*(4), e1501158. https://doi.org/10.1126/sciadv.1501158

Bucy, E. (2000). Emotional and Evaluative Consequences of Inappropriate Leader Displays. *Communication Research*, *27*(2), 194–226. https://doi.org/10.1177/009365000027002004

Center for the Digital Future. (2016). *The 2016 Digital Future Report, Surveying the Digital Future, Year Fourteen*. USC Annenberg School Center for the Digital Future. Retrieved from http://www.digitalcenter.org/wp-content/uploads/2013/06/2016-Digital-Future-Report.pdf

Chávez, M., Martinerie, J., & Le Van Quyen, M. (2003). Statistical assessment of nonlinear causality: application to epileptic EEG signals. *Journal of Neuroscience Methods*, *124*(2), 113–128.

Cho, J. (2013). Campaign Tone, Political Affect, and Communicative Engagement. *Journal of Communication*, *63*(6), 1130–1152. https://doi.org/10.1111/jcom.12064

Colleoni, E., Rozza, A., & Arvidsson, A. (2014). Echo Chamber or Public Sphere? Predicting Political Orientation and Measuring Political Homophily in Twitter Using Big Data. *Journal of Communication*, *64*(2), 317–332. https://doi.org/10.1111/jcom.12084

Cover, T. M., & Thomas, J. A. (2006). *Elements of Information Theory* (2nd Edition). Hoboken, NJ: Wiley-Interscience.

Covington, P., Adams, J., & Sargin, E. (2016). Deep Neural Networks for YouTube Recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems* (pp. 191–198). New York, NY, USA: ACM. https://doi.org/10.1145/2959100.2959190

Diakopoulos, N. (2015). Algorithmic Accountability. *Digital Journalism*, *3*(3), 398–415. https://doi.org/10.1080/21670811.2014.976411

Dreyfuss, E. (2016, November 19). Do-Gooder Technologists Are Trying to Burst the Post-Election Filter Bubble. *WIRED*, (Culture). Retrieved from https://www.wired.com/2016/11/coders-think-can-burst-filter-bubble-tech/

Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotion. *Science (New York, N.Y.)*, *164*(3875), 86–88.

Ellis, D. G., & Fisher, B. A. (1975). Phases of Conflict in Small Group Development: A Markov Analysis. *Human Communication Research*, *1*(3), 195–212. https://doi.org/10.1111/j.1468-2958.1975.tb00268.x

Garner, W., & McGill, W. (1956). The relation between information and variance analyses. *Psychometrika*, *21*(3), 219–228.

Garrett, R. K. (2009a). Echo chambers online?: Politically motivated selective exposure among Internet news users. *Journal of Computer-Mediated Communication*, *14*(2), 265–285. https://doi.org/10.1111/j.1083-6101.2009.01440.x

Garrett, R. K. (2009b). Politically Motivated Reinforcement Seeking: Reframing the Selective Exposure Debate. *Journal of Communication*, *59*(4), 676–699. https://doi.org/10.1111/j.1460-2466.2009.01452.x

Garrett, R. K., & Stroud, N. J. (2014). Partisan Paths to Exposure Diversity: Differences in Pro- and Counterattitudinal News Consumption. *Journal of Communication*, *64*(4), 680–701. https://doi.org/10.1111/jcom.12105

Gentzkow, M., & Shapiro, J. M. (2011). Ideological Segregation Online and Offline. *The Quarterly Journal of Economics*, *126*(4), 1799–1839. https://doi.org/10.1093/qje/qjr044

Glaser, J., & Salovey, P. (1998). Affect in Electoral Politics. *Personality and Social Psychology Review*, *2*(3), 156–172. https://doi.org/10.1207/s15327957pspr0203_1

Gleick, J. (2011). *The information: a history, a theory, a flood*. Random House Digital, Inc.

Goodall, C. E., Slater, M. D., & Myers, T. A. (2013). Fear and Anger Responses to Local News Coverage of Alcohol-Related Crimes, Accidents, and Injuries: Explaining News Effects on Policy Support Using a Representative Sample of Messages and People. *Journal of Communication*, *63*(2), 373–392. https://doi.org/10.1111/jcom.12020

Gottfried, J., & Shearer, E. (2016, May 26). News Use Across Social Media Platforms 2016. Retrieved December 23, 2016, from http://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/

Grimmer, J., & Stewart, B. M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, *21*(3), 267–297. https://doi.org/10.1093/pan/mps028

Hannak, A., Sapiezynski, P., Molavi Kakhki, A., Krishnamurthy, B., Lazer, D., Mislove, A., & Wilson, C. (2013). Measuring Personalization of Web Search. In *Proceedings of the 22Nd International Conference on World Wide Web* (pp. 527–538). Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee. Retrieved from http://dl.acm.org/citation.cfm?id=2488388.2488435

Hasell, A., & Weeks, B. E. (2016). Partisan Provocation: The Role of Partisan News Use and Emotional Responses in Political Information Sharing in Social Media. *Human Communication Research*, *42*(4), 641–661. https://doi.org/10.1111/hcre.12092

Hawes, L. C., & Foley, J. M. (1973). A Markov analysis of interview communication. *Speech Monographs*, *40*(3), 208–219. https://doi.org/10.1080/03637757309375798

Himelboim, I., Sweetser, K. D., Tinkham, S. F., Cameron, K., Danelo, M., & West, K. (2016). Valence-based homophily on Twitter: Network Analysis of Emotions and Political Talk in the 2012 Presidential Election. *New Media & Society*, *18*(7), 1382–1400. https://doi.org/10.1177/1461444814555096

Holyst, J. A. (Ed.). (2017). *Cyberemotions - Collective Emotions in Cyberspace*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-43639-5

Iyengar, S., & Hahn, K. S. (2009). Red Media, Blue Media: Evidence of Ideological Selectivity in Media Use. *Journal of Communication*, *59*(1), 19–39. https://doi.org/10.1111/j.1460-2466.2008.01402.x

James, R. G., Ellison, C. J., & Crutchfield, J. P. (2011). Anatomy of a bit: Information in a time series observation. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *21*(3), 037109. https://doi.org/10.1063/1.3637494

Jamieson, K. H., & Cappella, J. N. (2008). *Echo Chamber: Rush Limbaugh and the Conservative Media Establishment*. Oxford University Press.

Jigsaw. (2016, March 15). If you are reading this, we might be in the same news bubble. Retrieved February 24, 2017, from https://medium.com/jigsaw/if-you-are-reading-this-we-might-be-in-the-same-news-bubble-cb697270c698

Keegan, J. (2016). Blue Feed, Red Feed. Retrieved December 21, 2016, from http://graphics.wsj.com/blue-feed-red-feed/

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, *110*(15), 5802–5805. https://doi.org/10.1073/pnas.1218772110

Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, *111*(24), 8788–8790. https://doi.org/10.1073/pnas.1320040111

Kramer, G. (1998). *Directed information for channels with feedback* (Doctoral Theses under J.L. Massey). Zurich: Swiss Federal Institute of Technology. Retrieved from ftp://204.178.31.32/who/gkr/Papers/KramerThesis.ps.gz

Krippendorff, K. (2008). *On Communicating: Otherness, Meaning, and Information*. (F. Bermejo, Ed.). Routledge.

Lang, P. J. (1988). What are the data of emotion? In V. Hamilton, G. H. Bower, & N. C. Frijda (Eds.), *Cognitive perspectives on emotion and motivation* (pp. 173–191). Boston: Kluwer Academic Publishers.

Lazer, D. (2015). The rise of the social algorithm. *Science*, *348*(6239), 1090–1091. https://doi.org/10.1126/science.aab1422

Li, W. (1991). On the relationship between complexity and entropy for Markov chains and regular languages. In *Complex Systems*. Retrieved from http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.106.7537

Lizier, J. T., Heinzle, J., Horstmann, A., Haynes, J.-D., & Prokopenko, M. (2011). Multivariate information-theoretic measures reveal directed information structure and task relevant changes in fMRI connectivity. *Journal of Computational Neuroscience*, *30*(1), 85–107. https://doi.org/10.1007/s10827-010-0271-2

MacKay, D. J. C. (2003). *Information Theory, Inference and Learning Algorithms* (1 edition). Cambridge, UK ; New York: Cambridge University Press.

Marcus, G. E. (2000). Emotions in Politics. *Annual Review of Political Science*, *3*(1), 221–250. https://doi.org/10.1146/annurev.polisci.3.1.221

Massey, J. L. (1990). Causality, feedback and directed information. In *Proceedings of the 1990 Internantional Symposium on Information Theory and its Applications*. Waikiki, Hawai. Retrieved from http://www.isiweb.ee.ethz.ch/archive/massey_pub/pdf/BI532.pdf

Mutz, D. C. (2006). *Hearing the Other Side: Deliberative Versus Participatory Democracy*. Cambridge University Press.

Nabi, R. L. (2003). Exploring the Framing Effects of Emotion: Do Discrete Emotions Differentially Influence Information Accessibility, Information Seeking, and Policy Preference? *Communication Research*, *30*(2), 224–247. https://doi.org/10.1177/0093650202250881

Ortony, A., Clore, G. L., & Collins, A. (1990). *The Cognitive Structure of Emotions*. Cambridge University Press.

Paltoglou, G., Theunis, M., Kappas, A., & Thelwall, M. (2013). Predicting Emotional Responses to Long Informal Text. *IEEE Transactions on Affective Computing*, *4*(1), 106–115. https://doi.org/10.1109/T-AFFC.2012.26

Pariser, E. (2011). *The Filter Bubble: What the Internet Is Hiding from You*. Penguin.

Philippot, P. (1993). Inducing and assessing differentiated emotion-feeling states in the laboratory. *Cognition & Emotion*, *7*(2), 171–193. https://doi.org/10.1080/02699939308409183

Schreiber, T. (2000). Measuring Information Transfer. *Physical Review Letters*, *85*(2), 461–464. https://doi.org/10.1103/PhysRevLett.85.461

Shannon, C. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, *27*, 379–423, 623–656. https://doi.org/10.1145/584091.584093

Sherif, M., & Hovland, C. I. (1961). *Social judgment: Assimilation and contrast effects in communication and attitude change* (Vol. xii). Oxford, England: Yale Univer. Press.

Sunstein, C. R. (2001). *Echo Chambers: Bush v. Gore, Impeachment, and Beyond*. Princeton University Press. Retrieved from https://pup.princeton.edu/sunstein/echo.pdf

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, *61*(12), 2544–2558. https://doi.org/10.1002/asi.21416

Valentino, N. A., Brader, T., Groenendyk, E. W., Gregorowicz, K., & Hutchings, V. L. (2011). Election night's alright for fighting: The role of emotions in political participation. *Journal of Politics*, *73*(1), 156–170. https://doi.org/10.1017/S0022381610000939

Valenzuela, S., & Bachmann, I. (2015). Pride, Anger, and Cross-cutting Talk: A Three-Country Study of Emotions and Disagreement in Informal Political Discussions. *International Journal of Public Opinion Research*, edv040. https://doi.org/10.1093/ijpor/edv040

Ver Steeg, G., & Galstyan, A. (2012). Information Transfer in Social Media. In *Proceedings of the 21st International Conference on World Wide Web* (pp. 509–518). New York, NY, USA: ACM. https://doi.org/10.1145/2187836.2187906

Verdes, P. F. (2005). Assessing causality from multivariate time series. *Physical Review E*, *72*(2), 026222. https://doi.org/10.1103/PhysRevE.72.026222

Vicente, R., Wibral, M., Lindner, M., & Pipa, G. (2011). Transfer entropy—a model-free measure of effective connectivity for the neurosciences. *Journal of Computational Neuroscience*, *30*(1), 45–67. https://doi.org/10.1007/s10827-010-0262-3

Vincent, J., & Fortunati, L. (Eds.). (2009). *Electronic Emotion: The Mediation of Emotion via Information and Communication Technologies* (1st New edition edition). Bern; Berlin; Frankfurt am Main; Wien u.a.Pieterlen: Peter Lang AG, Internationaler Verlag der Wissenschaften.

Weeks, B. E. (2015). Emotions, Partisanship, and Misperceptions: How Anger and Anxiety Moderate the Effect of Partisan Bias on Susceptibility to Political Misinformation. *Journal of Communication*, *65*(4), 699–719. https://doi.org/10.1111/jcom.12164

Wong, J. C., Levin, S., & Solon, O. (2016, November 16). Bursting the Facebook bubble: we asked voters on the left and right to swap feeds. *The Guardian*. Retrieved from https://www.theguardian.com/us-news/2016/nov/16/facebook-bias-bubble-us-election-conservative-liberal-news-feed

Yan, C., Dillard, J. P., & Shen, F. (2012). Emotion, Motivation, and the Persuasive Effects of Message Framing. *Journal of Communication*, *62*(4), 682–700. https://doi.org/10.1111/j.1460-2466.2012.01655.x

Yeung, R. W. (1991). A new outlook on Shannon's information measures. *IEEE Transactions on Information Theory*, *37*(3), 466–474. https://doi.org/10.1109/18.79902

YouTube. (2016a). Statistics - YouTube. Retrieved December 20, 2016, from https://www.youtube.com/yt/press/statistics.html

YouTube. (2016b). Watch Time optimization tips. Retrieved December 23, 2016, from https://support.google.com/youtube/answer/141805?hl%84=%84en