Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media

Write Like I Write: Herding in the Language of Online Reviews

Loizos Michael and **Jahna Otterbacher**
Open University of Cyprus
{loizos, jahna.otterbacher}@ouc.ac.cy

Abstract

Our behaviors often converge with those of others, and language within social media is no exception. We consider reviews of tourist attractions at TripAdvisor (TA), the world's largest resource for travel information. Unlike social networking sites, TA review forums do not facilitate direct interaction between participants. Nonetheless, theory suggests that language is guided by writers' conception of their audience, and that their style can shift in response. We implement a model of herding as a local transmission process, exploring the hypothesis that a reviewer is influenced by how preceding reviews manifest a given stylistic feature (e.g., pronouns, paralinguistic devices). We find that reviewers are more likely to use unusual features when such characteristics appear in their local context. The extent to which reviewers are influenced by context is correlated to attributes shared in their profiles, as well as their sentiment toward the attraction reviewed. Our results suggest that language can be influenced by others, even in an asynchronous environment with little to no interpersonal interaction. In other words, our behaviors may be susceptible to manipulation in social media; it may not always be the case that we write like ourselves.

Introduction

User review sites continue to be a popular social medium, allowing participants to share thoughts and opinions on an unlimited range of products, services and experiences. The current research focuses on the writing style participants use in English-language reviews of tourist attractions at TripAdvisor, a site that prides itself on being international in scope, operating in 34 countries.[1] TripAdvisor is an ideal case study for us to explore the dynamics of language in a social medium characterized by the diversity of its participants and its huge scale, yet that offers few opportunities for direct interaction or dialog between participants.

Our work focuses on patterns in the *stylistic* features of language, such as one's use of pronouns, lexical markers of dialect, and emoticons. We study *how* users describe attractions of interest rather than *what* they say. Stylistic features are particularly interesting in the context of social media because of their correlation to an individual's identity, as well as to relationships between communicators. Linguistic style varies depending on a range of social characteristics, such as socio-economic status or gender (Labov 1990), and authorship attribution techniques rely on this social (i.e., inter-speaker) variance in order to predict the likely author of a given text (Argamon et al. 2006).

Recent research has applied such techniques to social media. For instance, language behaviors in Twitter have been used to infer user characteristics including gender (Fink, Kopecky, and Morawski 2012), age (Nguyen et al. 2013) and even ethnicity (Pennacchiotti and Popescu 2011). Ironically, a possible challenge for these techniques is the social aspect of such media, and the extent to which people write in a manner that reflects their true selves. In the current work, we consider the possibility that online reviewers influence one another, such that stylistic features (e.g., the use of emoticons (Park et al. 2013)) are diffused through the process of reading and writing, resulting in herding behavior.

Could Review Style be Influenced?

Consider the following reviews contributed on a top Amsterdam attraction, The Jordaan neighborhood:

> We rented an apartment in The Jordaan for a week and enjoyed every bit of it! It is such a lively place and can be explored endlessly...

> We were told that this is the 'it' place to live in Amsterdam, and it definitely has charm. It's a nice place to stroll for an hour or so, with lots of little shops and places to grab a bite to eat.

These reviews represent a typical style in which the reviewers use the first and third person point of view, describing what the author ("we") did as well as the place itself ("it"). As will be detailed later, most reviews are written from one of these points of views. Only a minority (7%) of reviews uses the second person, in which reviewers address the audience directly, suggesting what "you" should do or see, such as in the following:

> These small streets are well worth the effort. You will be rewarded by taking the time to get lost in the small side streets where you gain a real picture of how the locals live...

[1] http://www.tripadvisor.com/PressCenter-c6-About_Us.html

> Amsterdam is a city full of tourists all year round. If you want to find a quieter and more peaceful part of this city, Jordaan is the place for you...

Given that a reviewer has no direct interaction with others at TA, and that the use of first and third person pronouns is the norm, what might influence a reviewer to incorporate this uncommon feature?

Contributions of the Current Work

We hypothesize that stylistic features of reviews exhibit herding behavior. We first establish a "house style" with respect to the use of 12 stylistic features, and identify features that are relatively rare (i.e., deviate from the "house style"). We then ask whether reviewers can be induced to adopt such unusual features. Our experiments demonstrate that when reviewers are exposed to these features by others in close context (i.e., in the immediately preceding reviews), they are significantly more likely to adopt them. Having established that herding can and does happen, we then explore the question of who is likely to be influenced by others. We uncover significant relationships between reviewer experience at TA, as well as sentiment toward the attraction, and the likelihood that the reviewer will follow others' behaviors.

Our work contributes to the growing literature on the language of social media, offering evidence that people's language can be influenced even in the absence of direct or repeated interaction with others. Methodologically, our empirical framework can easily be scaled up and applied to even larger data sets in order to further explore herding behaviors.

Background

In this section, we establish the underpinnings of our model of herding in order to study the dynamics of linguistic style.

Language Accommodation

When interlocutors interact in computer-mediated settings, socially constructed rules governing the language they use emerge over time. This is true of both dyadic (Scissors, Gill, and Gergle 2008; Danescu-Niculescu-Mizil, Gamon, and Dumais 2011) and group communication (Postmes, Spears, and Lea 2000; Cassell and Tversky 2005).

The social identity approach is often evoked to explain why a communicator adapts her style to that of others. Many argue that stylistic change takes place subconsciously, and that its underlying driver is social acceptance (Giles, Coupland, and Coupland 1991). The result is that a communicator's language converges toward the style of those with whom she identifies. Similarly, power differences may also explain language accommodation. The language of communicators of a lower social status is expected to converge to the patterns of more powerful or prestigious interlocutors (Herring 1996; Wang, Fussell, and Setlock 2009; Danescu-Niculescu-Mizil et al. 2012).

Interestingly, changes in language patterns of participants in online communities can reveal information about their relation to the group. For instance, changes toward / away from group linguistic norms correlate to a communicator's "lifestage" within the group; e.g., a newcomer in the process of acquiring a group's vocabulary versus an older participant who has stopped adjusting to group norms and may soon cease participating (Danescu-Niculescu-Mizil et al. 2013).

Similarly, in the context of TripAdvisor, we observe evidence of language convergence with respect to specific stylistic features. However, the social identity approach does not offer a satisfactory explanation for this phenomenon; little to no interpersonal interaction is fostered and it is unclear with which group or groups within TA a communicator might identify. Therefore, we propose a model of language herding, in which stylistic features might be transmitted between individuals without assuming that they will engage in repeated and/or direct interaction, and with no expectations that they identify as a member of as established group.

Audience Design

Bell's theory of audience design (Bell 1984) provides an explanation for style shifting in the absence of direct interaction. Bell claims that the communicator's perception of the audience, and her relationship to it, is the strongest factor influencing language style. While additional factors such as topic are likely to influence style, Bell argues that other factors are dominated by the audience effect.

In many situations, communicators do not have an established relationship with their audience but rather, experience brief (e.g., contact between a salesperson and a client) or indirect (e.g., communication via mass media) encounters. In such cases, Bell claims that a "house style" is in place. The communicator converges not to a specific interlocutor's style but rather, to her *notion* of the ideal client or reader. As evidence for this phenomenon, he cites Labov's study of salespersons in New York City department stores in which coworkers were found to have linguistic styles indistinguishable from one another (Labov 1972). Similarly, he cites his own study of the styles of newscasters across two radio stations (Bell 1982). While newscasters exhibited variance in their style, over time their language converged to two house styles, with the more prestigious radio station's style being more standard / formal than that of the other radio station.

Recent work resonates with Bell's theory. Marwick and boyd (2010) conducted a qualitative study of Twitter users, demonstrating some peculiar properties of social media with respect to audience: social media audiences are largely imagined by participants, and "context collapse" — where participants find a diverse set of people, from close friends to business contacts to those they have never even met, in their audience — is a unique property of social media. Hu et al. (2013) developed a computational model of social media language that incorporated a number of stylistic features. On a data set of texts from more "casual" media (e.g., Twitter, SMS, email) as well as more formal media (e.g., blogs, magazines, newswire), they demonstrated significant differences between the language style of Twitter and other media. That is, Twitter was found to have a unique "house style".

Herding in Review Language

"Herding" is a general term used to describe the convergence of behaviors without the influence of any central coordination. The behaviors of individuals align over time as a result

Behavior	Connections between individuals	Transmission mechanism
Wikipedia article deletion: Do editors' decisions on article "notability" exhibit evidence of information cascades? (Taraborelli and Ciampaglia 2010)	Votes are cast in Article for Deletion discussions. Wikipedia editors appoint themselves to participate in discussions and voting.	Votes cast are publicized. Each editor observes previous votes cast, rationalizing as to whether or not the article is notable enough to be kept.
Comment ratings: Are crowdsourced ratings of user-contributed comments on a news site biased by previous ratings and/or author identity? (Muchnik, Aral, and Taylor 2013)	Users establish networks of friends and enemies at the site. They may read and rate the comments previously posted.	User observes the scores on comments assigned by others, and whether the comment author is a friend, enemy or other user. User then decides whether and how to rate the comment.
Online book purchases: Can decisions be influenced by product reviews? (Chen et al. 2010)	No explicit structure. Consumers are exposed to reviews written by others on books of shared interest.	Consumer takes a rational purchasing decision, using heuristics such as ratio of positive/negative reviews, title and keywords.
Email cascades: Do decisions to forward chain emails exhibit evidence of reputational cascades? (Barton 2009)	Network of contacts established online (email senders and receivers) and offline (e.g., colleagues).	Recipient of an email takes a rational decision as to the quality of the message, using sender's reputation as a heuristic.
Emoticons on Twitter: Is the use of emoticons a social norm, transmitted from user to user? (Park et al. 2013)	A user's network of contacts with whom she has exchanged replies at least twice.	Once exposed to an emoticon, the user might adopt it in her own tweets. The mechanism might be non-mentalizing (e.g., emotional contagion) or mentalizing (e.g., social conformity).
Stylistic features in online reviews: Can reviewers be influenced to use unusual features by their presence in preceding reviews? (this work)	Structure inferred by local context. Reviewers are exposed to preceding reviews contributed about the attraction of interest (i.e., the most recent 10 reviews).	Undetermined. The mechanism might be non-mentalizing (e.g., contagion, priming) or mentalizing (e.g., linguistic accommodation via audience design).

Table 1: Examples of herding behaviors in social media.

of their local interactions with one another, not as the result of rules or directions. Table 1 provides examples of herding behaviors described in various social media contexts, as well as the herding behavior we presently study.

Raafat and Chater (2009) advocate modeling collective behavior at two levels: pattern (i.e., the structure of connections between individuals) and transmission (i.e., the mechanisms that govern the transfer of information between individuals). They claim that many approaches emphasize one level over the other. For instance, social network analysis stresses the patterns of connections between individuals. Of course, an individual is more likely to be influenced by others in her immediate environment, and is more likely to change her behavior the more sources of such influence there are (Latane 1996). However, we also must explain how individuals, who have received signals from others, then process and act on this information. More specifically, Raafat and Chater propose two key categories of transmission mechanisms: mentalizing and non-mentalizing. In mentalizing mechanisms, the individual consciously considers the information from others in order to form an impression of the herd. Non-mentalizing mechanisms are more emotional in nature and do not involve conscious rationalization.

In our model of herding in the language of reviews at TripAdvisor, the technical infrastructure of the medium and the communicative affordances it provides determine the structure of connections between individuals. There is no explicit social network — users do not make "friends" or "followers". TA's user interface displays reviews ranked in reverse chronological order, with ten reviews presented per page. Previous work showed that users presented with such a ranked list of items are unlikely to look past the first page of results; additionally, the time spent engaging each item decreases as one moves down the list (Pan et al. 2007). Therefore, it is reasonable to infer that a reviewer at TA will read a few reviews that are conveniently presented to her; she is unlikely to attend to a large number of items. For these reasons, our model examines the most recently posted reviews as the reviewer's local context that may influence her behavior.

As indicated in Table 1, while our present work provides evidence of herding, we are not yet able to establish with certainty what the transmission mechanism is. Our data is observational and we did not have access to TA reviewers. Therefore, we did not have a clear way to gauge users' processing of the information transferred to them via the reviews they read. We believe that audience design, a mentalizing transmission mechanism, is a plausible explanation for the behavior we observe. Our current experiments position us to further explore this mechanism in future work.

Data and Feature Selection

We identified 27 cities across three continents for which TripAdvisor provides a travel guide. Our analysis focuses on the "top-rated attractions" section of the guide, where users may add details on an attraction, rate it, as well as submit original reviews. Table 2 details the number of attractions

City	Attractions	Total reviews
Amsterdam	201	29,344
Boston	173	26,740
Budapest	179	18,907
Chicago	278	35,448
Delhi	122	12,503
Dubai	105	30,397
Dublin	262	26,573
Edinburgh	165	42,591
Hong Kong	326	27,059
Kyoto	308	7,140
Lisbon	165	9,941
London	820	160,845
Los Angeles	275	23,406
Madrid	308	15,945
Mexico City	221	5,466
Montreal	181	14,361
New York	616	133,325
Paris	509	142,103
Rio de Janeiro	268	9,371
Rome	558	46,020
San Francisco	281	39,600
Seattle	161	21,993
Singapore	240	32,128
Sydney	238	28,030
Toronto	208	21,165
Vienna	257	16,110
Washington D.C.	202	47,242
Total	7,627	1,023,753

Table 2: TripAdvisor data set.

and user-contributed reviews in our data set. We collected all available reviews on the 7,627 attractions, including the unstructured text (review title and body) as well as metadata on the review (e.g., date posted) and the reviewer (e.g., total number of reviews contributed).

Stylistic Features of Reviews

Table 3 describes the 12 stylistic features examined in our experiments, and shows their frequency in users' reviews.

Pronoun Use The manner in which people use pronouns reflects the norms of communication of the medium. For instance, second person pronouns are characteristic of spoken conversation (Yates 1996) and therefore are likely to be less common in more formal media (Hu, Talamadupula, and Kambhampati 2013). In addition, pronoun use tends to correlate to factors such as the communicator's age (Nguyen et al. 2013). In order to examine pronoun use in reviews, we developed four features: whether or not first person pronouns are used at all, whether or not second person pronouns are used at all, the dominant point of view (i.e., whether the plurality of pronouns is first, second or third person) and finally, whether the plurality of pronouns is singular or plural.

Informal Language Previous work has demonstrated that social media have different norms governing users' language style, with some media exhibiting more formal language than others (Hu, Talamadupula, and Kambhampati 2013). We examined four characteristics of informal language:

Feature	Frequency over all reviews	Frequency in / after first 10 reviews
(1) **Uses first person**		
1 - no first person pronouns appear in review	488,565 (48%)	42% / 48%
2 - one or more first person pronouns appear	535,188 (52%)	58% / 52%
(2) **Uses second person**		
1 - no second person pronouns appear in review	602,428 (59%)	54% / 59%
2 - one or more second person pronouns appear	421,325 (41%)	46% / 41%
(3) **Dominant point of view**		
0 - no plurality	479,848 (47%)	41% / 47%
1 - plurality of pronouns is first person	200,316 (20%)	20% / 20%
2 - plurality of pronouns is second person	75,415 (7%)	6% / 7%
3 - plurality of pronouns is third person	268,174 (26%)	33% / 26%
(4) **Singular vs. plural pronouns**		
0 - no plurality	390,094 (38%)	32% / 38%
1 - plurality of pronouns is singular	527,327 (52%)	58% / 51%
2 - plurality of pronouns is plural	106,332 (10%)	10% / 11%
(5) **All capital letters**		
1 - no instances	844,578 (82%)	77% / 83%
2 - one or more words written in all caps	179,175 (18%)	23% / 17%
(6) **Multiple punctuation marks**		
1 - no instances	854,623 (83%)	82% / 84%
2 - one or more instances	169,130 (17%)	18% / 16%
(7) **Internet slang & acronyms**		
1 - no use of slang or acronyms	1,011,582 (99%)	98% / 99%
2 - one or more instances	12,171 (1%)	2% / 1%
(8) **Emoticons**		
1 - no emoticons used in review	994,021 (97.1%)	96% / 97%
2 - one or more emoticons used	29,732 (2.9%)	4% / 3%
(9) **Informal language**		
1 - no informal characteristics	705,040 (69%)	64% / 69%
2 - one or more informal characteristics used	318,713 (31%)	36% / 31%
(10) **British vocabulary**		
1 - no markers of British English used	988,703 (97%)	97% / 97%
2 - at least one marker of British English used	35,050 (3%)	3% / 3%
(11) **Review length**		
1 - two or fewer sentences	264,484 (26%)	19% / 26%
2 - three or more sentences	759,269 (74%)	81% / 74%
(12) **Review length**		
1 - five or fewer sentences	725,161 (71%)	56% / 72%
2 - six or more sentences	298,592 (29%)	44% / 28%

Table 3: Stylistic features and their frequency in the data set.

- All capital letters in a word (e.g., "LOVED it!")
- Multiple punctuation marks (e.g., "definitely see this!!!")
- Internet slang and acronyms (e.g., "LOL", "asap", "fab")[2]
- Emoticons[3]

In addition, an aggregate measure, *informal language*, considers the use of any (at least one) of the above features. As seen in Table 3, the use of such informal stylistic features is rare; only one third of the total reviews exhibits them.

Markers of Dialect While based in the United States, TA attracts a worldwide audience. Therefore, an interesting question is whether other dialects of English are in use. We examine the use of markers of British English, considering the incidence of words and phrases that are uncommon in American English (e.g., use of "queue" rather than "line", "concession" rather than "reduction in price").[4] As expected, we find that markers of British English are relatively rare; only 3% of the reviews contained such language.

Review Length Finally, we consider whether reviewers write relatively shorter or longer reviews, and we experiment with two definitions of this feature. When we define a "short" review as consisting of two or fewer sentences, this characteristic is relatively infrequent. When we define a "short" review as consisting of five or fewer sentences, we observe long reviews (6+ sentences) more infrequently.

[2]http://www.netlingo.com/acronyms.php

[3]http://en.wikipedia.org/wiki/List_of_emoticons

[4]http://en.wikipedia.org/wiki/List_of_British_words_not_widely_used_in_the_United_States

Feature	Bins	Relative frequency after first 10 reviews
Attraction rating	0	0.03%
	1	1.31%
	2	2.14%
	3	8.08%
	4	30.56%
	5	57.88%
Reviewer experience (rank)	1	6.98%
	2	1.85%
	3-10	19.62%
	550+	0.15%
Review position (rank)	1-100	19.3%
	500+	55.6%

Table 4: Review and reviewer metadata.

Review and Reviewer Metadata

We also exploit three characteristics obtained from review and reviewer metadata in our experiments: the attraction rating (one to five; zero if unrated), the reviewer experience (expressed as the total number of reviews written at TA), and the review position in the sequence of reviews on the given attraction. Their distributions are summarized in Table 4.

Empirical Framework

Let a denote an attraction, and, in particular, think of a as specifying the sequence of reviews available for that attraction, with $a[n]$ denoting the chronologically n-th review. We use the symbol r to reference a generic review when its position and the attraction it belongs to need not be identified.

A feature f is a mapping from reviews to values. Let $\mathtt{val}(f)$ be the set of values to which reviews can be mapped. Examples of features relating to the style of language were presented in the preceding section. Additional features that we consider are the rating of a review r, the total number of reviews written by the author of r, and the position / index of a review among all reviews for a given attraction.

For a feature f and a review r, we write $f(r)$ to mean the value in $\mathtt{val}(f)$ of that feature for the particular review. Given a feature f and a value $v \in \mathtt{val}(f)$, we write $f^{-1}(v)$ to mean the set of all reviews that are mapped to v under f.

The L-context of a review, denoted by $\mathtt{cont}(a[n], L) = \langle a[n-L], \ldots, a[n-2], a[n-1] \rangle$, is the sequence of its preceding L reviews; the context is well-defined if $n > L$.

Given a sequence $s = \langle r_1, r_2, \ldots, r_m \rangle$ of reviews, denote by $f(s) = \langle f(r_1), f(r_2), \ldots, f(r_m) \rangle$ the sequence of the values of reviews in s under feature f. Given a sequence s of values, let $\sharp_v [s] = |\{n \mid s = \langle v_1, v_2, \ldots, v_m \rangle, v_n = v\}|$ denote the number of appearances of value v in sequence s. Given a sequence s of values, denote by $s_{a..b}$ the set of values that results by first dropping all duplicates and ordering s (in our work the values are always integers, so the ordering relation is the usual one over integers), and then selecting only the elements between positions a and b in this order, inclusive. The definitions apply also when s is a multi-set.

Give the fixed set A of attractions considered in this work, denote by $R^t = \{a[n] \mid a \in A, n > t\}$ the set of all reviews that appear at positions after t, thinking of t as an eligibility threshold for a review in an attraction to be considered.

Given a feature f, a value $v \in \mathtt{val}(f)$, three non-negative integers t, L, k, with $k \leq L \leq t$, and a set $R \subseteq R^t$ of eligible reviews, denote by $\mathtt{match}(f, v, L, k, R) = \{a[n] \in R \mid \sharp_v [f(\mathtt{cont}(a[n], L))] = k\}$ the multi-set of reviews in R that have matching contexts in terms of the frequency of appearance of value v under f in these contexts.

Methodology

We seek to determine the probability that a review exhibits a certain value under feature f_i. The probability is computed conditioned on the context of the review, and on the values that the given review exhibits under another feature g_j.

For our experiments we consider the stylistic features from Table 3, and denote them, respectively, by f_i for $i = 1, 2, \ldots, 12$. In addition, we consider the features g_1, g_2, g_3 such that $g_1(r)$ denotes the rating of review r, $g_2(r)$ denotes the total number of reviews written by the author of a review r (according to the author's profile), and $g_3(r)$ denotes the position n of review $r = a[n]$ for attraction a.

We denote by $R^t_j[u]$ the subset of reviews in R^t that belong in $g_j^{-1}(u)$, for $j = 1, 2, 3$ and $u \in \mathtt{val}(g_j)$; thus $R^t_j[u]$ includes those of the reviews in R^t that are mapped to the value u under feature g_j. By extension, we let $R^t_j[U] = \bigcup_{u \in U} R^t_j[u]$ for a given a set $U \subseteq \mathtt{val}(g_j)$ of values.

In our data analysis we compute and plot the quantity

$$\mathtt{Pr}(f_i, v, L, k, R) = \frac{\sharp_v [\mathtt{match}(f_i, v, L, k, R)]}{|\mathtt{match}(f_i, v, L, k, R)|}$$

for varying values of the parameters f_i, v, L, k, and for $R = R^t$. The computed quantity corresponds to the empirical probability that a review in R has value v (as opposed to some other value) under feature f_i. Not all reviews in R are considered, but only those preceded by k appearances of the value v (under feature f_i) in the preceding L reviews (the L-context of the review in question). This treatment is in line with our hypothesis: *the probability that a given feature value appears in a given review correlates with the frequency of appearance of that same feature value in preceding reviews*. Positive results in this regard will provide empirical evidence for the presence of herding behavior in how feature f_i is expressed in reviews.

In addition to the case of $R = R^t$ as above, we undertake further analysis by repeating the investigation after first restricting attention to reviews $R = R^t_j[U]$ that exhibit a certain value in U under feature g_j. Thus, we investigate whether the manner in which herding behavior is manifested in a review might be affected by factors other than (and presumably independent of) the context of the review.

Figure 1 summarizes the setting: Whether review $a[n]$ exhibits a certain value for linguistic feature f_i is influenced "horizontally" by the values of the review's context under f_i, and "vertically" by factors g_j pertaining to the review itself. Future work may investigate the shaded region, where a review's context is analyzed under the factors g_j as well.

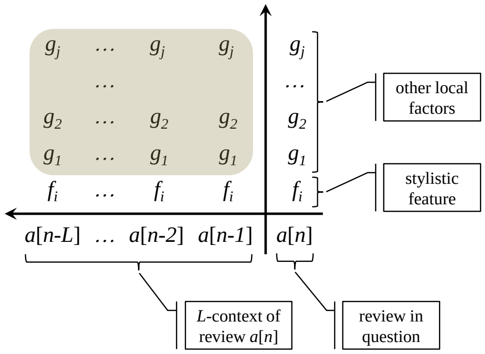

Figure 1: Context and factors affecting a review.

Empirical Results

Our empirical investigation concerns the case of $t = 10$, and spans the input space of function $\texttt{Pr}(\cdot,\cdot,\cdot,\cdot,\cdot)$. Selected results are illustrated in Figures 2–5, which correspond, respectively, to a choice of R to equal R^t or $R^t_j[U]$ for $j = 1, 2, 3$. Within each figure graphs are presented for various choices of values for i, v, and L. In particular, we present one graph for each value of i, choosing v to be that among the values of $\texttt{val}(f_i)$ that appears less often in $f_i(R^t)$, according to Table 3. L takes values in $\{1, 4, 7, 10\}$. Since larger values of L lead to fewer applicable data points from which $\texttt{Pr}(\cdot,\cdot,\cdot,\cdot,\cdot)$ is computed, we lower the value of L for those values of i and v where the data is scarce to begin with, so that the results remain statistically significant.

The value of k varies along the horizontal axis of each graph, while the quantity $\texttt{Pr}(\cdot,\cdot,\cdot,\cdot,\cdot)$ is measured along the vertical axis. Error bars correspond to the 95% confidence interval of the plotted values, assuming (as done typically) that samples are drawn from a normal distribution. Large error bars indicate settings where the presence of a feature was rare, and where, despite the lowering of the value of L, the data did not support statistically significant results.

Within each graph in Figures 3–5, different lines correspond to different choices of U. In some cases U is chosen explicitly as a set of values from $\texttt{val}(g_j)$. In other cases we find it easier to choose U in terms of the values of $\texttt{val}(g_j)$ that actually appear in our data. Thus, we use $g_j(R^t)$ to get these values, and then restrict attention to a subset $g_j(R^t)_{a..b}$ of these values according to their ordered position in $g_j(R^t)$. For example, $g_j(R^t)_{1..100}$ for $j = 3$ means the first 100 values of $g_3(R^t)$, which, since $t = 10$, actually correspond to values $\{11, 12, \ldots, 110\}$. This choice was made as a matter of convenience, with no bearing on the actual results.

We performed our analysis using the computational engine Mathematica 9.0. The code written is highly vectorized (i.e., relies on using vector operations rather than for-loops), and amenable, thus, to high parallelization. In addition, the computed quantity can be derived in time linear in the length of the data set. Overall, then, our approach can scale up to larger data sets if the need arises to do so for future work.

Discussion

The empirical framework was designed to investigate: *(i)* the existence of a "house style", *(ii)* the presence of herding behavior when deviating from this "house style", *(iii)* the effect of review-specific information on how such herding behavior is manifested. Each of these hypotheses is investigated for each of the stylistic features presented in Table 3.

(i) Is there empirical evidence of a "house style" with respect to incorporating stylistic features in textual reviews?

Table 3 shows that for most features considered, a certain choice among its possible values is predominant in terms of frequency of appearance in reviews. The use of first person pronouns ($i = 1$) might be taken to be an exception to the overall theme, although we point out that the frequencies of using versus not using first person pronouns are measurably apart, and their separation is statistically significant.

In line with the choice of the rarest feature value v to be used for our empirical analysis in Figures 2–5, it is useful for the rest of this section to think of "house style" as meaning "avoiding the use of the rarest feature value". The aim, then, is to examine whether and when this statement is falsified. In particular, it is not our intent to suggest that the particular "house style" considered is unique to TA, nor is the truth or falsity of this uniqueness property important for our results.

(ii) Is there empirical evidence of herding behavior with respect to incorporating stylistic features in textual reviews?

Evidence of herding behavior was observed for all of the stylistic features we considered, although the herding behavior is more pronounced for some of the features than for others; see Figure 2. In the case of using Internet slang and acronyms ($i = 7$) only a borderline herding behavior was observed. We would attribute this outlier in the overall theme to the very scarce data that were available for this feature, and suggest that further investigation on a much larger data set is warranted to derive safe conclusions for this feature.

Overall, there is a statistically significant increase in the probability that stylistic features will be incorporated into a given review in the same manner in which they were manifested in the preceding reviews. This is particularly interesting for features that are relatively unusual across the data: e.g., the use of second person point of view, plural pronouns, and short (≤ 2 sentences) or long (≥ 6 sentences) reviews.

Also of importance is the observed phase-transition (i.e., sharp non-linearity) on how local context affects the style of a review. For practically all stylistic features, the probability of deviating from the "house style" increases in a sigmoid fashion. This is particularly visible in the dominant voice used ($i = 3, 4$), where once a super-majority (typically somewhere between 70%–90%) of the reviews in the local context deviate from the "house style" there is a sharp increase in the probability that the next review will follow suit. In the case of many features this sharp increase leads to the probability of deviation becoming practically 1; e.g., in the dominant voice used ($i = 3, 4$), the use of atypical capitalization and punctuation ($i = 5, 6$), the use of informal language ($i = 9$), and the length of reviews ($i = 11$). The trend is clear even for the remaining features where this complete compliance is not observed, and we expect that given more data and the ability to examine sufficiently larger contexts (i.e., values of L) we would still get a complete compliance.

(iii) Might 'X' affect the extent to which reviewers are influenced by the writing styles of others in their local context?

Having established that reviews deviate from a "house style" depending on their local context, we now consider whether certain factors may affect the degree of this dependence, how easily deviation is triggered, how pronounced its effects are, etc. The results are illustrated in Figures 3–5.

We consider three factors 'X': review valence — reviewer sentiment toward the attraction, as expressed, in particular, through their rating ($j = 1$); reviewer experience — the number of reviews written in total by the reviewer ($j = 2$); and review placement — the position of the review in the sequence of reviews for a given attraction ($j = 3$). Rather than detailing the observed behavior in each of the 12 features (values of i), and for each of the three factors (values of j), we offer here some high level observations, and some possible explanations for the observed behaviors.

For most of the feature-factor combinations, reviews within each given group have a roughly monotone increase in adopting a feature value v when the local context heavily adopts v, following, in essence, the mean behavior shown in Figure 2. In a large subset of those cases, different groups exhibit statistically significant differences in their behavior.

Two broad categories of differences can be identified. The first category of differences is when a group exhibits a larger propensity towards having a review affected by its local context, with this propensity remaining roughly constant across different contexts. This phenomenon can be seen clearly for features $i = 5, 6, 9$ in Figure 3, where the plotted lines for the various groups are roughly parallel to each other, yet measurably apart. The second category of differences is when a group has a varying degree of propensity to be affected in relation to other groups. This phenomenon can be seen clearly for features $i = 1, 11$ in each of the Figures 3–5. Both categories of differences can be ultimately attributed to a general theme: each group exhibits its own sigmoid-like reaction of review style to their local context. Feature $i = 4$ in Figure 5, for instance, shows clearly the two sigmoid functions with a common origin and destination, but with a different phase-transition point. The same phenomenon is much more striking for feature $i = 11$ in Figure 4, where the phase transition for the first (blue) group is around 20%, whereas for other groups it is around 95%.

Future Investigation

To venture some general claims guided by our empirical results, it appears that reviewers with more extreme sentiments, reviewers with less experience, and reviewers writing the earlier reviews for an attraction, are more attuned to their local context, and more eager to adjust their writing style to follow suit. One could offer intuitive explanations as to why these phenomena arise. It could be argued that inexperienced reviewers, lacking a personal roadmap on the "expected" writing style, adopt and adapt to what others do. It could also be argued that earlier reviews are deemed to have more novel content, and so need be better stylistically-calibrated to convey their message.

Both aforementioned arguments rest on the premise that audience design is a possible explanation for the herding behavior that we observe. Future work may explore this question further, examining the patterns of individuals to see if and when their behaviors change, and what triggers this. This would enable us to describe more precisely the transmission mechanism behind language herding in this context.

It has been argued that style also changes with life experience (Nguyen et al. 2013). To the extent that relevant life experiences are recorded in a reviewer's profile, this information could be taken into account as a new factor (feature g_j) for further empirical work. Our work has, in fact, touched upon a — perhaps minor, but nonetheless relevant — type of life experience, namely the number of reviews previously written by the reviewer. We have found it easier in this work to approximate the number of *previously* written reviews by the *total* number of reviews, as this is the information that is available in the profile of a reviewer. Future work can compute directly the number of interest not by looking into the profile of the reviewer, but by identifying for each reviewer the ordered sequence of just her own-written reviews.

This last point relates to shifting attention from attractions as the entities of interest, to people. Such a shift would support investigating how a reviewer's personal style changes over time, and in particular whether it converges on average to the "house style", even if it subsequently momentarily deviates from it as a result of information in the local context.

In a direction orthogonal to the above, future work may investigate whether mechanisms other than herding could offer alternative, plausible explanations for our data. One could consider whether external contexts (e.g., attraction-specific or seasonal characteristics) or even some random process may lead reviewers to adapt their language style in a manner that would give rise to bursts of deviations from the "house style". On the empirical front, we expect that the effects of external contexts will be secondary to the effects of local contexts, particularly since we consider stylistic, rather than content, features in reviews. On the methodological front, it remains an interesting question whether a general methodology can be used to distinguish the effects of external contexts from the effects of local contexts, while using only observational data. While some researchers argue that it is "nearly impossible" to distinguish genuine herding behaviors from uninfluenced behaviors using observational data (Muchnik, Aral, and Taylor 2013), others have demonstrated that under certain conditions, it might be possible to infer causal relationships in such settings (Pearl 2000).

Conclusion

To our knowledge, we have presented the first study of herding behavior in language in a social medium where little to no direct interaction is facilitated. We have shown that herding might happen through the process of reading and writing content alone, without assuming that people identify with one another or the larger group, as in the case of more tightly-knit online communities. Given the popularity of quasi-interactive media such as online reviews, it is no surprise that there is much interest in mining this data to make inferences about people and their preferences. Here, our work offers one word of caution. It may not be the case that our writing style is consistent and specific to ourselves; instead, it could be influenced by others in unexpected ways.

References

Argamon, S.; Koppel, M.; Fine, J.; and Shimoni, A. R. 2006. Gender, Genre, and Writing Style in Formal Written Texts. *Text & Talk* 23(3):321–346.

Barton, A. M. 2009. Application of Cascade Theory to Online Systems: A Study of Email and Google Cascades. *Minnesota Journal of Law, Science & Technology* 10(2):473–502.

Bell, A. 1982. Radio: The Style of News Language. *Journal of Communication* 32(1):150–164.

Bell, A. 1984. Language Style as Audience Design. *Language and Society* 13(2):145–204.

Cassell, J., and Tversky, D. 2005. The Language of Online Intercultural Community Formation. *Journal of Computer-Mediated Communication* 10(2).

Chen, M.; Ma, Q.; Li, M.; Dai, S.; Wang, X.; and Shu, L. 2010. The Neural and Psychological Basis of Herding in Purchasing Books Online: An Event-Related Potential Study. *CyberPsychology, Behavior, and Social Networking* 13(3):321–328.

Danescu-Niculescu-Mizil, C.; Lee, L.; Pang, B.; and Kleinberg, J. 2012. Echoes of Power: Language Effects and Power Differences in Social Interaction. In *Proceedings of the 21st International Conference on World Wide Web*, 699–708.

Danescu-Niculescu-Mizil, C.; West, R.; Jurafsky, D.; Leskovec, J.; and Potts, C. 2013. No Country for Old Members: User Lifecycle and Linguistic Change in Online Communities. In *Proceedings of the 22nd International Conference on World Wide Web*, 307–318.

Danescu-Niculescu-Mizil, C.; Gamon, M.; and Dumais, S. 2011. Mark My Words!: Linguistic Style Accommodation in Social Media. In *Proceedings of the 20th International Conference on World Wide Web*, 745–754.

Fink, C.; Kopecky, J.; and Morawski, M. 2012. Inferring Gender from the Content of Tweets: A Region Specific Example. In *Proceedings of the 6th International AAAI Conference on Weblogs and Social Media*.

Giles, H.; Coupland, N.; and Coupland, J. 1991. Accommodation Theory: Communication, Context, and Consequences. In Giles, H.; Coupland, J.; and Coupland, N., eds., *Contexts of Accommodation. Developments in Applied Sociolinguistics*. Cambridge: Cambridge University Press.

Herring, S. 1996. Posting in a Difference Voice: Gender and Ethics in Computer-Mediated Communication. In Ess, C., ed., *Philosophical Perspectives on Computer-Mediated Communication*. Albany, NY: State University of New York Press. 115–145.

Hu, Y.; Talamadupula, K.; and Kambhampati, S. 2013. Dude, srsly?: The Surprisingly Formal Nature of Twitter's Language. In *Proceedings of the 7th International AAAI Conference on Weblogs and Social Media*.

Labov, W. 1972. *Sociolinguistic Patterns*. Philadelphia, PA, USA: University of Pennsylvania Press.

Labov, W. 1990. The Intersection of Sex and Social Class in the Course of Linguistic Change. *Language Variation and Change* 2(2):205–254.

Latane, B. 1996. Dynamic Social Impact: The Creation of Culture by Communication. *Journal of Communication* 46(4):13–25.

Marwick, A., and danah boyd. 2010. I Tweet Honestly, I Tweet Passionately: Twitter Users, Context Collapse, and the Imagined Audience. *New Media & Society* 13(1):114–133.

Muchnik, L.; Aral, S.; and Taylor, S. J. 2013. Social Influence Bias: A Randomized Experiment. *Science* 341:647–651.

Nguyen, D.; Gravel, R.; Trieschnigg, D.; and Meder, T. 2013. "How Old Do You Think I Am?" A Study of Language and Age in Twitter. In *Proceedings of the 7th International AAAI Conference on Weblogs and Social Media*.

Pan, B.; Hembrooke, H.; Joachims, T.; Lorigo, L.; Gay, G.; and Granka, L. 2007. In Google We Trust: Users' Decisions on Rank, Position, and Relevance. *Journal of Computer-Mediated Communication* 12(3):801–823.

Park, J.; Barash, V.; Fink, C.; and Cha, M. 2013. Emoticon Style: Interpreting Differences in Emoticons Across Cultures. In *Proceedings of the 7th International AAAI Conference on Weblogs and Social Media*.

Pearl, J. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge, United Kingdom: Cambridge University Press.

Pennacchiotti, M., and Popescu, A.-M. 2011. A Machine Learning Approach to Twitter User Classification. In *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media*.

Postmes, T.; Spears, R.; and Lea, M. 2000. The Formation of Group Norms in Computer-Mediated Communication. *Human Communication Research* 26(3):341–371.

Raafat, R. M., and Chater, N. 2009. Herding in Humans. *Trends in Cognitive Sciences* 13(10):420–428.

Scissors, L. E.; Gill, A. J.; and Gergle, D. 2008. Linguistic Mimicry and Trust in Text-based CMC. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work*, 277–280.

Taraborelli, D., and Ciampaglia, G. L. 2010. Beyond Notability. Collective Deliberation on Content Inclusion in Wikipedia. In *Proceedings of the 4th IEEE International Conference on Self-Adaptive and Self-Organizing Systems*, 122–125.

Wang, H.-C.; Fussell, S. F.; and Setlock, L. D. 2009. Cultural Difference and Adaptation of Communication Styles in Computer-mediated Group Brainstorming. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 669–678.

Yates, S. J. 1996. Oral and Written Linguistic Aspects of Computer Conferencing: A Corpus Based Study. In Herring, S. C., ed., *Computer-mediated Communication*. Amsterdam, The Netherlands: John Benjamins Publishing Company.

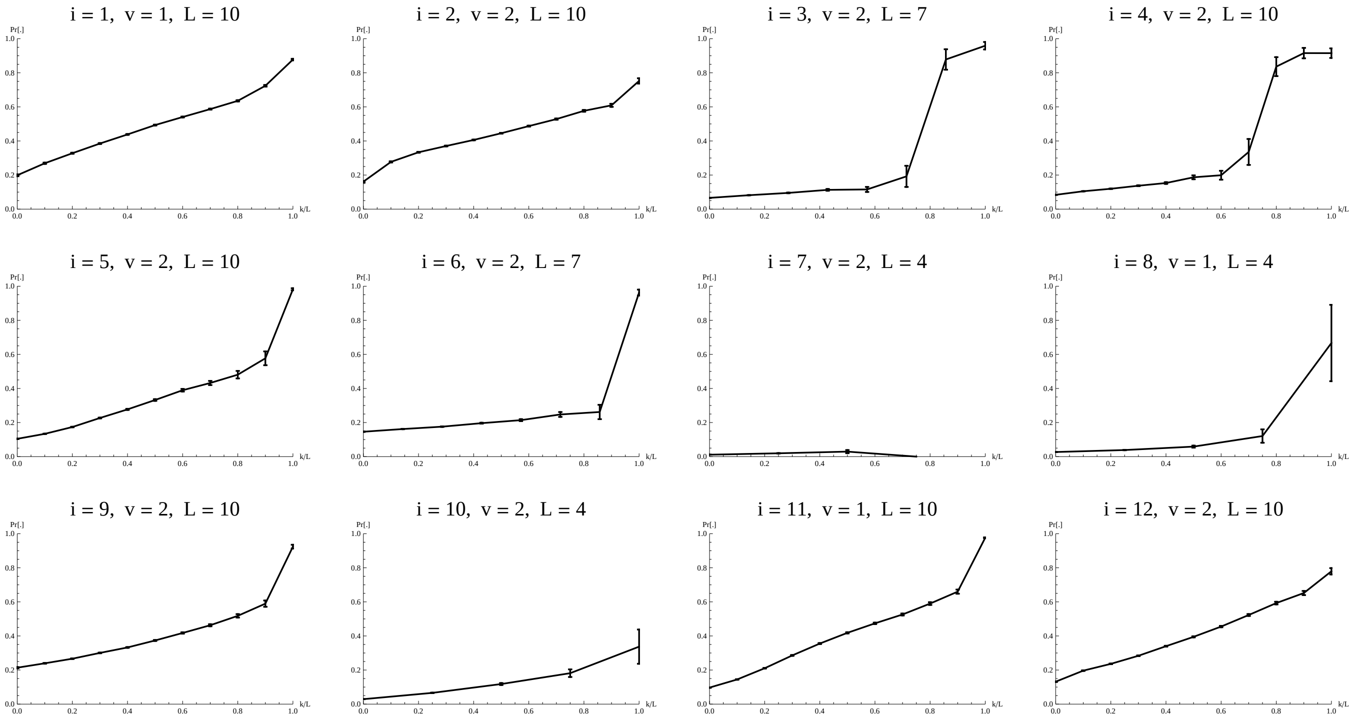

Figure 2: Plot of $\mathtt{Pr}(f_i, v, L, k, R^t)$ when $t = 10$. The values of i,v are interpreted as in Table 3.

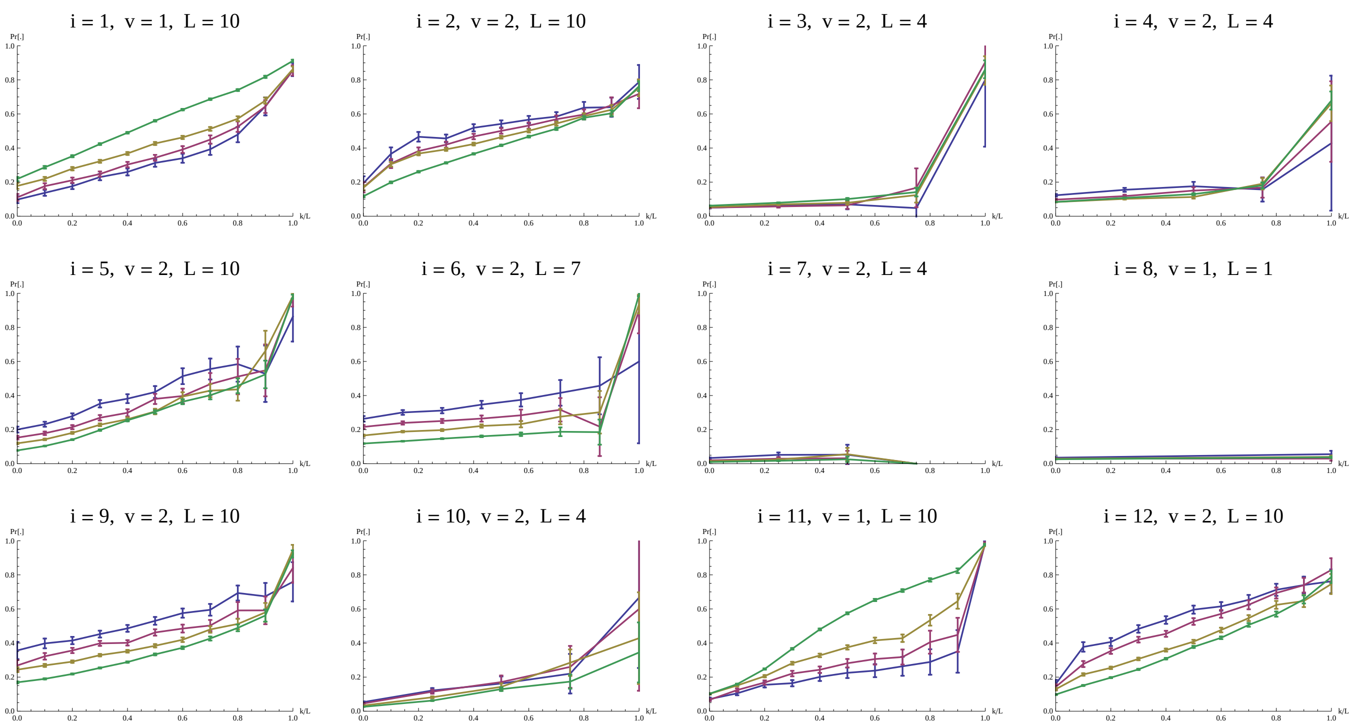

Figure 3: Plot of $\mathtt{Pr}(f_i, v, L, k, R_j^t[U])$ when $t = 10$ and $j = 1$ (grouped by review's rating). The values of i,v are interpreted as in Table 3. Colored lines correspond to: $U = \{2\}$ for blue; $U = \{3\}$ for red; $U = \{4\}$ for yellow; $U = \{5\}$ for green.

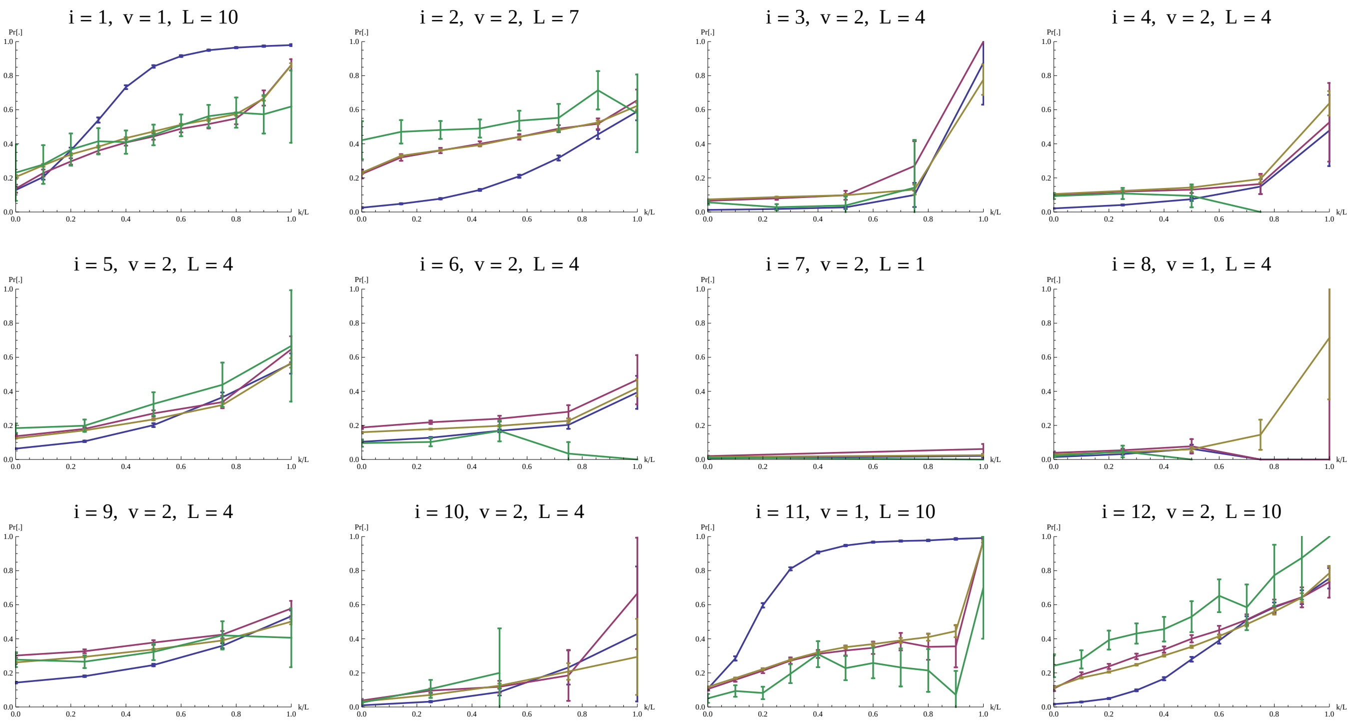

Figure 4: Plot of $\mathtt{Pr}(f_i, v, L, k, R_j^t[U])$ when $t = 10$ and $j = 2$ (grouped by reviewer's total reviews). The values of i,v are interpreted as in Table 3. Colored lines correspond to: $U = g_j(R^t)_{1..1}$ for blue; $U = g_j(R^t)_{2..2}$ for red; $U = g_j(R^t)_{3..10}$ for yellow; $U = g_j(R^t)_{550..\infty}$ for green.

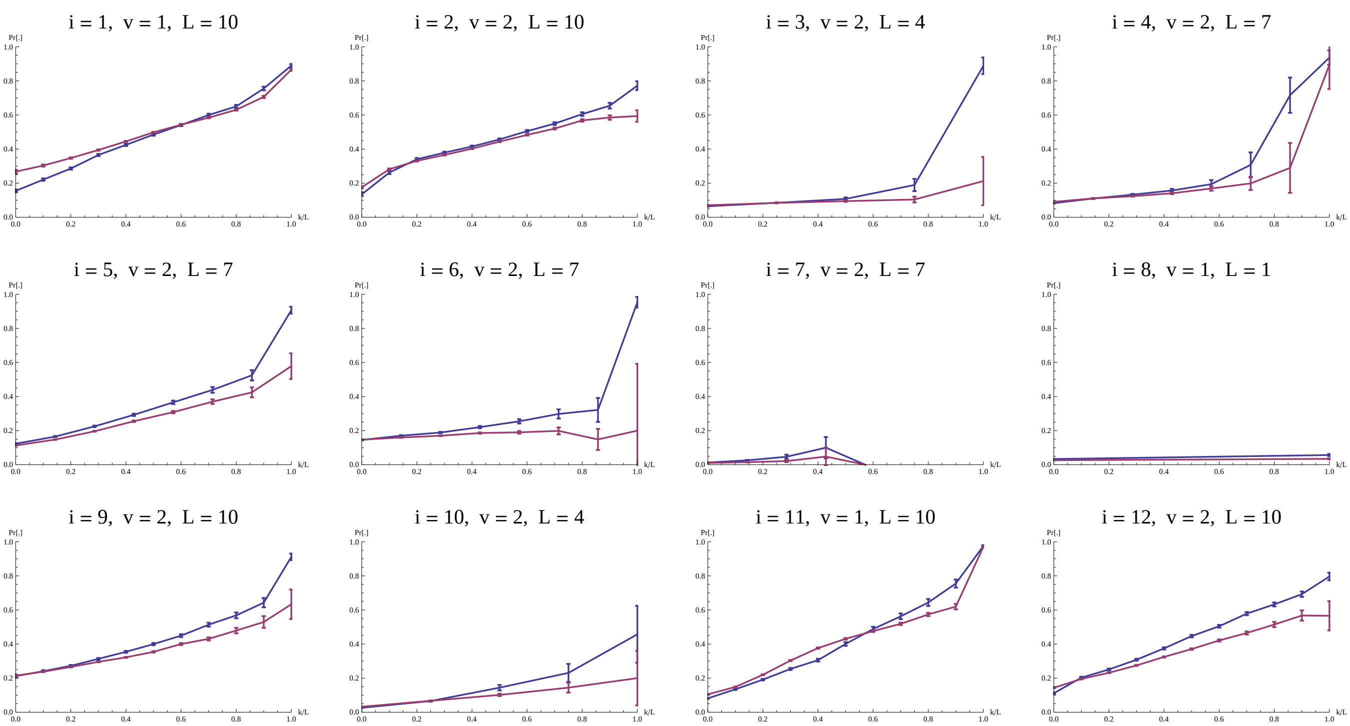

Figure 5: Plot of $\mathtt{Pr}(f_i, v, L, k, R_j^t[U])$ when $t = 10$ and $j = 3$ (grouped by review's position). The values of i,v are interpreted as in Table 3. Colored lines correspond to: $U = g_j(R^t)_{1..100}$ for blue; $U = g_j(R^t)_{500..\infty}$ for red.