The Journal of Web Science, 2020, 8: 1–12

Toward Understanding Online Impression Management: How Twitter Users Control Textual Expressions Over Time

Tomu Tominaga[1,a], Yoshinori Hijikata[2] and Naomi Yamashita[3]

[1] *Graduate School of Engineering Science, Osaka University, Japan, tominaga.tomu@gmail.com*

[2] *School of Business Administration, Kwansei Gakuin University, Japan, contact@soc-research.org*

[3] *NTT Communication Science Laboratories, Japan, naomiy@acm.org*

ABSTRACT

Impression management on social networking sites is becoming more important as people live in an increasingly connected world where they must initialize, develop, and maintain relationships with others online. Previous studies have shown that people form impressions differently depending on their relationship with their audience. However, few studies have focused on the longitudinal aspect of how people manage their impressions by controlling their expressions over time according to the audience. In this study, we investigated temporal changes in textual expressions (e.g., neurotic words) and then analyzed how such changes were related to a person's audience size (i.e., followers), density (i.e., mutual connections), and feedback (e.g., Likes). An analysis of 5 million posts collected from 1.6 thousand Twitter users over a period of 2.5 years revealed that users who had developed more mutual connections with their audience tended to use more neurotic and conscientious expressions. Meanwhile, users who received more Likes from their audience wrote fewer neurotic or conscientious expressions. Our findings indicate that Twitter users gradually adjust their use of expressions through their interactions with audiences, which may ultimately change the impressions that others have of them.

Keywords: Social networking sites, Twitter, Impression management, Audience

ISSN 2332-4031; DOI 10.1561/106.XXXXXXXX

1 INTRODUCTION

It is challenging for people in online spaces to adequately engage in impression management, which is the act of presenting oneself in a certain way to portray a desired image to the audience (Goffman, 1959). When managing impressions, people seek to gain benefits (e.g., gaining romantic partners on online dating sites (Zytko *et al.*, 2014b; Kapidzic, 2013) or making connections with friends on social networking sites (SNSs) (Ellison *et al.*, 2007)) and to avoid risks (e.g., losing a job (Wang *et al.*, 2011) or privacy (Gross *et al.*, 2005)) at the same time. Understanding how people form and maintain impressions on existing SNSs can provide insights for designing online platforms that allow people to better balance these benefits and risks.

Previous research on online impression management has revealed that people engage in different self-presentation strategies depending on their audiences. For example, when SNS users have a large audience, they tend to create more wall posts to maintain relationships with others (Rui and Stefanone, 2013) or share useful information to increase their visibility (Marwick and Boyd, 2011; Naaman *et al.*, 2010). If SNS users have a denser network with their audiences, they often express feelings of negative self-worth (e.g., *"feeling unloved"*) to obtain supportive comments from their friends (Burke and Develin, 2016). Moreover, when SNS users receive comments soon after joining an SNS, they tend to create many posts (Burke *et al.*, 2009). These findings show that the expressions that people use on SNSs are influenced by the size and density of their audience and the feedback they receive from them, which suggests that such changes in expressions may change the impressions that the audience has of the user.

However, most of these findings were derived from snapshot data collected at a specific time. Therefore, little is known about temporal change of users' expressions. In other words, we still lack an understanding of how people change their expressions and manage their impressions over time in response to changes in their audience size, density, and feedback. To design a sustainable social networking platform that helps people better manage their online impressions, it is important to gain a better understanding of how people engage in online impression management over a longer time frame.

Thus, we decided to explore the following research question: ***"How do SNS users manage their impressions by controlling their expressions over time according to changes in the size and density of their audiences and the feedback from the audiences?"*** By addressing this research question, we aim to obtain novel insights into the longitudinal aspects of online impression management.

To explore our research question, we studied 5 million Twitter posts collected from 1.6 thousand Twitter users that had been posted over 2.5 years. Using the collected data, we examined Twitter user changes in their use of expressions to

[a] The author is currently working at NTT Corp., Japan.

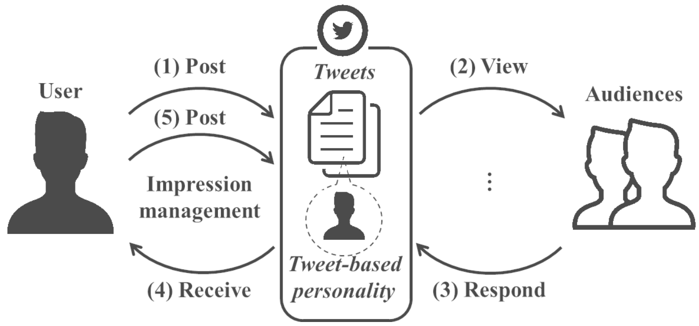


Figure 1: A diagram of impression management through interactions with audiences. (1) Users post a tweet, (2) audiences view the tweet, (3) audiences respond to the tweet, (4) users receive the responses, and (5) users post another tweet. Users manage their own impressions through the cycle of posting tweets while receiving responses from their audiences. A users' impression is personal characters that are formed from his/her tweets, which we call *tweet-based personality*.

see how they managed their impressions in association with changes in size and density of their audiences, and feedback from their audiences.

Based on literature reviews of online and offline impression management (Goffman, 1959; Marwick and Boyd, 2011), we developed a conceptual framework of impression management on Twitter (Figure 1). In the framework, we assumed that Twitter users manage their online impressions by creating posts while receiving signals of how audiences respond to their posts. In this study, we defined a user's impression as their provisional personality, a definition, which has been used in previous research (Vazire and Gosling, 2004) to verify whether impressions are conveyed to others. We also assumed that a provisional personality was created and inferred from the textual expressions in the user's posts. This is because online impressions are mainly formed from the users' content (Gosling *et al.*, 2011). We focused on the provisional personality projected in tweets, which we hereafter refer to as the *tweet-based personality.*

To observe users' tweet-based personalities, we followed the Five-Factor Model (McCrae and Costa, 1987). We measured this personality from expressions in Twitter posts using a computational personality prediction technique (IBM, 2017 (visited)). This prediction technique enabled us to analyze how users, regardless of their intent, expressed their personalities in their posts, and how these presentations were likely to be perceived by their audiences.

To measure the size and density of audiences, we used the number of followers and the ratio of mutual-following users, respectively. To quantify the feedback users receive from audiences, we focused on the number of replies, retweets, and Likes.

Using these measures, we identified two tweet-based personalities for each user: one for the past and one for the present (i.e., at the point of data collection). Tweet-based personality for the past was calculated from their tweet content posted approximately 2.5 years ago, which was one month after they started using Twitter, and their tweet-based personality at present was calculated from their tweet content posted at the point of our data collection. We then observed the within-user changes in the tweet-based personality from the past to the present and analyzed how the changes were associated with their audience size, density, and feedback.

Our primary results demonstrated that users with more mutual connections with their audiences were more likely to use neurotic or conscientious expressions, whereas users who received more Likes from their audiences had the opposite trend in their use of neurotic and conscientious languages. We also found that the users with more mutual connections tended to use more extraverted and agreeable expressions, which are characteristics associated with a sociable personality. To the best of our knowledge, this is the first study to quantitatively investigate online impression management in the long term. Our findings provide insights for developing impression management tools that provide users with feedback about their expressed personality.

2 RELATED WORK

2.1 Impression Management in Online Environments

Researchers have found that people engage in impression management in online (Dominick, 1999; Zytko *et al.*, 2014a; Zhao *et al.*, 2013) and offline settings (Goffman, 1959; Braginsky *et al.*, 1966). In both settings, the purpose of managing impressions is to portray a particular, desired image to other people (Goffman, 1959). However, the means of managing impressions in online settings is usually different from that in offline settings. This is because the environmental features of online settings differ from those of offline settings, which affects online impression management.

Some features of the online environment facilitate online impression management. For example, anonymity allows people to exaggerate their status when managing impressions. One specific example is that on online dating sites, men are more likely to exaggerate their height, whereas women are more likely to report their weight as lower than it is (Hancock *et al.*, 2007). Furthermore, SNS users selectively share their profile photos so that others see them as attractive (Deeb-swihart *et al.*, 2017; Kapidzic, 2013). As such, anonymity provides users with a greater chance of presenting themselves differently than how they are.

Another facilitatory feature of SNSs is asynchronicity. This feature enables people to edit the information that is transferred to others for an almost unlimited time (Walther, 2007) to find the optimal way of presenting themselves (Sunnafrank, 1986). On online dating sites, users often take care of small cues such as misspellings or the length of their messages because they aim to be perceived as educated or deliberate (Ellison *et al.*, 2006). SNS users edit their messages even after making posts when they care about those who can see the posts (Wang *et al.*, 2014). In the online asynchronous environment, users can manage their impressions more carefully and politely than they can in in-person, offline environments.

In contrast to anonymity and asynchronicity, the audience can be a restrictive feature for online impression management. This is because online audiences are more diverse than offline audiences, and they range from close friends to strangers (Litt and Hargittai, 2016; Vitak, 2012). Thus, when managing impressions by making posts that are publicly shared with such audiences, it is difficult for users to meet the standards of all audience members at once (Binder *et al.*, 2009; Sleeper *et al.*, 2013). To overcome this difficulty, users take several strategies for managing impressions. For example, some users abstain from self-expression to meet the strictest standards of their audience (Marwick and Boyd, 2011) by removing undesired content (Lampinen *et al.*, 2009; Lang and Barton, 2015; Sleeper *et al.*, 2013). On SNSs, users withdraw from making posts or comments when their content may sound negative to a specific part of their audience (Lampinen *et al.*, 2009; Sleeper *et al.*, 2013). Alternatively, other users manage their impressions only for sections of their audience that provide the most influential gains or losses (Marder *et al.*, 2016). For example, users might post content to seek help about trouble they are experiencing at work, even though they understand that this content may disturb their family members, if they have a strong motive to solve the problems (i.e., their gains).

To balance the facilitation and restriction of impression management in an online environment, users monitor signals from the audience. On SNSs, users usually see who is in their audience and how they react with their content. In the next subsection, we review prior work on how SNS users manage their online impressions while interacting with the audience.

2.2 Effect of Audience on Online Impression Management

Interaction with audiences when managing online impressions is highly related to the concept of an *imagined audience*, which is defined as a mental conceptualization of the people with whom users are communicating (Litt, 2012). Researchers have shown that impression management in SNSs varies by who and how many people users imagine are following their posts (Vitak, 2012; Rui and Stefanone, 2013; Marwick and Boyd, 2011; Tice *et al.*, 1995).

For example, Facebook users who imagined their audiences to be rich in diversity engaged in self-protective behaviors (Vitak, 2012), such as asking friends to delete wall posts that they disliked (Rui and Stefanone, 2013). Alternatively, Twitter users with public accounts showed a different trend: they shared more intimate, personal, and private information when they had more diverse groups of followers in their audience (Choi and Bazarova, 2015).

Concerning the size of an imagined audience, Rui and Stefanone (2013) found that Facebook users with larger audiences tended to manage their impressions more actively through multiple photo sharing and wall posting. Facebook users also sometimes refrained from posting messages about their private experiences when they thought that these messages would sound negative to their audience (Sleeper *et al.*, 2013). On Twitter, users with smaller audiences tended to post tweets that focused on themselves (to some extent, contrary to Choi and Bazarova (2015)), whereas users with more followers tended to share information that was useful for their audiences (Marwick and Boyd, 2011; Naaman *et al.*, 2010).

Similar to the composition and size of an imagined audience, previous studies have shown that feedback from an audience also affects the ways of presenting information in SNSs (Burke *et al.*, 2009; Liu and Brown, 2014). On Facebook, newcomers tend to post visual content more actively after they received many comments on their photos during the initial two weeks (Burke *et al.*, 2009). Likewise, within Renren (a Chinese SNS), the amount of content on profile pages was positively associated with the perceived frequency of receiving positive comments from others (Liu and Brown, 2014).

Although receiving feedback from audiences is generally related to active engagement, receiving Likes may not relate to active postings. Previous research has shown that Facebook users did not feel any particular excitement when receiving Likes from their audiences (Cheikh-Ammar and Barki, 2014). As a result, the number of Likes was not associated with active production of posts (Cheng *et al.*, 2014).

In sum, previous studies have shown that the ways people customize and present information to form online impressions are affected by audience-related factors such as size, density, or feedback. However, although most of these studies have focused on different methods of online impression management of different users, few have investigated the temporal changes of impression management within the same user. Inspired by these studies, we examined whether and how Twitter users altered their expressions to form their online impressions over a specified period. We further investigated how these changes were related to changes in audience-related factors during that period.

2.3 Temporal Changes in Impression Management on Twitter

Extensive research has been conducted to study the temporal changes in various online user behaviors (e.g., rating in recommender systems (Liu *et al.*, 2017; Dror *et al.*, 2011), churning in Q&A sites (Pudipeddi *et al.*, 2014), or engaging in SNSs (Grinberg *et al.*, 2016)). For example, Facebook users are reported to be more likely to comment on their friends' posts after they create their own posts (Grinberg *et al.*, 2016). While such studies help us gain a better understanding of online user behaviors, few have explored how such changes in user behaviors affect their subsequent online impressions. Our study is among the first to explore how people change their expressions for online impression management.

To investigate the temporal change in user expressions, we chose Twitter as our study platform. Because Twitter is a high-immediacy medium compared to other media such as Facebook (Fiesler *et al.*, 2017), we expected that users would receive more immediate feedback from others, which may foster quicker customization of their information. In addition to the highly immediate nature of the platform, Twitter has some notable characteristics that may impact user expressions in their posts.

First, Twitter is a post-based medium in which users primarily present private information about themselves. Revealing information such as one's current situation or ongoing personal statuses in tweets (Fiesler *et al.*, 2017; Jaidka *et al.*, 2018) may make users aware of the feedback they receive from others, which may trigger an adjustment of their contents. For example, users may start to use more intimate expressions in their tweets as they receive more Likes, as Likes on Twitter are positive reactions from the audience (Gorrell and Bontcheva, 2016), which do not appear as frequently when compared to other media sites (Hayes *et al.*, 2016). It is worth noting that such an effect was not observed on other media (Cheng *et al.*, 2014).

Another notable feature of Twitter that may impact users' expressions is that Twitter users are regularly followed by strangers but are not allowed to control which sets of their followers receive the information that they output. According to Marwick and Boyd (2011), having many strangers in an audience often causes "context collapse;" an issue that makes it difficult for users to customize and deliver information to different types of people who do not share the same context. Therefore, contrary to the positive correlation between audience size and active engagement (i.e., posting activities or attitudes) (Rui and Stefanone, 2013; Vitak, 2012), the audience size on Twitter might have a negative impact on active engagement. For example, Twitter users' expressions may become more conservative as the size of their audience grows because their audiences often include many strangers. In addition, although the inner nature of retweeting is mostly positive (e.g., entertainment or agreement) (Boyd *et al.*, 2010), we assumed that retweets from others may make Twitter users' contents more neurotic due to the context collapse (Marwick and Boyd, 2011) brought about by the retweets.

Based on the above considerations, we believe that the expressions of Twitter users would be associated with their audience and that this association might eventually alter the impressions they form on Twitter.

3 METHOD

3.1 Data Collection

For our data collection, we first defined our target users and then collected their data using Twitter APIs. In selecting the target users, we decided to focus on users who had similar levels of experience using Twitter. We explain details of the procedure below.

We first used the Twitter Sampling API to collect Twitter users posting in English from September 3rd to October 7th, 2016. Through this procedure, 1.1M users were collected. After that, we extracted users who had posted 2800-3200 tweets from the pool of 1.1M users. The upper limit was set to 3200 tweets because the Twitter REST API does not allow third parties to obtain more than 3200 tweets from each user. We then extracted users who had been using Twitter for 950-1050 days to control for the frequency of posting tweets among users. We specifically set the period of use to 950-1050 days because the number of users corresponding to that period of use was the largest among the users who posted 2800-3200 tweets. By limiting the number of posts to 2800-3200 tweets and the period of Twitter use to 950–1050 days (approximately 2.5 years from March–April 2014 to September–October 2016), 2510 users remained.

Afterward, we extracted the size and density of audiences and feedback from the audiences from the collected data. Concerning audience size and density, we used the lists of followers and friends at the time of data collection (September 3rd to October 7th, 2016). For audience feedback, we obtained retweets, replies, and Likes that target users received during the above period.

3.2 Measures

In this subsection, we explain how we measured tweet-based personality, the size and density of audiences, and feedback from audiences from the data we collected.

3.2.1 Tweet-based personality

Tweet-based personality is one aspect of impressions that is generally created and inferred from the textual content of Twitter posts. To capture tweet-based personality, we adopted the Five-Factor Model (McCrae and Costa, 1987), which is composed of five personality factors: neuroticism, extraversion, openness, conscientiousness, and agreeableness. We chose this model because the personality factors identified are significantly related to language choices and styles (Golbeck *et al.*, 2011b; Golbeck *et al.*, 2011a; Schwartz *et al.*, 2013), which are important cues for controlling impressions in online settings (Baym, 1995; Walther *et al.*, 1992; Walther, 2007; Marriott and Buchanan, 2014). Thus, we believe that it is a reasonable model for characterizing personalities expressed in Twitter posts. In the left part of Table 1, we describe the personal characteristics of each factor with adjective pairs (McCrae and Costa, 1987).

To measure tweet-based personalities, we used a computational personality prediction technique called IBM Watson Personality Insights (IWPI). Using IWPI, we were able to calculate the scores of the five personality factors (McCrae and Costa, 1987) from textual features of expressions in tweets (IBM, 2017 (visited)). These scores ranged from 0 to 1. This prediction technique was developed based on prior research (Schwartz *et al.*, 2013) that explored the relationships between linguistic features extracted from users' posts with LIWC (a dictionary summarizing words into linguistic categories and dimensions) (Pennebaker *et al.*, 2007) and users' personality traits obtained from questionnaires (Costa and MacCrae, 1992). The right part of Table 1 shows a list of sample words and phrases belonging to each personality trait that was identified in prior research (Schwartz *et al.*, 2013). The notations $^+$ and $^-$ indicate whether a word/phrase raises or drops the score of the personality trait to which it belongs.

To calculate a tweet-based personality with IWPI, we defined a set of tweets from which we calculated a user's tweet-based personality. In doing so, we first excluded the tweets that had been posted in the initial 30 days (exclusion period in Figure 2) to alleviate the newcomers' effect in which users got used to the environment and the norms of Twitter. We then identified the initial set of N tweets to calculate the user's tweet-based personality for the past, and the final set of M tweets to calculate their tweet-based personality at present (see Figure 2). The number of tweets in the initial set N and the final set M were determined so that each set of tweets contained more than 1200 words. Note that 1200 words is the

Table 1: Corresponding adjective pairs and words/phrases of each personality factor.

Factor	Adjective pairs (McCrae and Costa, 1987)	Words/Phrases (Schwartz *et al.*, 2013)
Neuroticism	relaxed–high-strung, unemotional–emotional, secure–insecure, at ease–nervous, calm–worrying	$^+$: depression, I hate; $^-$: success, beautiful day
Extraversion	retiring–sociable, aloof–friendly, cold–warm, sober–fun loving, quiet–talkative, passive–active	$^+$: party, love you; $^-$: anime, internet
Openness	conventional–original, narrow interests–broad interests, uncurious–curious, uncreative–creative	$^+$: dream, universe; $^-$: ur, dont
Conscientiousness	negligent–conscientious, sloppy–neat, late–punctual, lazy–hardworking, careless–careful	$^+$: thankful, great day; $^-$: fuck, bored
Agreeableness	ruthless–soft-hearted, suspicious–trusting, critical–lenient, rude–courteous, uncooperative–helpful	$^+$: wonderful, blessed; $^-$: fucking, shit

Figure 2: Overview of data collected (Per = *neu, ext, ope, con, agr*; AF = *Like, Rep, RT* (Equation 2); AS = *Flr, FF* (Equation 3))

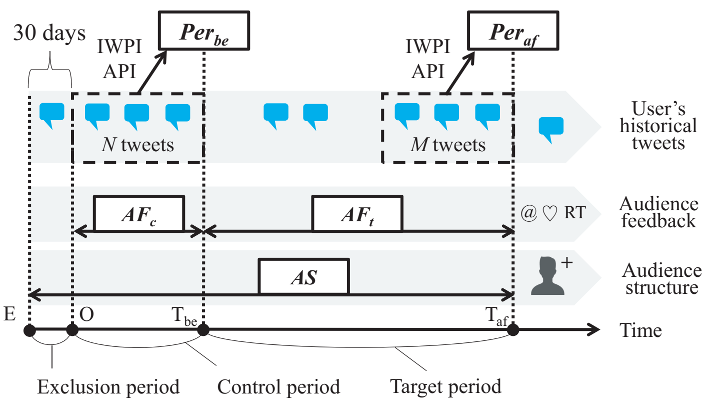


Table 2: Amount of data collected from target users

# of target users	1,618
# of use days	950 – 1,050
# of tweets	4,963,323 (2,800 – 3,200 per user)
Collection period	Sep. 3rd – Oct. 7th, 2016

minimum number required to obtain statistically reliable results to assess one's tweet-based personality using IWPI (IBM, 2017 (visited)). In addition, we excluded URL links from these tweet sets before counting the number of words.

Finally, we excluded users who had extreme posting patterns - those who posted 1200 words of tweets in less than seven days (one week) or those who took more than a year to post 1200 words of tweets. Eventually, 1618 users remained in the user pool, which we refer to as our "target users". Table 2 describes the amount of data collected from the target users and the collection period.

After collecting data from the target users, we calculated the scores of their tweet-based personalities in the past and present from each tweet set (N and M). As shown in Figure 2, Per_{be} and Per_{af} represent users' tweet-based personalities at time points in the past T_{be} and at present T_{af}, respectively. We calculated the changes in tweet-based personality by analyzing the differences in users' tweet-based personalities from the past to the present ($Per_{af} - Per_{be}$).

3.2.2 Audience size, density, and feedback

As discussed earlier, we focused on the size and density of audiences and feedback from the audience. We first defined two periods to measure the changes in audience feedback. As shown in Figure 2, we defined a "target period" as the period between T_{be} and T_{af}, and a "control period" as the period from the initial point O to T_{be}. Concerning audience feedback, we focused on the amount of feedback a user received during the target period relative to the control period. We paid attention to the "relative amount" of audience feedback rather than the absolute values because we were interested in understanding how the temporal within-user changes (i.e., increase/decrease) of audience feedback affected tweet-based personality. For example, suppose that a user received 10 retweets per day during the target period. Although this user may feel that the number is small if they had received 100 retweets per day during the control period, they may think the opposite if they had received only one retweet per day during the control period. To account for this potential difference, we used the relative frequency of receiving feedback in our analysis.

For audience size, we used changes in the number of followers. For audience density, we adopted the ratio of mutual-following users, defined as the Jaccard index of followers and friends. Although we wanted to calculate changes in the audience size and density in the same manner as for audience feedback, the Twitter REST API does not allow us to collect the history of followers/friends. Therefore, we assumed that the number of followers at E when the target users joined Twitter was zero, and simply used the number of followers and a Jaccard index of friends and followers at T_{af} instead of using changes in the number of followers and the ratio of mutual-following users from the control period to the target period. We introduce the mathematical definitions in Section 4.

4 ANALYSIS

For simplicity, we refer to audience size and density collectively as "audience structures". With the terms introduced before, our research question can be phrased as follows: "*How are temporal changes in tweet-based personality related to audience structures and feedback?*" To answer this question, we first observed the distributions of temporal changes in the tweet-based personality. We then conducted a series of linear multi-regression analyses in which the objective variable was the change in tweet-based personality, and the explanatory variables were audience structures and feedback. We explain the details of these analyses below.

4.1 Temporal Changes in Tweet-Based Personality

To capture the overall description of the changes in users' tweet-based personalities, we examined the user distribution for its change, calculated as follows:

$$\begin{aligned} \Delta Per &= Per_{af} - Per_{be} \\ Per &= neu, ext, ope, con, agr \end{aligned} \tag{1}$$

Since Per_{be} and Per_{af} range from 0 to 1, ΔPer ranges from -1 to 1. A positive Δneu means positive changes in neuroticism in a user's tweet-based personality, that is, an increase in neurotic expressions in his/her tweets.

4.2 Effects of Audience Structures and Feedback on Changes in Tweet-Based Personality

To understand whether audience structures and feedback correlated with tweet-based personalities, we performed a series of multiple regressions with the changes in the five features of tweet-based personality ΔPer as objective variables and audience structures and feedback as explanatory variables. All explanatory variables for the regression analysis were standardized such that the mean was 0, and the variance was 1. Below, we describe how we calculated audience feedback and structures.

4.2.1 Audience feedback

We used relative frequencies of receiving feedback as explanatory variables of audience feedback. Note that "relative frequency" is the degree of change in the frequency of receiving feedback between the control and target periods. We defined the relative frequencies of receiving Likes ($\delta Like$), replies (δRep), and retweets (δRT) as follows:

$$\begin{aligned} \delta AF &= \frac{AF_t/Days_t}{AF_t/Days_t + AF_c/Days_c + \alpha} \\ AF &= Like, Rep, RT, \end{aligned} \tag{2}$$

Here, AF_c and AF_t are the frequencies of receiving feedback in the control and target periods, respectively; $Days_c$ and $Days_t$ are the numbers of days in the control and target periods, respectively; and α is a supplementary term to make the denominator non-zero (for this analysis, we set α as 0.0001). The numerator indicates the daily frequency of receiving feedback during the target period, and the denominator is the summation of the daily frequencies of receiving feedback during both the target and control periods. Note that the relative frequency of receiving feedback δAF ranges from 0 to $\frac{1}{1+\alpha}(\approx 1)$. Higher daily frequencies of receiving audience feedback in the target period lead to a larger δAF (i.e., closer to 1).

4.2.2 Audience structures

We defined changes in the number of followers (δFlr) and the ratio of mutual-following users (δFF) as explanatory variables of audience structures:

$$\begin{aligned} \delta Flr &= |followers| \\ \delta FF &= \frac{|friends \cap followers|}{|friends \cup followers|} \end{aligned} \tag{3}$$

Here, $friends$ and $followers$ represent a set of friends and followers, respectively. δFF takes a larger value when friends and followers have a greater overlap.

4.2.3 Control variables

To understand how audience feedback and structures relate to changes in tweet-based personality, we should control for the effects of users' active behaviors, such as tweeting and following. Thus, we introduced the relative frequency of posting tweets δTw and an increase in the number of friends δFrd as control variables in the regression models, and defined them as:

$$\begin{aligned} \delta Tw &= \frac{Tw_t/Days_t}{Tw_t/Days_t + Tw_c/Days_c + \alpha} \\ \delta Frd &= |friends| \end{aligned} \tag{4}$$

Here, Tw_c and Tw_t represent the number of tweets posted in the control and the target period.

5 RESULTS

5.1 Temporal Changes in Tweet-Based Personality

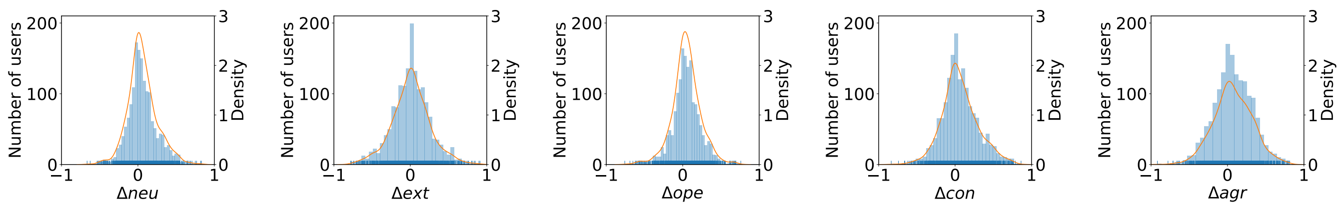
To observe temporal changes in the tweet-based personality, Figure 3 shows the user distributions for the temporal changes of each factor. Table 3(A) shows the descriptive statistics of changes in tweet-based personality. To see these distributions more specifically, Table 3(B) shows the percentage of the number of users corresponding to each interval when the amount of change in tweet-based personality is separated by -1.0, -0.20, -0.10, 0.0, 0.10, 0.20, and 1.0.

As shown in Figures 3(a) and 3(c), Δneu and Δope had similar distributions. The mean values were around 0.050, and the standard deviations were around 0.190. The standard deviations were small compared to the other factors.

The distributions of Δcon and Δagr were similar: the mean values were approximately 0.070, and the standard deviations were approximately 0.240. The standard deviation of Δext was approximately 0.240, but the mean value was smaller than those of Δcon and Δagr. Among the intervals of the change in conscientiousness and agreeableness, the percentage of users ranging from 0.20 to 1.0 was the highest. This suggests that changes in conscientiousness and agreeableness tended to be larger than changes in the other factors.

5.2 Effects of Audience Structures and Feedback on Changes in Tweet-Based Personality

Table 4 shows the regression coefficients (βs) with significant probabilities (*...$p < 0.05$, **...$p < 0.01$, ***...$p < 0.001$) and



(a) Neuroticism (b) Extraversion (c) Openness (d) Conscientiousness (e) Agreeableness

Figure 3: Distributions of the number of users for changes of (a) neuroticism, (b) extraversion, (c) openness, (d) conscientiousness, and (e) agreeableness in tweet-based personality.

Table 3: Descriptive statistics (Ave.: average, S.D.: standard deviation, Min.: minimum, 1Q.: first quantile, Med.: median, 3Q.: third quantile, Max.:maximum) and percentages of users in each interval of changes in tweet-based personality. C_s^l represents the percentage of users in a range from s to l (e.g., for 6.74% of users, the change in neuroticism ranges from -1.0 to -0.20).

	(A) Descriptive statistics							(B) Percentage of users					
	Ave.	S.D.	Min.	1Q.	Med.	3Q.	Max.	$C_{-1.0}^{-0.2}$	$C_{-0.2}^{-0.1}$	$C_{-0.1}^{0.0}$	$C_{0.0}^{0.1}$	$C_{0.1}^{0.2}$	$C_{0.2}^{1.0}$
Δneu	0.057	0.192	-0.665	-0.051	0.036	0.155	0.831	6.74	10.38	22.31	25.46	16.13	18.97
Δext	0.010	0.250	-0.781	-0.135	0.008	0.156	0.912	18.11	12.73	16.38	20.64	12.98	19.16
Δope	0.048	0.187	-0.758	-0.051	0.042	0.150	0.760	7.54	9.27	20.58	27.26	17.80	17.55
Δcon	0.063	0.236	-0.836	-0.079	0.037	0.193	0.863	11.25	10.57	19.47	19.47	14.83	24.41
Δagr	0.073	0.252	-0.917	-0.086	0.063	0.248	0.814	13.29	10.32	14.96	17.49	13.10	30.84

standard errors (S.E.s) for each regression model. The partial regression coefficients indicate the influence of an explanatory variable on the objective variable when the other explanatory variables are assumed to be constant. Here, we define β_x^y as the partial regression coefficient of an explanatory variable δx for an objective variable Δy.

In the regression model that explains temporal changes in tweet-based neuroticism, the relative frequency of receiving Likes showed negative coefficients ($\beta_{Likes}^{neu} = -0.129^*$), and the relative frequency of receiving retweets and the changes in the mutual-following ratio showed positive coefficients ($\beta_{RT}^{neu} = 0.102^*$, $\beta_{FF}^{neu} = 0.136^{***}$). These results indicate that an increase in the frequency of neurotic language use corresponds to an increase in the number of mutual-following users, an increase in the frequency of retweets received, and a decrease in the frequency of Likes received.

We found that the regression model for temporal changes in tweet-based extraversion had a negative coefficient for the relative frequency of posting tweets ($\beta_{Tw}^{ext} = -0.113^{**}$) and positive coefficients for the relative frequency of receiving Likes and for changes in the mutual-following ratio ($\beta_{Likes}^{ext} = 0.142^{**}$, $\beta_{FF}^{ext} = 0.073^*$). These results indicate that an increase in the frequency of use of extraverted expressions corresponds to a decrease in the frequency of posting tweets, an increase in the frequency of receiving Likes, and an increase in the number of mutual-following users.

For the regression model explaining tweet-based openness, the changes in the number of followers showed a negative coefficient ($\beta_{Flr}^{ope} = -0.116^{**}$). This means that an increase in the frequency of using open-minded language corresponds to a decrease in the number of followers.

The regression model for tweet-based conscientiousness was found to have a negative coefficient for the relative frequency of receiving Likes ($\beta_{Likes}^{ope} = -0.114^*$) and a positive coefficient for changes in the number of followers ($\beta_{FF}^{ope} = 0.149^{***}$). These results suggest that an increase in the frequency of using deliberate and cooperative expressions corresponds to a decrease in the frequency of receiving Likes and an increase in the number of mutual-following users.

In the regression model for tweet-based agreeableness, the changes in the mutual-following ratio showed a positive coefficient ($\beta_{FF}^{agr} = 0.173$). This indicates that an increase in the frequency of using agreeable expressions corresponds to an increase in the number of mutual-following users.

6 DISCUSSION

We performed a series of regression analyses to infer the influence of audience feedback and audience structures on temporal changes in tweet-based personality; however, it should be noted that our results do not necessarily indicate causal relationships.

6.1 Interpretations

Some of our results are consistent with previous findings. Users with increased audience density were found to use neurotic, extraverted, conscientious, and agreeable words more frequently over time. The correlation between audience density and neurotic language use can be explained by previous research (Burke and Develin, 2016), which indicated that Facebook users with denser networks tended to use more negative expressions to receive supportive comments from others. As the connections with the audience became denser, the frequency of using negative words and phrases increased, which may have led to the

Table 4: Linear regression models identifying effects of audience properties (i.e. audience feedback and structures) on temporal changes in (1) neuroticism, (2) extraversion, (3) openness, (4) conscientiousness, and (5) agreeableness in tweet-based personality. All p-values are adjusted with Bonferroni correction ($N = 1618$, $^{*}...p < 0.05$, $^{**}...p < 0.01$, $^{***}...p < 0.001$).

	(1) Δneu		(2) Δext		(3) Δope		(4) Δcon		(5) Δagr	
	β	$S.E.$	β	$S.E.$	β	$S.E.$	β	$S.E.$	β	$S.E.$
Intercept	0.000	0.025	0.000	0.025	0.000	0.025	0.000	0.024	0.000	0.024
Control variables										
δTw	-0.021	0.034	-0.113**	0.034	0.055	0.034	-0.022	0.034	-0.060	0.034
δFrd	0.020	0.032	0.013	0.032	0.059	0.032	-0.019	0.032	0.017	0.032
Audience feedback: relative frequency of receiving the feedback										
$\delta Like$	-0.129*	0.042	0.142**	0.042	-0.079	0.042	-0.114*	0.042	0.026	0.042
δRT	0.102*	0.037	0.002	0.037	0.027	0.037	0.084	0.037	-0.017	0.037
δRep	-0.012	0.027	0.002	0.027	-0.025	0.027	-0.038	0.027	0.037	0.027
Audience structures: increase/decrease of audience size (followers) or density (mutual-following ratio)										
δFlr	-0.010	0.032	-0.009	0.032	-0.116**	0.032	-0.007	0.032	-0.034	0.032
δFF	0.136***	0.025	0.073*	0.025	0.019	0.026	0.149***	0.025	0.173***	0.025

formation of a neurotic impression. Alternatively, since the users frequently used neurotic expressions, followers who had been unidirectionally following the user unfollowed them, resulting in a higher percentage of mutual-following users in the audience.

The correlation between audience density and conscientious language use is similar to the findings of a previous study (Vitak, 2012), which reported that the network size of Facebook users was associated with the degree to which users were conscious of what they wrote in their posts. Moreover, the association between audience density and extraverted and agreeable expressions can be explained by previous findings (Rui and Stefanone, 2013), which showed that the number of friends on Facebook (i.e., mutual connections) was related to active self-presentation. It is not clear whether the use of conscientious, extraverted, and agreeable words increased after audience density increased or vice versa. However, it seems more natural to assume that users engaged in more considerate and sociable expressions as their connections with the audience became denser.

We found that the frequency of neurotic expressions increased for users who received more retweets. Because retweets can spread posts regardless of the user's intentions and can increase the user's anxiety about context collapse (Marwick and Boyd, 2011), it is possible that the frequency of use of neurotic language increased after receiving many retweets. The reverse scenario (i.e., a user receiving more retweets because they frequently used neurotic words) seems unlikely because retweets are motivated by positive motives (Boyd *et al.*, 2010) and are thus less likely to be made in response to negative posts.

Some of our results appeared to be inconsistent with those of previous studies. Although Likes were anticipated to have less of an impact on online expressions or activities (Cheikh-Ammar and Barki, 2014; Cheng *et al.*, 2014), we found that receiving more Likes on Twitter was associated with a decrease in neurotic and conscientious expressions and an increase in extraverted expressions. We speculated that these different results stem from the different uses of Likes across different social media platforms. Specifically, receiving Likes from others is not considered special on Facebook, whereas receiving them on Twitter is a rather special occasion; it is more common for Twitter users to see tweets without Likes from others (Hayes *et al.*, 2016). A decrease in neurotic words, an increase in extraverted words, and a decrease in conscientious words meant that there was an increase in casual language expression. Considering that Likes on Twitter express positive attitudes of audiences (Gorrell and Bontcheva, 2016), we thought that the experience of receiving many Likes had the effect of making the user's linguistic expression more casual. Alternatively, it is also possible that users began to express themselves more frequently in a casual manner, which led to receiving more Likes. We believe that both of the above scenarios are happening concurrently.

Moreover, we found that a decrease in openness-related expressions was related to an increase in the number of followers. This result is somewhat inconsistent with previous findings that audience size is positively associated with active self-presentation (Rui and Stefanone, 2013; Vitak, 2012). Again, we think that this inconsistency stems from the different social media platforms studied: Twitter and Facebook. Rui and Stefanone (2013) and Vitak (2012) studied users on Facebook, where users can control the range of their audience. On Facebook, a user's audience consists of people whom the user recognizes and accepts as friends. In contrast, Twitter users cannot control the range of their audience. Because tweets are regularly read or seen by both friends and strangers, it is more difficult for Twitter users to estimate "who is reading my posts" than it is for Facebook users. Such uncertainty may have led Twitter users in the present study to exhibit a decrease in daring or liberal expressions (i.e., openness-related expressions) in their tweets as their number of followers increased.

6.2 Implications

6.2.1 Long term and short term

The key focus of our study is on the temporal changes in expression when forming impressions through Twitter posts within the same user. Overall, our results indicated that users changed their use of expressions on Twitter during the observed 2.5 years. Although previous studies have shown that SNS users often adjust their expressions based on feedback from others, this prior work largely focused on short-term adjustments (Burke *et al.*, 2009; Marwick and Boyd, 2011; Liu and Brown, 2014). Our study showed that such adjustments are also made over a longer period, likely affecting others' impressions of them. We infer that such long-term adjustments are made unconsciously because these adjustments were not triggered by specific incidents or feedback. This points to the possibility that the adjustments users make to manage their impressions may not be controlled entirely by the users themselves but may also be affected by other factors, such as audience structures and the accumulation of audience feedback. Such long-term adjustments may result in the formation of online impressions that deviate from expectations.

Although long-term impression formation can be considered as an accumulation of short-term impression formations, it is important to note that the short-term changes are subtle and often unnoticeable. Thus, to successfully support users' long-term impression formation, it would be effective to present information about the impressions that the user has given to the audience in the past and at present. For example, if users were presented with the impression that their current audience would have of them based on their past postings and interactions with the audience, they could adjust their impressions to be consistent with their past impressions and avoid forming online impressions that differ from their expectations.

6.2.2 Personality prediction techniques

Another implication from our study concerns personality prediction techniques which are increasingly gaining popularity these days. To date, the textual data retrieved from Twitter has often been used to explore how the use of words or phrases relates to personality traits (Golbeck *et al.*, 2011b; Schwartz *et al.*, 2013; Golbeck *et al.*, 2011a) and to develop personality-prediction techniques (IBM, 2017 (visited)). For example, people with an agreeable character were found to talk about others or talk to others using words such as "you" or "your" (Golbeck *et al.*, 2011a).

As implied from our results, linguistic patterns in tweets may be affected by audience-related factors. Therefore, we argue that researchers or developers attempting to build personality prediction techniques with higher precision should consider the effects of audience structures or feedback on people's use of expressions because linguistic patterns may change over time due to audience-related factors, even if their personality traits may have remained constant. For example, suppose that an individual was predicted to have an agreeable character from their tweets. However, because the degree of agreeableness in their tweet text is influenced by audience properties (e.g., the ratio of mutual-following users), it is challenging to identify whether it is indeed their actual personality or the effects of the audience properties.

Although further investigation is needed to address this issue, we believe that our study serves as a first step to uncover the potential impact of audience-related factors on users' impression formation.

6.3 Limitations

The main limitation of our study is that we fixed the target period to approximately 2.5 years. Our results indicated that half of our users showed changes in tweet-based personality during that period; however, some users may have changed their tweet-based personality in a shorter period, whereas others may have taken a longer time to change their tweet-based personality. Although the frequency of changes in tweet-based personality may be much higher or lower, this was not considered.

Next, the granularity of observation points for tweet-based personality was limited. As suggested by prior work (Burke *et al.*, 2009; Marwick and Boyd, 2011; Liu and Brown, 2014), users adjust their manner of expressing themselves to manage impressions for audiences over a short time span. Contrary to those studies, we focused on impression management over a more extended period. If we had more observation points, we would have been able to capture more details of the changes. Future work should investigate the appropriate level of granularity to capture the detailed change of linguistic expressions in users' impression formation.

Another limitation is that the target users were English speakers on Twitter. Whether our results can be applied to users with different languages or cultural backgrounds is an open question that should be assessed in the future.

6.4 Future Directions

In addition to resolving the above limitations, we will address the following issues in the future.

6.4.1 Provisional personality and actual personality

Whether such changes in one's provisional personality formed on one platform (e.g., Twitter) are also found in their actual personality is an open question. Results from previous studies on this subject are mixed. On the one hand, as discussed by Marriott and Buchanan (2014), Back *et al.* (2010), and Gosling *et al.* (2011), the impressions people tend to form of someone in online settings are closely related to the actual personality traits of that person. This insight indicates that changes in a user's personality expressed on one platform may also hold for their actual personality. On the other hand, other researchers (Norman, 1963; Costa and MacCrae, 1992) argue that personality traits are temporally stable factors in humans. Based on this argument, the change in a user's linguistic expression on an SNS platform is superficial and is not related to the user's intrinsic personality.

To address this point, we need to examine whether a user's personality traits change in the same way as their tweet-based personality changes, using traditional methodologies such as questionnaire surveys. Such an examination might show the future potential of research on impression formation using SNS data because it would examine the extent to which people's actual personality traits are manifested in SNSs.

6.4.2 Participation in different SNSs

We believe that, including audience properties, the use of different SNSs for different purposes is one of the factors that implicitly affects the temporal changes in tweet-based personality. For example, by compartmentalizing the use of different SNSs, a user may gradually use more extraverted expressions on Twitter or Facebook while using more introverted expressions on a different site (e.g., a healthcare SNS). In fact, Twitter users were found to express their extraverted personality more often than Disqus users (Maruf *et al.*, 2015). By expanding our research to multiple SNS sites, we may be able to achieve a better understanding of how users form and maintain their impressions in online settings.

6.4.3 Contents and senders of reply

Contrary to Likes and retweets, replies from audiences were found to have no association with tweet-based personality. Considering that the act of replying is a more direct and intimate form of communication with the audience than Likes or retweets, we expect that the lack of a significant effect of the audience's reply on tweet-based personality may be because the content or sender of the replies has a greater effect than the frequency of receiving a reply.

For example, users may express themselves in a more introverted manner when they start to receive more critical replies but may express themselves in a more extraverted manner when they receive more affirmative replies. They may also use more casual language when they receive more replies from friends and acquaintances, whereas they may use more formal language when they receive more replies from complete strangers. The above effects of content and sender are not necessarily independent, and there is a large possibility that they are interdependent. In our future research, we aim to describe the conditions under which tweet-based personality changes in more detail by conducting an analysis that considers the effects of the content of the reply and the relationship with the sender of the reply.

7 Conclusion

To understand how users control their linguistic expressions for impression management, we studied the relationship of within-subject temporal changes in tweet-based personality and audience properties, using 5 million posts from 1.6 thousand Twitter users over 2.5 years. The primary results indicated that temporal changes in the frequency of using casual expressions corresponded to temporal changes in the frequency of receiving Likes. Moreover, we found a correspondence between the changes in the frequency of using nervous, extraverted, conscientious, and agreeable language and the changes in the density of the relationship with the audience. Our results provide evidence that users adjust their linguistic expressions over time through their interaction with the audience. Based on these findings, we discussed the effects of such long-term changes in linguistic expressions on impression formation.

This is the first study to investigate temporal changes in linguistic expressions for impression management over a long period of time. In our future work, we will assess whether the audience has the same impressions of a user that are intended by the user, and how users control linguistic expressions according to different cultural backgrounds or in different SNSs. We believe that this study will lead to a better understanding of the mechanisms of impression formation among people online.

References

Back, M. D., J. M. Stopfer, S. Vazire, S. Gaddis, S. C. Schmukle, B. Egloff, and S. D. Gosling (2010). "Facebook Profiles Reflect Actual Personality, Not Self-Idealization". *Psychological Science*. 21(3): 372–374.

Baym, N. K. (1995). "The emergence of community in computer-mediated communication."

Binder, J., A. Howes, and A. Sutcliffe (2009). "The Problem of Conflicting Social Spheres: Effects of Network Structure on Experienced Tension in Social Network Sites". In: *Proceedings of the 27th International Conference on Human Factors in Computing Systems*. New York, New York, USA: ACM Press. 965–974.

Boyd, D., S. Golder, and G. Lotan (2010). "Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter". In: *Proceedings of the 43rd Hawaii International Conference on System Sciences*. 1–10.

Braginsky, B. M., M. Grosse, and K. Ring (1966). "Controlling outcomes through impression-management: An experimental study of the manipulative tactics of mental patients". *Journal of Consulting Psychology*. 30(4): 295–300.

Burke, M. and M. Develin (2016). "Once More with Feeling: Supportive Responses to Social Sharing on Facebook". In: *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. 1460–1472.

Burke, M., C. Marlow, and T. Lento (2009). "Feed Me: Motivating Newcomer Contribution in Social Network Sites". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 945–954.

Cheikh-Ammar, M. and H. Barki (2014). ""Like" is More than Just a Thumbs up: The Role of Feedback and Sociability in SNS Usage". In: *Proceedings of the 20th Americas Conference on Information Systems*. 1–13.

Cheng, J., C. Danescu-Niculescu-Mizil, and J. Leskovec (2014). "How Community Feedback Shapes User Behavior". In: *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*. 41–50. arXiv: 1405.1429.

Choi, Y. H. and N. N. Bazarova (2015). "Self-Disclosure Characteristics and Motivations in Social Media: Extending the Functional Model to Multiple Social Network Sites". *Human Communication Research*. 41(4): 480–500.

Costa, P. T. and R. R. MacCrae (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) Manual*. Psychological Assessment Resources, Incorporated.

Deeb-swihart, J., C. Polack, E. Gilbert, and I. Essa (2017). "Selfie-Presentation in Everyday Life: A Large-Scale Characterization of Selfie Contexts on Instagram". In: *Proceedings of the Eleventh International AAAI Conference on Web and Social Media*. 42–51.

Dominick, J. R. (1999). "Who Do You Think You Are? Personal Home Pages and Self-Presentation on the World Wide Web". *Journalism & Mass Communication Quarterly*. 76(4): 646–658.

Dror, G., N. Koenigstein, and Y. Koren (2011). "Yahoo! Music Recommendations: Modeling Music Ratings with Temporal Dynamics and Item Taxonomy". *Proceedings of the Fifth ACM Conference on Recommender Systems*: 165–172.

Ellison, N., C. Steinfield, and C. Lampe (2006). "Spatially Bounded Online Social Networks and Social Capital: The Role of Facebook". *Annual Conference of the International Communication Association*. 36(1-37).

Ellison, N. B., C. Steinfield, and C. Lampe (2007). "The Benefits of Facebook "Friends:" Social Capital and College Students' Use of Online Social Network Sites". *Journal of Computer-Mediated Communication*. 12(4): 1143–1168.

Fiesler, C., M. Dye, J. L. Feuston, C. Hiruncharoenvate, C. Hutto, S. Morrison, P. Khanipour Roshan, U. Pavalanathan, A. S. Bruckman, M. De Choudhury, and E. Gilbert (2017). "What (or Who) Is Public? Privacy Settings and Social Media Content Sharing". In: *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 567–580.

Goffman, E. (1959). *The presentation of self in everyday life.* New York.

Golbeck, J., C. Robles, M. Edmondson, and K. Turner (2011a). "Predicting Personality from Twitter". In: *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom)*. IEEE. 149–156.

Golbeck, J., C. Robles, and K. Turner (2011b). "Predicting Personality with Social Media". In: *CHI '11 Extended Abstracts on Human Factors in Computing Systems*. New York, New York, USA: ACM. 253–262.

Gorrell, G. and K. Bontcheva (2016). "Classifying Twitter favorites: Like, bookmark, or Thanks?" *Journal of the Association for Information Science and Technology*. 67(1): 17–25. arXiv: 0803.1716.

Gosling, S. D., A. A. Augustine, S. Vazire, N. Holtzman, and S. Gaddis (2011). "Manifestations of Personality in Online Social Networks: Self-Reported Facebook-Related Behaviors and Observable Profile Information". *Cyberpsychology, Behavior, and Social Networking*. 14(9): 483–488.

Grinberg, N., P. A. Dow, L. A. Adamic, and M. Naaman (2016). "Changes in Engagement Before and After Posting to Facebook". In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 564–574.

Gross, R., A. Acquisti, and H. J. Heinz (2005). "Information Revelation and Privacy in Online Social Networks". In: *Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society*. New York, New York, USA: ACM Press. 71.

Hancock, J. T., C. Toma, and N. Ellison (2007). "The Truth about Lying in Online Dating Profiles". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, New York, USA: ACM Press. 449–452.

Hayes, R. A., C. T. Carr, and D. Y. Wohn (2016). "One Click, Many Meanings: Interpreting Paralinguistic Digital Affordances in Social Media". *Journal of Broadcasting and Electronic Media*. 60(1): 171–187.

IBM (2017 (visited)). "IBM Watson Personality Insights".

Jaidka, K., S. C. Guntuku, and L. H. Ungar (2018). "Facebook vs. Twitter: Differences in self-disclosure and trait prediction". In: *Proceedings of the Twelfth International AAAI Conference on Web and Social Media*. 141–150.

Kapidzic, S. (2013). "Narcissism as a Predictor of Motivations Behind Facebook Profile Picture Selection". *Cyberpsychology, Behavior and Social Networking*. 16(1): 14–19.

Lampinen, A., S. Tamminen, and A. Oulasvirta (2009). ""All My People Right Here, Right Now": Management of Group Co-Presence on a Social Networking Site". In: *Proceedinfs of the ACM 2009 International Conference on Supporting Group Work*. New York, New York, USA: ACM Press. 281–290.

Lang, C. and H. Barton (2015). "Just untag it: Exploring the management of undesirable Facebook photos". *Computers in Human Behavior*. 43: 147–155.

Litt, E. (2012). "Knock, Knock. Who's There? The Imagined Audience". *Journal of Broadcasting & Electronic Media*. 56(3): 330–345.

Litt, E. and E. Hargittai (2016). "The Imagined Audience on Social Network Sites". *Social Media + Society*. 2(1): 20563051166334

Liu, D. and B. B. Brown (2014). "Self-disclosure on social networking sites, positive feedback, and social capital among Chinese college students". *Computers in Human Behavior*. 38: 213–219.

Liu, Y., Y. Liu, Y. Shen, and K. Li (2017). "Recommendation in a Changing World: Exploiting Temporal Dynamics in Ratings and Reviews". *ACM Transactions on the Web*. 12(1): 1–20.

Marder, B., A. Joinson, A. Shankar, and K. Thirlaway (2016). "Strength matters: Self-presentation to the strongest audience rather than lowest common denominator when faced with multiple audiences in social network sites". *Computers in Human Behavior*. 61: 56–62.

Marriott, T. C. and T. Buchanan (2014). "The true self online: Personality correlates of preference for self-expression online, and observer ratings of personality online and offline". *Computers in Human Behavior*. 32: 171–177.

Maruf, H. A., N. Meshkat, M. E. Ali, and J. Mahmud (2015). "Human Behaviour in Different Social Medias: A Case Study of Twitter and Disqus". In: *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015. ASONAM '15*. Paris, France: Association for Computing Machinery. 270–273.

Marwick, A. E. and D. Boyd (2011). "I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience". *New Media & Society*. 13(1): 114–133.

McCrae, R. R. and P. T. Costa (1987). "Validation of the Five-Factor Model of Personality Across Instruments and Observers". *Journal of Personality and Social Psychology*. 52(1): 81–90.

Naaman, M., M. Naaman, J. Boase, J. Boase, C. Lai, and C. Lai (2010). "Is it Really About Me? Message Content in Social Awareness Streams". In: *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*. 189–192.

Norman, W. T. (1963). "Toward an Adequate Taxonomy of Personality Attributes: Replicated Factor Structure in Peer Nomination Personality Rating". *The Journal of Abnormal and Social Psychology*. 66(6): 574–583.

Pennebaker, J. W., C. K. Chung, M. Ireland, A. Gonzales, and R. J. Booth (2007). "The development and psychometric properties of LIWC2007". *LIWCNET*. 1: 1–22.

Pudipeddi, J. S., L. Akoglu, and H. Tong (2014). "User Churn in Focused Question Answering Sites : Characterizations and Prediction". In: *Proceedings of the 23rd International Conference on World Wide Web*. 469–474.

Rui, J. and M. A. Stefanone (2013). "Strategic self-presentation online: A cross-cultural study". *Computers in Human Behavior*. 29(1): 110–118.

Schwartz, H. A., J. C. Eichstaedt, M. L. Kern, L. Dziurzynski, S. M. Ramones, M. Agrawal, A. Shah, M. Kosinski, D. Stillwell, M. E. P. Seligman, and L. H. Ungar (2013). "Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach". *PLoS ONE*. 8(9): e73791. Ed. by T. Preis.

Sleeper, M., R. Balebako, S. Das, A. L. McConahy, J. Wiese, and L. F. Cranor (2013). "The Post that Wasn't: Exploring Self-Censorship on Facebook". In: *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*. Vol. 13. New York, New York, USA: ACM Press. 793–802.

Sunnafrank, M. (1986). "Predicted outcome value during initial interactions: A reformulation of uncertainty reduction theory". *Human Communication Research*. 13(1): 3–33.

Tice, D. M., J. L. Butler, M. B. Muraven, and A. M. Stillwell (1995). "When Modesty Prevails: Differential Favorability of Self-Presentation to Friends and Strangers". *Journal of Personality and Social Psychology*. 69(6): 1120–1138.

Vazire, S. and S. D. Gosling (2004). "e-Perceptions: Personality Impressions Based on Personal Websites". *Journal of Personality and Social Psychology*. 87(1): 123–132.

Vitak, J. (2012). "The Impact of Context Collapse and Privacy on Social Network Site Disclosures". *Journal of Broadcasting & Electronic Media*. 56(4): 451–470.

Walther, J. B. (2007). "Selective self-presentation in computer-mediated communication: Hyperpersonal dimensions of technology, language, and cognition". *Computers in Human Behavior*. 23(5): 2538–2557. arXiv: 0803973233.

Walther, J. B., J. F. Anderson, and D. W. Park (1992). "Interpersonal Effects in Computer-Mediated Interaction". *Communication Research*. 19(1): 52–90.

Wang, Y., P. G. Leon, A. Acquisti, L. F. Cranor, A. Forget, and N. Sadeh (2014). "A Field Trial of Privacy Nudges for Facebook". In: *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. New York, New York, USA: ACM Press. 2367–2376.

Wang, Y., G. Norcie, S. Komanduri, A. Acquisti, P. G. Leon, and L. F. Cranor (2011). ""I Regretted the Minute I Pressed Share": A Qualitative Study of Regrets on Facebook". In: *Proceedings of the Seventh Symposium on Usable Privacy and Security*. 1–16.

Zhao, X., N. Salehi, S. Naranjit, S. Alwaalan, S. Voida, and D. Cosley (2013). "The Many Faces of Facebook: Experiencing Social Media As Performance, Exhibition, and Personal Archive". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, New York, USA: ACM Press. 1–10.

Zytko, D., S. A. Grandhi, and Q. G. Jones (2014a). "Impression Management Through Communication in Online Dating". In: *Proceedings of the Companion Publication of the 17th ACM Conference on Computer Supported Cooperative Work and Social Computing*. 277–280.

Zytko, D., S. A. Grandhi, and Q. Jones (2014b). "Impression Management Struggles in Online Dating". In: *Proceedings of the 18th International Conference on Supporting Group Work*. New York, New York, USA: ACM Press. 53–62.