**Title**
Evidence for a warning bias in information transmission in social networks



**Author**
Altshteyn, Ilya

**Publication Date**
2014

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

# Evidence for a warning bias in information transmission in social networks

A thesis submitted in partial satisfaction

of the requirements for the degree Master of Arts

in Anthropology

by

Ilya Altshteyn

2014

ABSTRACT OF THE THESIS

Evidence for a warning bias in information

transmission in social networks

by

Ilya Altshteyn

Master of Arts in Anthropology

University of California, Los Angeles, 2014

Professor H. Clark Barrett, Chair

Information about environmental dangers is valuable and the cost of transmitting such information to social partners is minimal. Furthermore, an information transmitter who helps a social partner to avoid harm can later reap the benefits of a continued social relationship, and does not have to pay the costs of helping an injured friend. This cost-benefit assymetry suggests there is a positive selective pressure on a propensity to socially transmit information about danger at especially high rates compared to information that is not about danger. We call this predicted propensity a *warning bias*. Here we report the results of tests for this bias using data from the social networking site Twitter. Two coders rated each of 13,203 tweets (publicly-shared 140 character utterances) for whether or not each tweet contained information about danger. The number of retweets for each tweet indexes that tweet's transmission rate, and was our outcome variable. Results of negative binomial regressions showed that tweets about danger have up to 3.13 times as many retweets as tweets that are not about danger.

The thesis of Ilya Altshteyn is approved.

Daniel Fessler

Joseph Manson

H. Clark Barrett, Committee Chair

University of California, Los Angeles

2014

Table of Contents

Main text … p. 1

Supplementary information: Detailed methods of studies 1 and 2, and additional notes … p. 7

Figure 1 … p. 17

Figure 2 … p. 18

Table 1 … p. 19

Appendix A … p. 20

Appendix B … p. 21

References … p. 22

List of Figures

Figure 1 … p. 17

Figure 2 … p. 18

List of Tables

Table 1 … p. 19

## Acknowledgements

This paper is a version of Altshteyn, I. and Barrett, H. C. (2014). *Evidence for a Warning Bias in Information Transmission in Social Networks*. Manuscript in preparation.

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1144087.

Like biological evolution, cultural evolution – change in the frequencies of knowledge, beliefs, attitudes and practices over time in populations – requires differential copying of information. This differential copying occurs as information passes from mind to mind during cultural transmission – the process of social sharing and learning of information – and can be the result of different kinds of cognitive biases (Richerson & Boyd, 2005). *Context biases* result in differential copying of information as a function of who transmits it (e.g., high or low status individuals). *Content biases* result in differential copying of information as a function of what that information is about. Because not all information is equally fitness-useful to the mind mentally representing it, content biases are especially likely to exist for types of information whose acquisition and transmission is useful for survival and reproduction: for example, information about danger. Consistent with this idea, studies have shown that people preferentially attend to dangers in visual arrays (Ohman, Flykt & Esteves, 2001), remember danger information better than comparable non-danger information (Barrett & Broesch, 2012), and are especially likely to believe claims about environmental dangers (Fessler, Pisor & Navarrete, in press).

A less studied question is *where* in the cultural transmission process particular biases might occur. A complete cycle of cultural transmission – reproduction of a cultural norm, belief, or other piece of information – happens in several stages. In particular, information must first be *acquired and retained* (attended to, processed, stored), and then *transmitted* (produced or communicated in a way that allows others to acquire it). For practical reasons, many studies treat the cycle as a black box, without trying to disentangle different possibilities for where in the cycle biases occur. Separate (but potentially overlapping) sets of cognitive mechanisms make each of the two stages possible, and biases, if any, must be located in these mechanisms.

While there is widespread belief that biases exist in the human cultural transmission process, much prior research, including work that documents differential spread of ideas through networks (e.g. Nichols, 2002), is unable to disentangle acquisition/retention and transmission biases, since findings are consistent with biases in the cognitive mechanisms that enable either stage (e.g. Mesoudi, Whiten & Dunbar, 2006). Some studies, such as studies that carefully control information exposure and test later recall, have isolated biases to parts of the acquisition/retention stage, e.g., differential memory storage and / or recall (Barrett & Broesch, 2012). However, to our knowledge, no study to date has definitively demonstrated a content bias in cultural transmission specifically at the transmission stage, which is the purpose of the present studies (although several studies have shown that people *believe* that they would transmit some types of information at especially high rates, without actually doing so, e.g. Heath, Bell & Sternberg, 2001).

Here we report two studies that looked for a possible *warning bias* in information transmission, which we define as an increased probability of transmitting information when that information is about danger, relative to other kinds of information, *ceteris paribus* (note that we are not proposing that this is the *only* content bias in transmission). There are several reasons to expect a warning bias to exist in humans. First, there are the potential adaptive benefits of transmitting fitness-useful information to others: because information about environmental dangers is valuable and the cost of transmitting it to social partners is minimal, sharing such information is an inexpensive form of altruism. An information transmitter who helps a social partner to avoid harm can later reap the benefits of a continued social relationship, and does not have to pay the costs of helping an injured friend. Second, there is the (not mutually exclusive) possibility that a warning bias in humans is a homologous form of a more ancient, primate-wide or mammalian mechanism, given that alarm calls, a form of warning about danger, are widespread in animals (Hollen & Radford, 2009).

Here we look for evidence of a warning bias in social transmission of information on the social media site Twitter. We use number of retweets – reposts of previously posted tweets – as a measure of Twitter users' differential propensity to transmit information as a function of whether or not it is about danger. This method has several virtues. First, unlike laboratory studies of transmission chains (e.g. those based on Bartlett's 1932 chain method), reposting of tweets reflects real-world behavioral decisions to transmit data to one's social network, and has the potential to warn others of actual, fitness-relevant dangers. Second, because Twitter feeds are present onscreen at the time decisions to retweet occur, retweeting reflects behavioral decisions at the transmission stage of social transmission, eliminating factors such as differential retention of information in memory prior to transmission (an effect at the acquisition/retention stage of cultural transmission).

**Study 1: Police and fire department Twitter feeds**

Our first study compared the frequency of retweeting of both danger and non-danger tweets from two kinds of Twitter sources that contain real-world, potentially fitness relevant information about danger: police and fire department Twitter feeds. First, a complete sample of 9,388 tweets (140 character, typed utterances posted on the publically viewable social network Twitter) was obtained from 13 U.S. police and fire department Twitter pages between March, 2009 and June, 2013. Independent coders rated all tweets according to whether they did or did not contain danger information (with a third category, "not sure"). 198 tweets were rated by both coders as containing danger information, and 7979 tweets were rated by both coders as not containing danger information ("not sure" and tweets without agreement were discarded for a final n = 8620; see supplementary materials for details). Examples of danger and non-danger tweets were a scam warning from the police, and an announcement of a change of address for the police impound lot, respectively (see appendix A for samples tweets). Our dependent measure was the number of retweets, the number of times a given tweet was shared by another user on his/her own Twitter page – each reflecting an individual decision to transmit the information in the tweet to others. We modeled the effect of danger content on retweet frequency using negative binomial regression, and found that tweets that are about danger receive 1.35 times as many retweets as those that are not about danger [Table 1, and see supplementary online materials for more details].

### Study 2: Expanded sample of Twitter feeds

It is possible that the kinds of people who view police and fire department Twitter pages are also the kinds of people most concerned with danger. If this were true, and if concern for danger translated into an unusually high probability of sharing danger information with others, then our effect could be a result of sampling bias. In study 2, we extended our test to tweets from banks, parenting magazines, local news and weather services. In addition, Study 1 suffered from a relatively small proportion of danger tweets, so we produced a more balanced sample using a three-step method. First, one coder read each of 25 Twitter feeds and identified potentially danger-containing tweets, using a broad criterion. Each of these was included in the second stage, along with the tweet that immediately preceded it, whether the preceding tweet was about danger or not. Next, an additional randomly selected sample of tweets, 15-100% as large as the step-one sample from each Twitter feed, was added to the pool to add variety and make the proportion of danger to non-danger tweets less transparent. Finally, two other coders implemented the same procedure used in study 1 on the resulting pool of source tweets. Out of an initial 3,815 tweets in the pool, 753 tweets were rated by both secondary coders as being about danger, and 1437 were rated by both coders as being not about danger (disagreements and "not sure" excluded for final n = 2829; see supplementary materials for information). Again, we modeled effects of danger content on retweet frequency using negative binomial regression, and found that tweets about danger receive 3.13 times as many retweets as tweets that are not about danger (see Table 1 and supplementary materials for details, and appendix B for sample tweets).

## Discussion

The existence of a warning bias in the social transmission of information is intuitively plausible, makes adaptive sense given the potential fitness-relevance of danger information, and represents a potential homology with warning behaviors in other mammalian species. Our studies support its existence in humans, and show its effect on the distribution of information in an online social network. We replicated the effect in two independent samples of tweets, using two different sampling methods, in real-world decisions of individuals transmitting information about danger risks rather than under artificial laboratory conditions. To our knowledge, this is the first study that clearly isolates the bias in question to the transmission stage of cultural transmission, as opposed to, e.g., bias in information acquisition and retention on the part of transmitters. It is also one of the few examples of a change in the distribution of cultural variants over time that is clearly due to a cognitive bias (for others see: Gould, 1980; Hinde, 1985). However, we show the effect over only a single transmission step: in most cases, tweets are passed from their original authors to readers who then replicate them. More research is necessary to demonstrate the effects of a warning bias in a longer transmission chain.

If a warning bias exists in face-to-face interpersonal communication networks, it has the potential to shape the direction of cultural evolution, amplifying the frequency of some kinds of information over others. Although it remains to be seen whether the rates at which information is reproduced translate into increased frequency of such information over the long term, the size of the content bias that we documented is fairly large, with danger content amplifying the rate of re-transmission between about one and a half to three times (Study 1 and Study 2, respectively). Other content effects in cultural transmission likely exist and may act in a similar way – for example, researchers have proposed that a "social bias" may increase the rate at which social and gossip information is passed in cultural transmission (Mesoudi, Whiten & Dunbar, 2006). Continuing to investigate such biases and the extent of their influence in cultural transmission is a critical task for social scientists interested in a more complete understanding of the psychological forces that shape trajectories of cultural change.

Supplementary information: Detailed methods of studies 1 and 2, and additional notes

## Detailed methods, Study 1

In this study, we collected tweets from the Twitter feeds of 10 police departments and 3 fire departments from the top 10 US cities by population. 3 cities—Chicago, Houston and Philadelphia were represented by both a police department Twitter and a fire department Twitter. Twitter accounts included in this study had a mean of 22,488 followers, with a range of 2,998 to 103,000 followers.

Twitter limits how far back through the feed of a given account a viewer can look at any one time. We collected all tweets between the date of collection (which was in May and June, 2013) and the earliest tweet we could view from each Twitter feed. In total, we collected we collected 10,435 tweets. Coders did not rate 1,047 of these tweets because of time constraints. This left 9,388 rated tweets, which is the number we report in the main text.

Two coders rated each tweet for whether or not it was about danger. All coders knew the hypothesis in the study. Coders were given the following guidance to determine whether or not a tweet was about danger: "If you lived in the place where the tweet is coming from, would the information in the tweet be (at least theoretically) useful to you in avoiding being harmed?" They were told that harm could be economic or bodily, to oneself or to friends or family. In this study, all tweets came from sources that were located in a specific city.

After completing the danger coding, one coder recorded how many retweets each tweet had, and coded each tweet for whether it contained a number of content types. These included whether the tweet referenced another Twitter user (in which case it would be part of a Twitter conversation, and could appear not only on the original tweeter's page but also on the page of the recipient), whether it contained an image or a video, and whether it contained a link. Importantly, coders did not have access to information about the number of retweets a tweet had while they were coding tweets for danger content.

Only tweets that both danger content coders agreed on and neither coder used an "unsure" code for were used in the analysis. We tested coder agreement only for tweets for which neither coder used the "unsure" code because different coders likely have different thresholds for using the "unsure" code. Where one coder might indicate that a tweet is about danger, another one might indicate that he is unsure. Similarly, if both coders use the "unsure" code for a given tweet, it does not mean that the coders agree about the tweet's content. Such cases therefore do not tell us about coder agreement. Cohen's kappa, a test of inter-rater agreement for categorical items, takes into account the base rates at which coders assign different codes to calculate the proportion of inter-rater agreement beyond what would be expected by chance given those base rates. Cohen's kappa in our data was low, $k = .33$, but significant, $p < .0001$. Most tweets in this dataset are not about danger. The number of codes indicating that a tweet does not contain danger is therefore much higher than the number indicating that a tweet does contain danger. This drives the percent chance that two coders agree on a given tweet up to 88%, leaving very little room for coders to agree above that chance percentage. Our coders actually agreed on 92% of tweets. That means that in 8% of cases, one coder indicated that a tweet was not about danger while the other indicated that it was. This rate is unsurprising given that coders have to detect a rare signal (only ~2% of tweets in this dataset were rated by both coders as being about danger) in a large dataset, and that the operational definition of danger content was not always easily applicable (see appendix A for a sample of tweets from the study 1 dataset).

This reduced the number of possible tweets to analyze from 9,388 to 8969. Of these, 203 were agreed by both coders to be danger tweets, and 8766 were agreed to be non-danger tweets. Tweets that were part of a conversation with another user, identified by their inclusion of an @ symbol preceding another Twitter user's username (this is the syntax the Twitter platform uses to have users send tweets to each other; when a tweet is sent to another user, the receiving user is privately notified of the tweet) were removed, leaving 198 danger tweets and 7979 non-danger tweets. We chose to remove tweets that were considered part of a conversation between users because it is unclear how being a part of a private conversation could influence the likelihood of a Tweet being retweeted in our dataset. This removal shifted the data slightly against our hypothesis—the mean retweet count of danger tweets was reduced by .20, while that of non-danger retweets was reduced by only .07.

The remaining tweets' retweet counts were a negative binomial distribution because the distribution had a long right-side tail and a variance much larger than the mean [Figure 1]. 49% (n = 97) of danger tweets and 55% (n = 4421) of non-danger tweets had zero retweets, so 55% (n = 4518) of study 1 tweets had zero retweets. Mean retweet counts were low for all tweets: 2.34 for danger tweets and 1.74 for non-danger tweets. We report the results of the negative binomial regression as an exponentiated beta value because this value represents the relationship between whether or not a tweet contains danger information and its retweet count. In other words, the reported exponentiated beta value of 1.35 indicates that in the model based on our data, tweets that are about danger get 1.35 times as many retweets as tweets that are not about danger. This result is from a model that includes whether or not a tweet contains a photo as a predictor, which previous research has shown to be an important factor in predicting retweet rates (Zarrella, n.d.). We replicated Zarrella's finding—tweets with a photo received 2.67 times as many retweets as tweets without a photo. Excluding photo as a factor in the analysis did not change the size of the effect of danger content on retweet rates.

One problem with our dataset is that it contains a disproportionate number of danger tweets (7979, compared to only 198 non-danger tweets). Another possible problem is that the date on which tweets were written could affect their retweet rates, because Twitter accounts occasionally have large, sudden changes in number of followers (users who are updated each time the account they are following posts a tweet). For example: the day before the Boston Marathon bombing, the Boston Police Department's Twitter account had approximately 54,600 followers. One week later, that number had grown to over 331,000 (Bindley, 2013). If tweets from the Boston Police department were more likely to be about danger after the bombing than before it, then those tweets would have been exposed to a much wider audience, increasing the chances that they would be retweeted. Even without such large jumps, it is worth controlling for the timing of danger and non-danger tweets because Twitter account followers counts tend to grow over the lifespan of an active account, and it is possible that tweets about danger become more likely as their usefulness grows with the number of followers. To address both of these issues, we excluded all non-danger tweets that were not immediately followed by a tweet about danger. Most Twitter accounts in our sample tweet multiple times per day, so this method gave us a large degree of control over the date that tweets included in the new model were tweeted.

The new subset of the data included 144 danger tweets – the number of danger tweets out of the previous dataset's 198 that were immediately preceded by a tweet that was not about danger – and 144 non-danger tweets. The mean retweet count for danger tweets in this dataset was 2.60, and for non-danger tweets was 2.29. The exponentiated beta value for danger content in this model was 1.37, essentially the same as in the original dataset. The exponentiated beta for the effect of whether or not a tweet had a photo was 10.56 in this model. One non-danger tweet was a significant outlier, having over 100 retweets. This tweet was a photograph of a frozen building. Excluding this tweet from the analysis reduced the retweet count for non-danger tweets to 1.60, increased the exponentiated beta for danger to 1.46 and decreased the exponentiated beta for photo to 4.01.

## Detailed methods, Study 2

In this study we used a different sampling method that increased the number and proportion of danger tweets in our dataset. Tweets were collected from a total of 25 Twitter feeds from the accounts of banks, parenting magazines, local news sources and weather services. We chose these four types of Twitter accounts because we reasoned that they would be likely to occasionally tweet about danger, but that their followers would not be following them specifically *because* they tweet about danger. The accounts in each category (banks, parenting magazines, etc) were chosen because they had relatively many followers.

Twitter accounts included in this study had a mean of 172,236 followers—approximately 8 times more than the mean of 22,488 from accounts included in Study 1. Follower counts ranged from approximately 15,200 to 830,000. Because number of followers is a measure of the size of the audience that a tweeter reaches, this dataset has the potential to provide a more accurate measure of the effect size of danger content on retweet rates, minimizing the influence of variation in retweet rates that is unrelated to danger content (some tweets have unusually large retweet counts for idiosyncratic reasons and can have dramatic effects on the outcome of a binomial regression in a smaller sample, as exemplified by the change in the size of the effect of a tweet containing a photo on its retweet rate when a single outlier tweet was removed from the analysis in study 1).

As mentioned in the main text, the sampling method in this study included three steps. In the first step, coders read through the Twitter feeds of the 25 accounts and picked tweets that they thought could possibly be about danger. They used the same operational definition as coders in study 1, but were instructed to apply it loosely in the first step, including tweets liberally. The purpose of this step was to create a dataset that was more heavily populated with danger tweets than the Twitter feed they came from, decreasing the amount of coder hours required to gather a sizeable collection of danger tweets. Each tweet that might be about danger, along with the tweet immediately preceding it (whether it was about danger or not), was included in the dataset used in the second step. The coder in this step did not include information about which tweets might be about danger in the dataset she created.

In the second step, we added a pseudo-random sample (copy-pasted clusters) of tweets, 15-100% as large as the step-one sample from each Twitter feed (mean: 39%; each of the 25 Twitter feeds had a mean of 110 original tweets collected and a mean of 43 were added in this second stage; the number of original tweets ranged from 12 to 484 across the 25 Twitter feeds). We added these tweets to make the proportion of danger to non-danger tweets less transparent to coders. At this point the dataset contained 3,815 tweets.

In the third step, coders implemented the same procedure as in study 1, now using the operational definition of danger more strictly than in step 1. The coder who had collected the tweets for a given account in step 1 was never one of the two coders who rated the tweets in this third step. 828 tweets were rated by both coders as being about danger, and 1828 were rated by both coders as being not about danger. After excluding (75 danger and 391 non-danger) tweets that were part of a conversation between users, 2190 tweets remained (753 danger and 1437 non-danger tweets).

20% of the tweets in this dataset were about danger. This made danger content an easier signal for coders to detect in this dataset than in the study 1 dataset, and Cohen's kappa increased to .86, $p < .0001$. Expected (by chance) agreement in this dataset was 56%, and coders actually agreed in 94% of cases where neither coder used the code for "unsure." This interrater agreement percent is similar to that in study 1, suggesting that it may be at a ceiling that exists because of coder error. This provides support for the conjecture that Cohen's kappa was low in study 1 because of the high, base rate-induced expected inter-rater agreement.

The retweet counts in this dataset were also in a negative binomial distribution because the distribution had a long right-side tail and a variance much larger than the mean [Figure 2]. In this dataset only 4% ($n = 32$) of danger tweets and 18% ($n = 272$) of no-danger tweets had zero retweets, for a total of 14% ($n = 304$) of the dataset with zero retweets. This is much smaller than the 55% of tweets with zero retweets in study 1, and is likely due largely to the fact that Twitter accounts used in this study had many more followers than those included in study 1. The larger mean number of followers likely also accounts for the much higher mean retweet counts for danger and non-danger tweets in this study: 23.24 retweets for danger tweets (compared to only 2.34 in study 1) and 7.41 retweets for non-danger tweets (compared to 1.74 in study 1). Again, we report the results of the negative binomial regression as an exponentiated beta, showing that tweets that are about danger get 3.31 times as many retweets as those that are not about danger in our model. This result is again from a model that includes whether or not a tweet contains a photo as a predictor, again replicating Zarrella's (n.d.) finding: tweets with a photo receive 2.82 times as many retweets as tweets without a photo. Again, excluding photo as a factor in the analysis did not change the size of the effect of danger content on retweet rates.

## Additional notes

*Does our method isolate effects of a bias at the transmission stage of social transmission?*

In the main text, we claim that our method can isolate an effect at the transmission stage of social transmission. Twitter feeds are present onscreen at the time decisions to retweet occur, so we argue that acquisition/storage stage biases such as differential retention of information in memory prior to transmission cannot explain differential transmission. One possible criticism of this claim is that although memory biases cannot be responsible for the effect, differential attention to or processing of danger and non-danger tweets can explain the result. This would be the case if readers process tweets about danger more thoroughly, or spend more time reading them, and that extra processing leads to more retweeting of danger tweets.

It is an empirical question whether or not readers have a processing bias for danger information. It is plausible that they might. However, there is still the question of what cognitive mechanism(s) drive(s) that processing bias. If the mechanism driving it is the same mechanism that elects information for possible transmission (i.e. if at least a part of this "processing" is the selection of information for transmission), then the bias is still located in the transmission stage. If the mechanism is one that only increases attention to the stimulus and the increased chance of transmission is a result only of that increased attention, then the bias we argue for can be said to be located in the mechanisms supporting the acquisition/storage stage, but still leads to differential transmission.

These possibilities are difficult to disentangle *a priori* in part because the sets of cognitive mechanisms that enable different stages of social transmission likely overlap. In other words, some of the mechanisms and biases that make information acquisition/storage possible and biased also make information transmission possible and biased. Finding out what mechanisms support which stage(s) requires further research.

Potentially relevant to this discussion are the lengths of danger and non-danger tweets. Although processing time and intensity does not have to be only a function of a tweet's length (they can be a function of its content), it is still possible that danger tweets might be longer, leading to increased reading/processing time. This is not the case. For example: in study 2, in a sample of the 258 danger tweets that are immediately preceded by tweets that are not about danger, non-danger tweets are marginally longer than danger tweets: 107 characters compared to 104 characters, respectively. If considering all tweets in that study, danger tweets (n = 753) are marginally longer than non-danger tweets (n = 1437), 110 characters compared to 104 characters, respectively.

*Why an acquisition/storage and transmission model of cultural transmission?*

The purpose of the two-stage model we briefly describe here is to point out that the process of cultural transmission involves multiple sub-processes, which must be supported by partially overlapping sets of cognitive mechanisms, and to organize a research program whose aim is to describe these mechanisms. The research program explicitly investigating the cognitive mechanisms that make cultural transmission possible is only beginning, and a two-stage model is sufficient to make the points in the main text. In the future, finer-grained distinctions will have to be made. We are not tied to a two-stage model of cultural transmission.

*Would the existence of a warning bias suggest that the pool of cultural information should move toward an ever-increasing proportion of danger information?*

No. Our study shows only an advantage in the rate of reproduction of danger information, and not a change in the frequency of danger information in the pool of socially shared information over time. There are likely many cognitive biases and constraints of the natural world that influence what information is transmitted during cultural transmission. Different information types compete for space in the pool of culturally transmitted information. The competition is an approximately zero-sum game because (i) the size of the pool is limited and (ii) although a given cultural variant can activate multiple biases at once, it is unlikely that any cultural variant can satisfy all biases at once. In the pool, proportions of information that is about danger and information that is not about danger likely oscillate over time. This may be partially due to the influence of contextual factors on the rates of reproduction of danger information. For example: if people discuss the dangers of lax (or strict) gun laws after a terrorist attack, then the proportion of danger information in the pool of culturally shared information might temporarily increase.

*What are some future directions for research?*

We propose that there are specialized abilities that make possible each of the different sub-processes that comprise cultural transmission. A clear next step is to continue the research program investigating what these mechanisms are, how they work and what biases they contain. Some of the mechanisms are likely to be evolved specifically to enable cultural transmission (e.g. abilities to create and understand language); others will exist for reasons that may be unrelated to the evolution of cultural transmission, but will still influence what information people socially transmit (e.g. biases in attention toward danger-related stimuli: Ohman, Flykt & Esteves, 2001). Parts of the proposed project have already been accomplished. For example: many biases that researchers study can be thought of as content biases that could impact cultural evolution (e.g. present-at-birth preference for face-like stimuli, biases in perception, etc.). Another next step would be to review such effects with a focus on their influence in cultural transmission. Finally, an important step in the research program investigating human cultural transmission will be to expand on a multi-stage model of cultural transmission. The model lacks details that could make it more useful for studying the cognitive mechanisms that support and bias cultural transmission.

Figure 1: Distribution of Retweets in Study 1

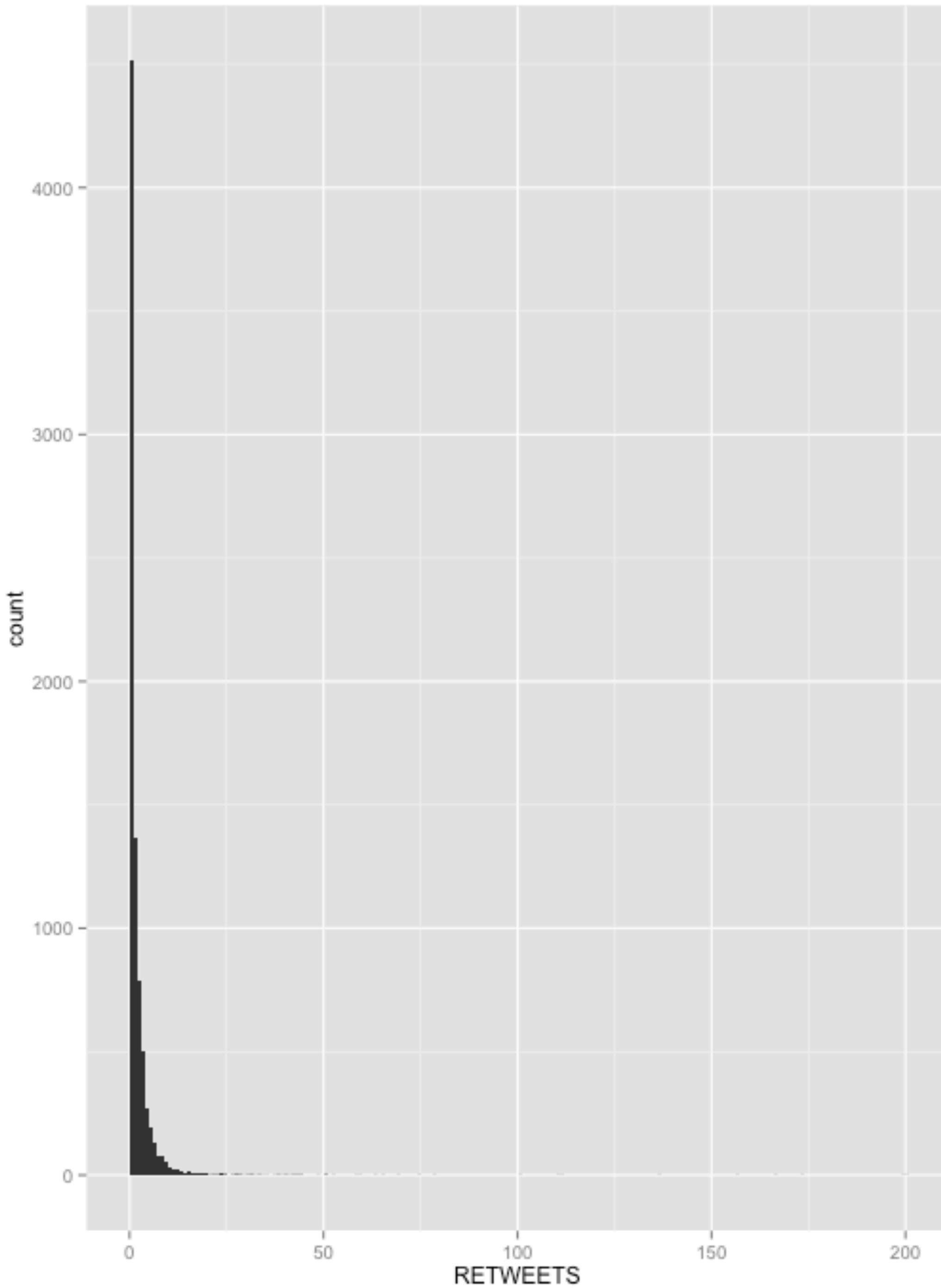

Figure 2: Distribution of Retweets in Study 2

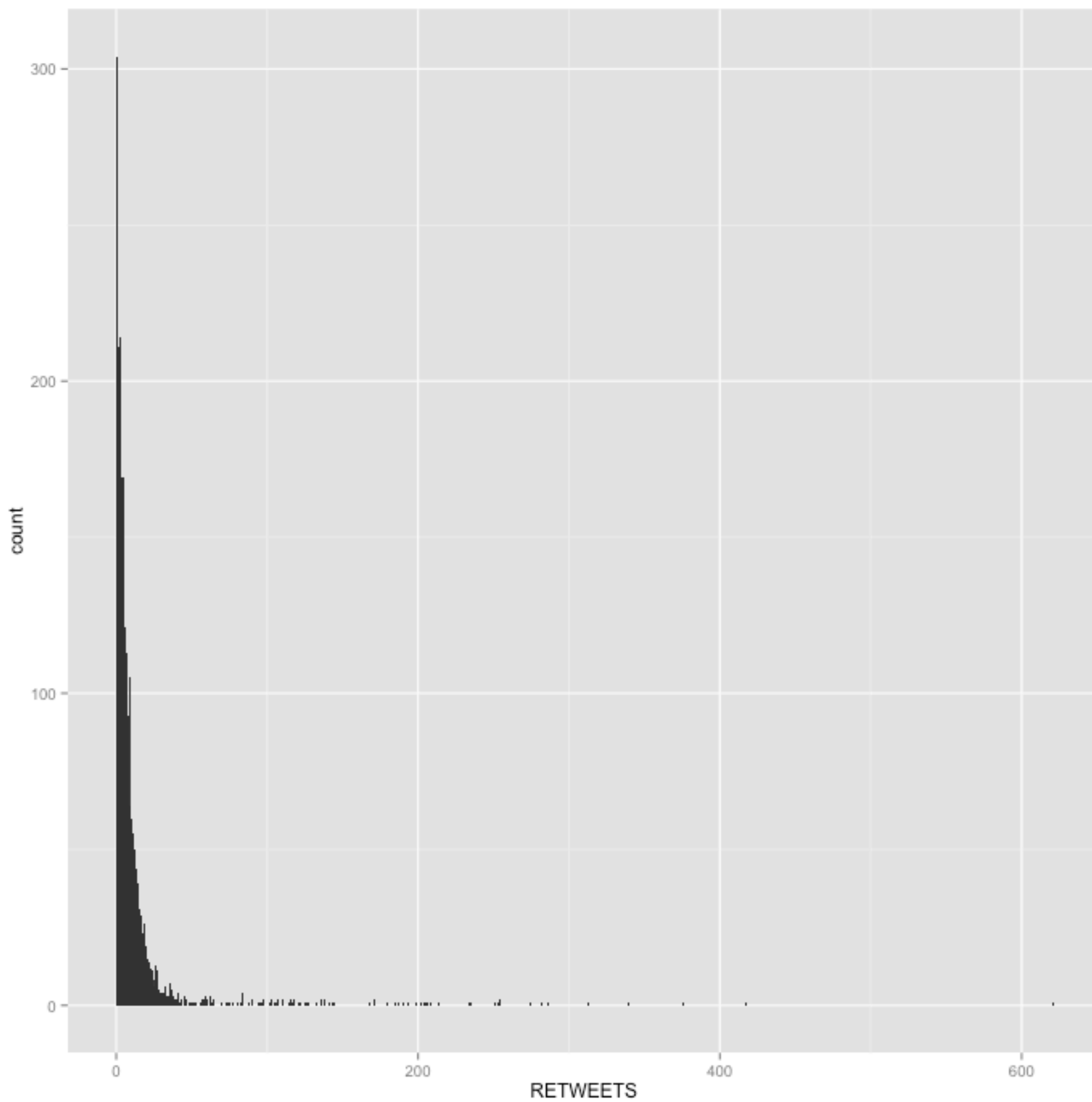

Table 1: Summary of Studies 1 and 2

| Tweet sources | Mean # of followers per feed | # of danger tweets | # of non-danger tweets | Mean danger retweets | Mean non-danger retweets | Exponentiated negative binomial beta estimate | Pr(>\|z\|) |
|---|---|---|---|---|---|---|---|
| Study 1 | 22,488 | 198 | 7979 | 2.34 | 1.74 | 1.35 | >.01 |
| Study 2 | 172,236 | 753 | 1437 | 23.24 | 7.41 | 3.31 | >.001 |

Appendix A: Sample Tweets From Study 1

<u>Rated by both coders as being about danger:</u>

“Level 2 hazmat. 6040 Harper. Co level of 200 on upper floor. No report of illness yet.”

“3 Alarm Fire - Downtown YMCA 1600 Louisiana is currently on fire. Several streets closed. Avoid area.”

“LIVE WIRES: Wires down on 1000 Ivy Hill Rd. #UseExtremeCaution”

“**UPDTAE** Attempted abduction of two year old girl http://bit.ly/VsMmU9  #SJPD”

<u>Rated by both coders as being not about danger:</u>

“San Jose Police Receive State-Funded Grant to Continue Traffic Safety Program http://bit.ly/tEuONx  #SJPD”

“Morning Philly! Keep it quiet out there! #notonfire”

“PhillyPolice Inspector Retires After 35 Years of Service http://bit.ly/XRN6k5”

“Happening Now:Hundreds of local kids get back to school haircuts/styles & school supplies as part of HPD's 4th annual "Look Good, Feel Good"”

Appendix B: Sample Tweets From Study 2

Rated by both coders as being about danger:

"THIS JUST IN: Malibu High School reports caulk samples containing PCB levels above legal limit to EPA, agency says."

"#Recall alert on HALO SleepSack wearable blankets! Get the details here: http://bit.ly/184n6X7  #sleepsack"

"For Tri-State Area Burglars The Holidays Are The Season For Stealing http://cbsloc.al/ICQxdt "

"Tornado Warning issued for Bienville, Natchitoches, and Red River Parish in Louisiana until 4:30pm CDT."

Rated by both coders as being not about danger:

"NWS Austin/San Antonio reports 9.57 inches of rain at San Antonio Intl. Airport already today! Now the 2nd all time daily rainfall #record"

"Former El Segundo teacher charged with molesting 13 boys http://abc7.la/1l6Qu60 "

"The owner of the Indian Point nuclear power plant reached a tentative contract agreement, avoiding a strike http://4.nbcny.com/D1CxGHs"

"This childhood activity could lead to your child becoming and entrepreneur: http://bit.ly/1dS3hdy"

## References


Barrett, H. C., & Broesch, J. (2012). Prepared social learning about dangerous animals in children. *Evolution and Human Behavior*, *33*(5), 499-508.

Bartlett, F. C. (1932). *Remembering*. Oxford: MacMillan.

Bindley, K. (2013, April 26). Boston Police Twitter: How Cop Team Tweets Led City From Terror To Joy. The Huffington Post. Retrieved June 2, 2014, from http://www.huffingtonpost.com/2013/04/26/boston-police-twitter-marathon_n_3157472.html

Fessler, D. M., Pisor, A. C., & Navarrete, C. D. (2014). Negatively-biased credulity and the cultural evolution of beliefs. *PloS one*, *9*(4), e95167.

Gould, S.J. (1980). A biological homage to Mickey Mouse. In *The Panda's Thumb: More reflections in Natural History*, pp. 95–107. WW Norton and Company.

Heath, C., Bell, C., & Sternberg, E. (2001). Emotional selection in memes: the case of urban legends. *Journal of personality and social psychology*, *81*(6), 1028.

Hinde, R. A., & Barden, L. A. (1985). The evolution of the teddy bear. *Animal Behaviour*, *33*(4), 1371-1373.

Hollen, L. I., & Radford, A. N. (2009). The development of alarm call behaviour in mammals and birds. *Animal Behaviour*, *78*(4), 791-800.

Mesoudi, A., Whiten, A., & Dunbar, R. (2006). A bias for social information in human cultural transmission. *British Journal of Psychology*, *97*(3), 405-423.

Nichols, S. (2002). On The Genealogy Of Norms: A Case For The Role Of Emotion In Cultural Evolution*. *Philosophy of Science*, *69*(2), 234-255.

Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: detecting the snake in the grass. *Journal of experimental psychology: general*, *130*(3), 466.

Richerson, P. J. and R. Boyd (2005). Not by Genes Alone: How Culture Transformed Human Evolution (University of Chicago Press, Chicago).

Zarrella, D. (n.d.). Use Images on Twitter to Get More ReTweets. *Dan Zarrella RSS*. Retrieved June 2, 2014, from http://danzarrella.com/use-images-on-twitter-to-get-more-retweets.html