What's in a @name?
How Name Value Biases Judgment of Microblog Authors

Aditya Pal
Deaprtment of Computer Science
University of Minnesota
Minneapolis, MN 55455, USA
apal@cs.umn.edu

Scott Counts
Microsoft Research
One Microsoft Way
Redmond, WA 98052, USA
counts@microsoft.com

Abstract

Bias can be defined as selective favoritism exhibited by human beings when posed with a task of decision making across multiple options. Online communities present plenty of decision making opportunities to their users. Users exhibit biases in their attachments, voting and ratings and other tasks of decision making. We study bias amongst microblog users due to the value of an author's name. We describe the relationship between name value bias and number of followers, and cluster authors and readers based on patterns of bias they receive and exhibit, respectively. For authors we show that content from known names (e.g., @CNN) is rated artificially high, while content from unknown names is rated artificially low. For readers, our results indicate that there are two types: *slightly biased*, *heavily biased*. A subsequent analysis of Twitter author names revealed attributes of names that underlie this bias, including effects for gender, type of name (individual versus organization), and degree of topical relevance. We discuss how our work can be instructive to content distributors and search engines in leveraging and presenting microblog content.

Introduction

A cognitive bias is the human tendency to err systematically when making judgment based on heuristics rather than thorough analysis of evidence (Tversky and Kahneman 1974). Online media presents venues where a user has to choose one of the several options presented to her. Decision making in these environments can result in users making choices influenced by biases. As an example, consider product ratings in online marketplaces like Amazon.com. Certainly these ratings can influence purchasing decisions, but even the ratings a user would have made may be biased due to anchoring around existing ratings and can expose biases (Lauw, Lim, and Wang 2006).

Users of social media such as Twitter, make analogous choices at various points, ranging from selecting users to follow to what content to retweet, each of which might incorporate bias. The behavior of *following*, for example, shows preferential attachment (Howard 2008), such that users are more likely to follow other users with high follower counts, thereby increasing their follower count further. Recent research also explain that user's perception about retweeeting could be biased based on several characteristics like url, hashtags, etc (Suh et al. 2010). In general, the assessment of the quality of the content or the user is based in part on biases of the users evaluating them.

In the current work, we study how bias due to name value impacts the perception of quality of Twitter authors. Understanding this bias is increasingly important as microblogs become a vehicle for content distribution (Kwak et al. 2010) and consumption moves to domains like search results pages where readers must evaluate content from authors they do not follow. In these scenarios readers have very little information to draw on when making these content evaluations. In fact, often a reader sees only the user name and photo of the author. Thus, while the username may seem like a mere detail, the scarcity of contextual information around content in these environments makes it a key piece of information that may be biasing content consumers.

Indeed our results show exactly that: users are biased in their perception of the quality of the content based solely on the name of the author. That is, judgments of an author's content are biased either positively or negatively, as compared to judgments of that author's content made in the absence of the author's name. We describe several author name attributes that affect this bias. We also discuss the kind of impact this can have for content providers and how they can tap user biases to improve the presentation of microblog content.

In particular, our main contributions are as follows:

- We discover how name value affects ratings of the quality of an author's tweets. We show how some authors benefit due to their names, how some are disadvantaged, and the degree to which this happens. In particular, we analyze the relationship between name value with follower count of Twitter authors and show that famous people are beneficiaries.
- We measure to what degree Tweet *readers* are biased, showing that they generally follow a bi-model distribution, one mode representing heavily influenced and another mode representing minimally influenced.
- We consider several different attributes to characterize author names, such as gender analysis. We show that cer-

tain qualities of names are more effective in influencing reader's perception about the quality of their content over other kinds of names.

We present our analyses, and then discuss applications to microblog use scenarios, such as follower recommendation.

Related Work

Bias has been studied in a variety of domains. For instance, (Bechar-Israeli 1995) studied how people adopted names in IRC relay chat and the association of the names corresponding to the role the person intends to play in the community. The y suggested that in order to reflect technical skill and fluency users deliberately violated conventional linguistic norms in their user names. (Jacobson 1999) studied how users in text-based virtual communities develop impression of one another. They show that these impressions are based not only on cues provided, but also on the conceptual categories and cognitive models people use in interpreting those cues.

In the domain of online commerce, Lauw et al. (Lauw, Lim, and Wang 2006) formulated bias of users in customer rating websites. They proposed a model to capture the bias of users based on the controversy that a product creates. They showed that biases can be effectively measured from the product ratings of users in a shopping website such as *Amazon.com*.

Bias in social media contexts has been studied in different forms, such as *trust* or *influence*. (Kittur and Kraut 2008) showed that the initial edits of Wikipedia pages set the tone of the article and subsequent edits do not deviate much from that tone. This hints at the use of anchoring heuristics such that later editors do not make larger changes that are misaligned with the initial pitch of the document. Also in Wikipedia, (Leskovec, Huttenlocher, and Kleinberg 2010) recently showed that during wikipedia elections people prefer to vote for candidates who have much higher or much lower achievements that their own. In other words this work shows that people vote against the candidates whose achievements matches their own achievements, where achievements is measured in terms of the number of edits a person makes on an article. Again we see something of an anchoring process in which people use their own achievements as a benchmark from which to judge others.

(Pal and Konstan 2010) studied question selection biases of users in question answering communities. They show that the experts in these communities are more biased towards picking questions with no answers or answers with low value. The authors show that the selection biases of users can be used to identify experts in these communities effectively. Cosley et al. (Cosley et al. 2003) showed that recommender systems affect the opinions of their users in that users can be manipulated towards the predictions that system shows.

Bias in microblog environments has not been studied extensively, though (Pal and Counts 2011) confirmed an initial hypothesis that microblog readers are biased towards an overly positive evaluation of content authored by celebrities. They also showed that non-celebrities can suffer as a result of lack of name value. Our research builds on the initial findings of (Pal and Counts 2011) and we explore the name value bias in greater detail. To our knowledge, this is the first in depth study of bias due to the value of an author's name in a microblog context.

Dataset and User Study Design

Twitter Dataset

We selected three topics: *iPhone*, *Oil Spill*, *World Cup* that were amongst the top trending topics on Twitter during June 2010. Through an agreement with Twitter, we had at our disposal all the publicly accessible tweets posted on Twitter from 6th-June-2010 to 10th-June-2010 (5 days). We extracted all the tweets that mentioned the above topics using a keyword match. Table 1 presents the statistics of the extracted tweets and the users who posted them.

	iphone	oil spill	world cup
no. of users	430,245	64,892	44,387
no. of tweets	1,029,883	148,364	385,073

Table 1: Basic statistics of the extracted dataset.

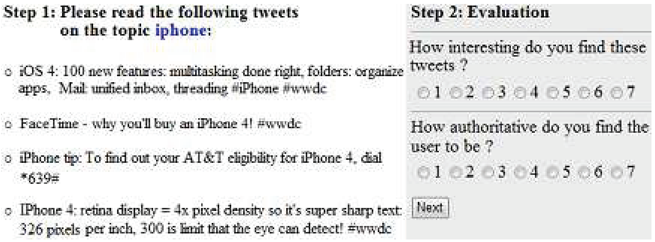
Next, we selected 40 Twitter authors per topic from this dataset, 30 of which were selected using an expert identification algorithm (Pal and Counts 2011). While all authors were considered topically relevant according to the selection algorithm, we ensured that the 30 contained a mix of those with small and large numbers of followers. The number of followers for these 30 authors varied from 29 to 2,161,200. Finally we selected 10 authors randomly from a set of authors who have tweeted on topic but were not considered authorities according to the algorithm. This was done to add noise to the stimuli sample.

For each author, we picked 4 of their topical tweets to use as stimuli. We ensured that the 4 tweets per author were not retweets of other tweets, by discarding tweets containing "RT" or by checking meta-data that tells about the origin of the tweet. If an author had less than 4 topical tweets (8 out of 120 authors), we picked from their non-topical tweets to get 4 tweets per author.

Procedure and Evaluation Criteria

Participants 48 participants (25% female) were recruited via email distribution lists within our company, and were compensated with a $10 lunch coupon. The median age of participants was 31 with an average age of 32.1 and a standard deviation of 5.9. Participants were required to be familiar with the concept of tweets and use Twitter at least once per week.

Procedure Each participant was first randomly assigned a topic. The 40 authors for that topic were then presented to the participant for evaluation one by one. In each case, the four tweets of the author were shown, and after reading the tweets, participants rated how *interesting* they found the tweets and how *authoritative* they found the author (see Figure 1 and 2). Each rating was made on a 7-point Likert scale.



Step 1: Please read the following tweets on the topic iphone:

- iOS 4: 100 new features: multitasking done right, folders: organize apps, Mail: unified inbox, threading #iPhone #wwdc
- FaceTime - why you'll buy an iPhone 4! #wwdc
- iPhone tip: To find out your AT&T eligibility for iPhone 4, dial *639#
- IPhone 4: retina display = 4x pixel density so it's super sharp text: 326 pixels per inch, 300 is limit that the eye can detect! #wwdc

Step 2: Evaluation

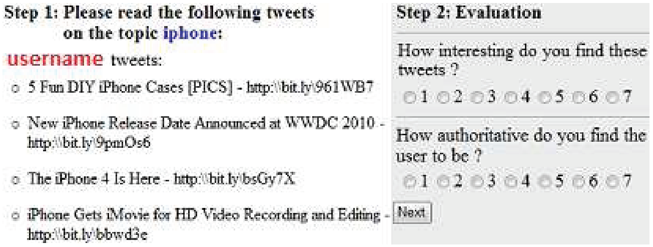
How interesting do you find these tweets ?

○1 ○2 ○3 ○4 ○5 ○6 ○7

How authoritative do you find the user to be ?

○1 ○2 ○3 ○4 ○5 ○6 ○7

Figure 1: Anonymous rating screen. In this condition, author who posted the tweets is not shown to the participant.



Step 1: Please read the following tweets on the topic iphone:

username tweets:

- 5 Fun DIY iPhone Cases [PICS] - http:\\bit.ly\961WB7
- New iPhone Release Date Announced at WWDC 2010 - http:\\bit.ly\9pmOs6
- The iPhone 4 Is Here - http:\\bit.ly\bsGy7X
- iPhone Gets iMovie for HD Video Recording and Editing - http:\\bit.ly\bbwd3e

Step 2: Evaluation

How interesting do you find these tweets ?

○1 ○2 ○3 ○4 ○5 ○6 ○7

How authoritative do you find the user to be ?

○1 ○2 ○3 ○4 ○5 ○6 ○7

Figure 2: Non-Anonymous rating screen. In this condition, author who posted the tweets is shown to the participant.

The first 20 authors were presented anonymously, without the author's Twitter user name shown (see Figure 1). The second 20 authors were evaluated non-anonymously, with their Twitter user name's shown (see Figure 2). This difference (showing or not showing the author's name) was the only difference between our two conditions. Author presentation order was randomized across participants.

Note that each participant rated a given author only once, either in the anonymous or non-anonymous condition. Also, each author was rated an equal number of times in the anonymous and non-anonymous conditions. We received 8 ratings per author per condition. Each author's tweets remained the same across the two conditions, and as a result, any rating difference between the two conditions would arise from showing or not showing the author name. Finally, we removed variables from the tweets that could identify the Twitter authors, such as @username. Participants were advised only to look at the provided text and judge authors based on that.

The above design procedure helps us capture several things. First we capture the merit of the tweets produced by the author irrespective of any other confounding factors such as author name and other meta-data available through deeper inspection of the authors page on Twitter (author bio, picture, follower count, etc). Second, it allows us to capture the variability that participants exhibit when they analyze content without knowing the source from when they know the source. An informal interview with a few participants revealed that they found it hard to estimate authoritativeness of authors in the anonymous setting and found the role of names negligible in estimating the interestingness of the tweets. This leaves us expecting no significant change in interestingness ratings of an author in the two conditions and a significant change in their authority ratings in the two conditions, but the actual results were more surprising than this initial guesstimate. The next section presents the actual outcome and our analysis of the results.

Rating Density Estimation

Each author received 8 ratings on a 7-point Likert scale per condition. The discrete scale doesn't allow users to rate at a higher granularity that would enable finer comparison between authors and the conditions. On the other hand a scale more flexible than a standard Likert scale could have confused users and led to fine-tuned ratings that would have made inter-rater reliability difficult to assess.

For a richer comparison between the ratings, we used a density estimation technique to build a probability distribution over the ratings instead of considering point estimates like taking the mean. This enables us to distinguish ratings of $\{1, 7\}$ from $\{4, 4\}$ which are indistinguishable if we consider their mean.

We used Gaussian kernel to estimate the continuous density of the ratings. Consider N ratings $r = \{r_1, r_2, ..., r_N\}$. For r_i we consider a Gaussian distribution with $\mu = r_i$ and $\sigma = 0.5$. N ratings are combined as follows:

$$P(r) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sqrt{2\pi\sigma^2}}\,exp\{-\frac{1}{2\sigma^2}\,(r - r_i)^2\} \quad (1)$$

where σ is a smoothing parameter in this case because as σ^2 is increased, the distribution smoothes and local peaks get diluted. As σ^2 is decreased, the distribution degenerates to a sharp peaked discrete distribution (like a histogram without binning). We set $\sigma = 0.5$ empirically.

There are several benefits of a continuous probability density function. It is defined at all the points on the number line rather than just 8 of them (1,2,..., 7). It enables us to choose continuous plots over histograms, improving presentation in our case. It smoothes the ratings to reflect better approximation to true ratings and reduces the effect of outliers on the small number of ratings. It enables the use of distance measures (such as KL-divergence) over differences of point estimates (mean, sum). It is better over discrete probability distributions, because KL-divergence for a discrete distribution could be undefined due to being divided by zero. We use symmetric KL-divergence for comparing two probability distributions, defined as follows:

$$KL(p||q) = \frac{1}{2}\int [p(x)\log\frac{p(x)}{q(x)} + q(x)\log\frac{q(x)}{p(x)}]\,\mathrm{d}x \quad (2)$$

Figure 3 shows the comparison of the discrete and the smooth density distribution. We can see that the smooth distribution tries to reduce sharp variations and yet closely capture the observed distribution for $\sigma = 0.5$.

Author Rating Analysis Results

Notation

We employ the following notation: I stands for *interestingness*, A stands for *authority*, α stands for *anonymous*, β stands for *non-anonymous*.

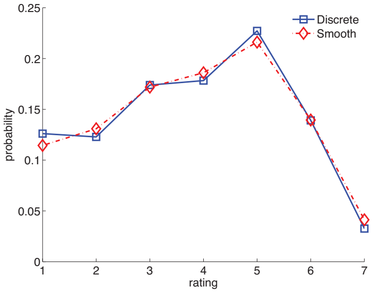


Figure 3: For this illustration the probability value of smooth distribution is scaled at 8 points and rescaled to sum to 1.

So in our context, I and A represent smooth probability density of interestingness and authority ratings. Consider an author a with smooth rating distribution I^a_α over interestingness ratings received anonymously, $\bar{I}^a_\alpha = \int_{-\infty}^{\infty} r \cdot I^a_\alpha \cdot dr$. The expected value would be the same as the mean of all the ratings received by a under the same measure.

Similarly, $\bar{I}_\alpha$ stands for the average of anonymous interestingness rating (expected value) for all the authors and so on.

Anonymous vs Non-Anonymous Ratings

The Pearson correlation between $\bar{I}^a_\alpha - \bar{I}^a_\beta$ is 0.53 and between $\bar{A}^a_\alpha - \bar{A}^a_\beta$ is 0.57 indicating that α and β measures share a positive but moderate linear relationship. This means that for an individual author her anonymous and non-anonymous ratings are correlated but not very strongly, suggesting that other than the quality of the tweets, her name influenced participants' judgment.

Table 2 presents the rating averages across all authors showing that authors were awarded higher ratings anonymously than non-anonymously. We confirm the statistical significance of this difference by running a paired one-sided t-test. $\bar{I}_\alpha$ is significantly higher than $\bar{I}_\beta$ ($p < 0.001$), whereas for $\bar{A}_\alpha$ was only marginally significantly higher than $\bar{A}_\beta$ ($p = 0.07$).

$\bar{I}_\alpha$	$\bar{I}_\beta$	$\bar{A}_\alpha$	$\bar{A}_\beta$
3.72	3.44	3.55	3.42

Table 2: Average author ratings.

Figure 4 shows the plot of average rating distributions. We observe two key things from this plot:

- The average rating distributions have two modes: one towards high rating value and another towards low rating value, indicating that participants collectively could identify that there are two types of authors that are getting rated. Indeed that is the case, as we mixed top authors with randomly selected authors.
- As we move from α to β ratings, the distribution shifts towards the left. The likelihood of an author receiving a 5 or more on interestingness is much higher when their content is rated anonymously than when their names are shown.

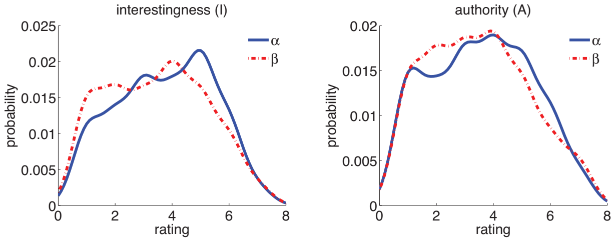


Figure 4: Average rating distribution for authors.

Rating Distribution Comparisons

The previous result presents an aggregate picture of how ratings vary from one condition to other. It suggests that authors are better off getting rated anonymously for both the interestingness of their tweets and their authoritativeness, rather than when their names are visible.

Here we aim to discover if all authors are better of being rated anonymously or do some authors benefit on the contrary. We use KMeans clustering algorithm over the I^a_β rating distributions of authors by using symmetric KL-divergence as the distance criteria between two authors. We allowed clusters to be merged with another cluster if it had few points and cluster to split if it had a lot of points. The clustering algorithm almost always converged to 3 clusters for our initial choice of clusters set to either of 2, 3, 4, 5.

There were roughly equal number of authors per cluster and we labeled them: {bad, average, good }. Authors in the bad cluster and good cluster received lowest and highest $\bar{I}_\beta$ ratings, respectively. Table 3 shows the average author ratings per cluster. It shows that the good authors received higher rating when their names were made visible. This is contrary to what we saw earlier. Indeed this presents evidence that there is a considerable chunk of authors (38%) that benefited slightly from having their names shown. The remaining clusters suffered, especially the author's in bad cluster (drop in rating by 12-20%), when their names were shown.

cluster	% authors	$\bar{I}_\alpha$	$\bar{I}_\beta$	$\bar{A}_\alpha$	$\bar{A}_\beta$
bad	34	3.14	2.49	2.97	2.60
average	28	3.93	3.55	3.74	3.52
good	38	4.21	4.34	4.04	4.21

Table 3: Average ratings per cluster.

Figure 5 shows the rating distribution over anonymous measures for authors in different clusters. The trend we see is that bad authors were rated bad even anonymously indicating that the quality of their content was bad. The average authors were not rated very much differently from the good authors anonymously (KL divergence of 0.02) yet when their names were shown, they lagged by a difference of almost 1 point from the good authors (KL divergence of 0.23, ttest

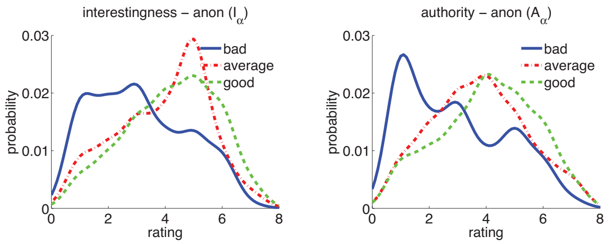


Figure 5: Average AI ratings given by participants to authors in different clusters.

$p < 0.001$). This indicates that even though "average" authors provided high quality tweets, there their name created a detrimental impact on the participants.

Rating and Follower Count

The previous result suggests that anonymous ratings benefited some authors, while non-anonymous ratings benefited others. To further breakdown the influence of author names on ratings of their content, we turn to arguably the most prominent attribute of a person in Twitter: number of followers. Prior work (Pal and Counts 2011) showed that those authors with more than 50,000 followers benefited significantly when their names were shown, while those with fewer than 50,000 followers were rated significantly lower when their names were revealed. Here we examine the relationship between follower count and ratings more comprehensively [1].

Figure 6 shows the result of linear regressions between author ratings and their number of followers. The positive slope for α reinforces the belief that *popular authors post good content* or alternately, *good content producers become popular*. Our current interest, however, lies in the comparison of these regressions in the anonymous (on the left in Figure 6) and non anonymous (right half of Figure 6) cases. First we note that for those authors with the highest follower counts, above about 12 on the log scale x-axis, most people are below the regression line when their content is rated anonymously but above the line when their content is rated non-anonymously. Thus, these people, likely celebrities or organizations, go from below to above expected ratings simply by virtue of their name being seen.

Second, for both measures, the intercepts are lower and the slopes slightly steeper in the case of non-anonymous ratings, again indicating an effect for greater increases in ratings corresponding to increases in follower counts when the user names are shown. Finally, we note that the fits are better in the case of non-anonymous ratings [$R^2(I_\alpha)$=0.16, $R^2(I_\beta)$=0.19, $R^2(A_\alpha)$=0.16, $R^2(A_\beta)$=0.21]. While the amount of variance explained was small, we are concerned here with the relative improvement in fit when ratings are made non-anonymously: 19% improvement for interestingness and 31% improvement for authoritativeness. This moderately tighter relationship between ratings and follower count implies that raters were using the name (whether it was known or unknown) as a common heuristic that biased their ratings.

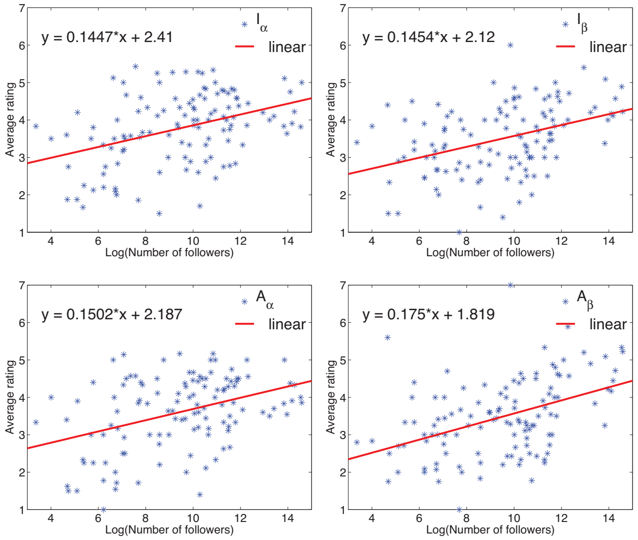


Figure 6: Scatter plot of author's mean ratings and number of her followers along with the details of Linear regression for the underlying data. The Pearson correlation is $\rho(I_\alpha)$=0.39, $\rho(I_\beta)$=0.42, $\rho(A_\alpha)$=0.39, $\rho(A_\beta)$=0.45

Participant Rating Analysis Results

So far we established that 1) authors with high follower counts are rated higher than they would be on the quality of their tweets alone, 2) evaluations of the roughly 30% of authors clustered in the average group are degraded simply be the presence of their usernames, and 3) overall authors received higher ratings when their names are not shown to the participants. We now turn to analyses of our raters.

Figure 7 shows the average rating distributions provided by the participants across different conditions. These distributions look remarkably similar to the average distributions for authors (see Figure 4), suggesting that participants on average were more biased towards lower ratings when author names were presented to them. This could be because they failed to recognize the author based on their Twitter user names.

The aggregate distributions for participants still seem to

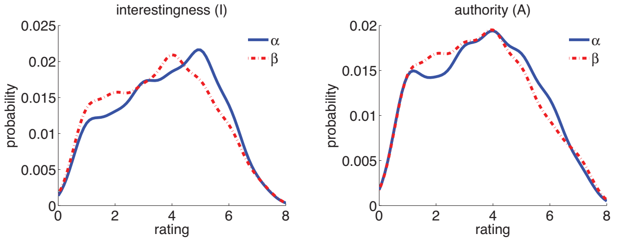


Figure 7: Average rating distribution for the participants.

[1] Note that the evaluators were not provided the follower count of the authors.

coincide quite nicely across the α and β conditions, suggesting that though participants got restrictive in β condition, but they did so only very slightly. To see if this is the case, we computed the KL divergence of aggregate distribution and also the aggregate of KL divergence of individual differences (i.e., the KL divergence for each user was computed separately, and then averaged). Table 4 presents the difference in the divergence values of the two distributions.

	$I(\alpha - \beta)$	$A(\alpha - \beta)$
KL-divergence of average rating distribution (Figure 7)	0.012	0.005
Average KL-divergence of individual rating distribution	0.353	0.305

Table 4: Symmetric KL-divergence between participant rating distributions.

This result shows that the rating distribution of participants under our α and β conditions were in fact quite different (2-sided ttest, $p < 0.001$), and thus unlike what the average distribution suggests, each individual participant is influenced by the presence of the username. We cannot yet draw conclusions about this result because each participant rated different authors anonymously than she rates non-anonymously. For a better test, we consider authors with follower count larger than 50,000 (as a way to create matched sets across the two conditions) and still we observe that the participants' rating distributions differ significantly across the two conditions.

Measuring Participant Bias

Earlier we saw that participants were biased due to the showing of the authors' names. In some cases this creates a negative impact on authors' ratings and in other cases it takes a positive turn. In particular, we saw that the authors in the "average" cluster received poorer ratings non-anonymously, yet they were not perceived that badly when their names were not shown. Here we aim to estimate this shift in perception of participants.

To do this, we consider the anonymous ratings received by authors as the true rating they deserve for the quality of their tweets. Then for a given participant, the divergence between her rating (recall that each rating is a Gaussian distribution with $\sigma = 0.5$) made when the author's name was shown and this true rating indicates the bias she was shown as the result of name value. This bias is averaged for each participant across all the authors she rated non-anonymously. Figure 8 shows the density distribution of biases indicating that majority of the participants (65-70%) are only slightly biased with KL-divergence 1, yet there is a small group of users who are very strongly biased with KL-divergence close to 4.5.

We also see that participants are more biased on the interestingness measure (avg KL-div. = 2.6159) than authoritativeness (avg KL-div. = 1.5335). This can also be seen in the user probability mass of A being higher in the first peak than the first peak of I, meaning that more users were minimally biased when rating authority as compared to rating interestingness. This result is contrary to what we thought initially based on the interview of participants. They suggested that it was easier to rate authors on authority by their names, while the names were irrelevant for rating interestingness.

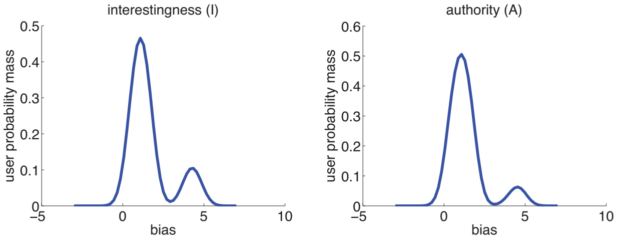


Figure 8: Biases of participants.

Measuring Other factors for bias

So far we established that in general users get more conservative in giving ratings to authors when their names are shown in comparison to when not. This scheme helps prominent authors who users can recognize or have an association with in some way, and hurts nearly 30% of authors who are in fact tweeting high quality content. This result is intriguing and leads us to explore further how authors' names influence raters.

Author Names Study

In this section, we explore attributes that can be derived from the author names and whether they contribute towards swaying reader's ratings. We consider the following 3 attributes: gender, type of author (individual or organization), and topical fit (extent to which the name sounds on-topic).

The authors of this paper rated the Twitter author names on these three attributes. In order to break ties between the authors, we used the codings provided by an outside rater.

Gender Table 9 shows the author ratings for the three categories of gender: male, female, cant tell by the username. This partition of authors based on gender indicates that a majority of top authors are gender neutral. They are either news agencies or associated with organizations and their name does not reflect their gender (e.g. time, mashable, nwf). Both male and female authors experience a slightly larger drop in their authority ratings (6-8%) than gender neutral authors (drop of 2-3%). Male authors were perceived as providing better content than the female authors (higher α ratings) and were seen to be more authoritative as well. We also observe that for interestingness ratings, female authors saw a very slight increase from anonymous to non-anonymous condition, though the number of female authors is substantially low and the increase is not significant, so we cannot say much about this result. Overall this result indicates that users could be influenced to varying degrees depending on the gender of the author.

Type Here we compare three types of usernames: individual humans, organizations, and cant tell by the username.

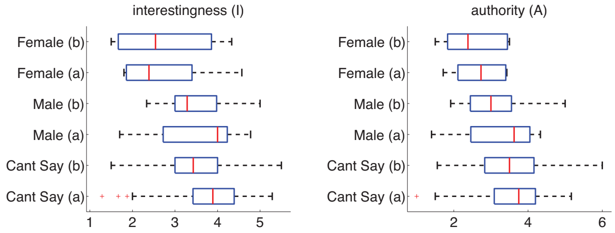


Figure 9: Average author ratings categorized based on $Q1$. Note here (a) represents α condition and (b) represents β condition.

Figure 10 shows that the names associated with organizations are better across all rating categories. Interestingly, we see that the authority ratings of organizations improve from α to β conditions, the "can't tell" category records a significant drop of .35 (10%), and human authors receive a drop of 0.26 (8%). The drop in authority ratings for these two categories is significant at $p < 0.001$ (two-tailed ttest). Overall we see that organizations do not suffer the drop in ratings that individuals do when ratings are made non-anonymously.

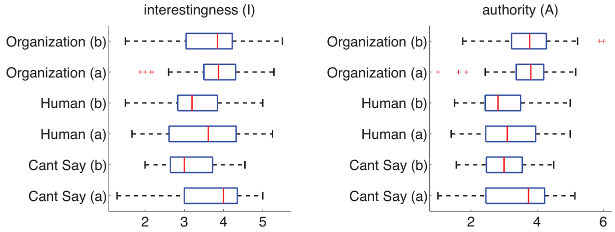


Figure 10: Average author ratings categorized based on $Q2$. Note here (a) represents α condition and (b) represents β condition.

Topical Another dimension to study the effectiveness of the author names is to estimate how much the author names tell about the topic they tweet on. For a topic like "iphone" an author name like "iphone" is highly topically sounding. Additionally, for a topic like "worldcup" the author name "mikecnn" gives an impression that the author is a sports news correspondent at cnn and the name is slightly topically relevant.

Figure 11 indicates the average ratings of authors categorized based on how topically sounding their name is. We observe that highly topically relevant author names (yes++) receive slightly lower ratings than the slightly relevant names (yes). The yes category authors see a significant jump in their ratings and the no category authors see a significant drop in their ratings (both statistically significant at $p < 0.025$). This result suggests that it is better to have a name only slightly relevant to topic name. One possibility for this is that names that are moderately on topic convey a sense that they are authorities on more than just the single topic (e.g., @Apple is an authority on all things about Apple Computer, not just the iPhone).

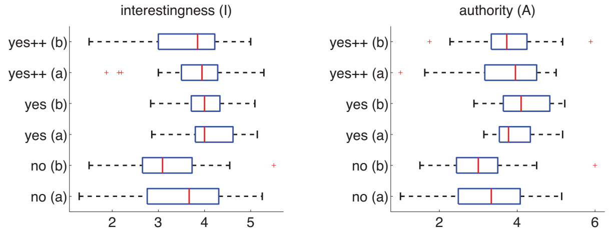


Figure 11: Average author ratings categorized based on $Q3$. Note here (a) represents α condition and (b) represents β condition.

Discussion

We started by demonstrating a main effect for anonymity in ratings: on average, ratings of Twitter authors and their content were slightly, but significantly, lower when their names were shown. Further exploration showed this to be an averaging out of several effects across both authors and raters. First, we showed that authors who were rated lower when their names were shown, see increased ratings when rated anonymously, while those rated highest non-anonymously receive lowered ratings when rated anonymously. Given that the only difference in the two cases was the presence of their user names, the names must have pushed raters to more extreme ratings.

This alone does not necessarily correspond to the concept of "name value" of popular or known people artificially getting a boost due to their name. Thus we used follower counts as a proxy for the likelihood a user name would be known to raters. Here we show that authors with high follower counts receive ratings below expectation when rated anonymously, but above expectation when rated non-anonymously. Also, the better fit of the regressions to the data in the non-anonymous ratings conditions provides support for the idea that the user name serves as a common heuristic used by participants when making ratings.

We characterized our raters in terms of the degree and direction of bias shown. Roughly, this distribution is normally distributed around a group of users that are biased, but only minimally, and a second group of much more biased raters. Finally, we showed how different types of usernames (e.g., male versus female) led to different amounts of bias.

Thus, our take-away findings are:

- On average, authors are actually hurt by the presence of their name.
- "Average" and "Bad" authors are particularly hurt by the presence of their name, despite authoring content that is of roughly comparable quality to "good" authors.
- Authors with high follower counts reap the most benefit from their names being shown.
- Names that are male, organizations, and moderately topical see the most shift in the positive direction in the eyes of readers.

How then can these findings be used by system designers who wish to deliver the highest quality content to a user? For many end user scenarios, such as social search and friend recommendation, acknowledging the phenomenology of the user and *working with* bias may be most appropriate: if users actually think content is better simply by virtue of the names of certain users, then showing content from those users is a good idea. On the other hand, including some content from lesser known authors can elevate the quality of the system as a whole by promoting truly high quality content and authors. Finding the exact balance and method for showing end users content from names they expect and content free of name value bias is an area for future work.

One possibility relevant to presenting microblog content in search results pages is to provide the reader with information beyond the username and photo. That is, we have shown here that the username alone can lead to bias, so perhaps there are other pieces of information that could counter this bias. Of course, these pieces of information may induce their own biases, but in general we see exploring secondary information about users, including their bio, profile picture, and so on, as fruitful possibilities for inducing the fairest assessment of microblog authors and their content. Additionally, our work suggests the importance of the user names and a carefully crafted user name can lead to more desirable results for an author.

Microblogging systems might take name value bias into account in non-end user facing scenarios. For example, suppose a microblog system was attempting to compile a list of shared links, possibly in support of other analyses like search results ranking. The system could incorporate name value into its author weightings during this selection process by using follower counts as a proxy for name value. Given our results, this would mean a slight lowering of weights on users with high follower counts. We plan to explore other author metrics, or combinations of metrics, such as incorporating retweet rates, as proxies for name value.

Conclusion

In analyzing bias in microblogging due to name value, we show that some authors are hurt, others helped simply by virtue of their user name. Most raters were at least somewhat biased, and about 10% were strongly biased, indicating an effect of significant scale on the part of microblog consumers. Our findings could be incorporated by both system developers building interfaces as well as "backend" processes, such as when ranking content. The role of the username is particularly relevant as search over microblog content becomes more commonplace (Post 2010) and users increasingly are making judgments about content outside their personal feed. Future design work is needed to determine how to best present content end users will perceive as being of highest quality, both including and irrespective of bias due to name value.

Acknowledgments

We thank Microsoft Research for providing resources that facilitated the research. The first author performed this work while interning at Microsoft Research. We would also like to thank the anonymous reviewers for providing their valuable feedback.

References

Bechar-Israeli, H. 1995. From bonehead to clonehead: Nicknames, play, and identity on internet relay chat. *Journal of Computer-Mediated Communication* vol. 1, issue 2.

Cosley, D.; Lam, S. K.; Albert, I.; Konstan, J. A.; and Riedl, J. 2003. Is seeing believing?: how recommender system interfaces affect users' opinions. In *Proceedings of the 2003 Conference on Human Factors in Computing Systems, CHI 2003*, 585–592. ACM.

Howard, B. 2008. Analyzing online social networks. *Commun. ACM* 51:14–16.

Jacobson, D. 1999. Impression formation in cyberspace: Online expectations and offline experiences in text-based virtual communities. *Journal of Computer-Mediated Communication*.

Kittur, A., and Kraut, R. E. 2008. Harnessing the wisdom of crowds in wikipedia: quality through coordination. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work, CSCW 2008*, 37–46. ACM.

Kwak, H.; Lee, C.; Park, H.; and Moon, S. 2010. What is twitter, a social network or a news media? In *Proceedings of the 19th International Conference on World Wide Web, WWW 2010*, 591–600. ACM.

Lauw, H. W.; Lim, E.-P.; and Wang, K. 2006. Bias and controversy: beyond the statistical deviation. In *Proceedings of the Twelfth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 625–630. ACM.

Leskovec, J.; Huttenlocher, D. P.; and Kleinberg, J. M. 2010. Governance in social media: A case study of the wikipedia promotion process. In *Proceedings of the Fourth International Conference on Weblogs and Social Media, ICWSM 2010*. AAAI Press.

Pal, A., and Counts, S. 2011. Identifying topical authorities in microblogs. In *Proceedings of the 4th ACM International Conference on Web Search and Web Data Mining, WSDM 2011*. ACM.

Pal, A., and Konstan, J. A. 2010. Expert identification in community question answering: exploring question selection bias. In *Proceedings of the 19th ACM Conference on Information and Knowledge Management, CIKM 2010*, 1505–1508. ACM.

Post, H. 2010. Twitter user statistics revealed, http://www.huffingtonpost.com/2010/04/14/twitter-user-statistics-r_n_537992.html.

Suh, B.; Hong, L.; Pirolli, P.; and Chi, E. H. 2010. Want to be retweeted? large scale analytics on factors impacting retweet in twitter network. In *Proceedings of the 2010 IEEE Second International Conference on Social Computing, SocialCom / IEEE International Conference on Privacy, Security, Risk and Trust, PASSAT 2010*, 177–184. IEEE Computer Society.

Tversky, A., and Kahneman, D. 1974. Judgment under uncertainty: Heuristics and biases. *Science* 185(4157):1124–1131.