Anonymity, Intimacy and Self-Disclosure in Social Media

Xiao Ma
Jacobs Institute, Cornell Tech
New York, NY, USA
xiao@jacobs.cornell.edu

Jeff Hancock
Stanford University
Stanford, CA, USA
jeff.hancock@stanford.edu

Mor Naaman
Jacobs Institute, Cornell Tech
New York, NY, USA
mor@jacobs.cornell.edu

ABSTRACT

Self-disclosure is rewarding and provides significant benefits for individuals, but it also involves risks, especially in social media settings. We conducted an online experiment to study the relationship between content intimacy and willingness to self-disclose in social media, and how identification (real name vs. anonymous) and audience type (social ties vs. people nearby) moderate that relationship. Content intimacy is known to regulate self-disclosure in face-to-face communication: people self-disclose less as content intimacy increases. We show that such regulation persists in online social media settings. Further, although anonymity and an audience of social ties are both known to increase self-disclosure, it is unclear whether they (1) increase self-disclosure baseline for content of all intimacy levels, or (2) weaken intimacy's regulation effect, making people more willing to disclose intimate content. We show that intimacy always regulates self-disclosure, regardless of settings. We also show that anonymity mainly increases self-disclosure baseline and (sometimes) weakens the regulation. On the other hand, an audience of social ties increases the baseline but *strengthens* the regulation. Finally, we demonstrate that anonymity has a more salient effect on content of negative valence. The results are critical to understanding the dynamics and opportunities of self-disclosure in social media services that vary levels of identification and types of audience.

Author Keywords
Anonymity; intimacy; self-disclosure; valence; social media.

ACM Classification Keywords
H.4.m Miscellaneous

INTRODUCTION

Self-disclosure on social media platforms has become an important part of one's social life. Self-disclosure is the act of "revealing personal information to others" [29], and can be both intrinsically rewarding [25, 42, 43, 51] and socially beneficial. Social media technologies greatly facilitate such sharing to a large audience through non-direct status updates [4],

supporting goals such as social validation, relational development, social control and resource gain, as well as goals related to benefiting others [5]. At the same time, self-disclosure – especially on social media – involves inherent risks. The ill-defined audience creates challenges such as context collapse [36], increased vulnerability, and loss of privacy [5], limiting the benefits that self-disclosure could bring.

Anonymity is known to increase self-disclosure, both in offline and online settings. Early evidence that anonymity increases self-disclosure is the "stranger on a train" phenomenon [47] where people may self-disclose quite intimately to fellow passengers on a train. Similarly, in online settings, Suler observed the disinhibition effect [53] where people self-disclose more in many online environments than in person. For the HCI community, anonymity, as a design choice, can be leveraged to influence online interaction and norms, such as in the case of Reddit or 4/chan/ [7, 32].

However, in all these cases, the effect of anonymity is strictly coupled with the target audience and context of self-disclosure. It is not clear that people self-disclose more to others, online or offline, because of anonymity, or because the audience in each particular settings results in different benefit/risk dynamic. For example, it is possible that people self-disclose more to those in physical proximity because of propinquity or because they share common knowledge about what is happening around them.

In other words, anonymity and audience could have different effects on self-disclosure. Social media platforms that leverage anonymity as a design choice while allowing people to disclose to a specific audience can present a wide set of new possibilities to support self-disclosure. Several recent mobile applications have exemplified such new possibilities. Applications like Rumr, Wumi, Yik Yak, Whisper, and Secret (now defunct) allow users to anonymously self-disclose to an audience whose relationship is known to the users (social ties or people nearby). These platforms provide new affordances and could be inspirations for more nuanced social media platform designs, making it critical to understand the separate effects of anonymity and audience in influencing self-disclosure. Different combinations of the two can also bring new possibilities, e.g. a Facebook confession board where people could reach out to ones social ties anonymously [8].

Here we use an online experiment to examine how the identification and audience features of the platform design moderate the relationship between content intimacy and willingness to self-disclose. It is known that content intimacy regulates

self-disclosure in real-name conditions: people self-disclose less as content intimacy increases [1, 30]. We are interested in examining whether this regulation will persist in social media settings with different identification, and towards different audiences. Compared to the observational approach of analyzing people's use of self-disclosure systems, the experimental approach we take here allows us to collect data on attitudes about what people do and *do not* feel comfortable disclosing, which will form a more complete picture for understanding patterns of self-disclosure.

We use the same experiment to also study the patterns of self-disclosure about positive content and negative content.[1] People share both positive and negative aspects in social media [10, 31], and these different disclosures are important for different aspects of individuals well-being [5, 10]. Are people more likely to share positive content with social ties? Does anonymity have a more salient effect on encouraging sharing for negative-valence content? Answering such questions could offer insights about people's behavior and self-disclosure in social media, help understand and predict system dynamics, and guide the future design of new social media platforms.

To address these questions, we used a mixed factorial design with two between-subject and two within-subject factors. The between-subject factors focus on features of social media sites: *identification* (real name vs. anonymous) and *audience* (social ties vs. people nearby). The within-subject factors focus on features of the disclosure: *intimacy* (low through high) and *valence* (negative vs. positive). In each condition, we collected answers about comfort level of self-disclosure for content of different levels of intimacy, for example: "posting about missing home or old friends on this platform". The self-disclosure content items were developed based on previous work from the field of social psychology [30, 48], and serve as an additional contribution of this work.

BACKGROUND AND HYPOTHESES

It is somewhat puzzling that people sometimes self-disclose quite intimately to total strangers, such as fellow passengers on a train [47]. As Simmel notes, "the stranger who moves on often receives the most surprising openness-confidences which have the character of a confessional and which would be carefully withheld from a more closely related person [50]." Such "openness" to strangers compared to "a more closely related person" could potentially have two explanations: the difference in identification (anonymous), or the difference in the relationship to the audience (based on physical proximity). The former and common view is that interacting with strangers provides a degree of anonymity, which disinhibits self-disclosure by reducing risk, e.g. harming one's self-presentation. An alternative explanation, however, is that the physical proximity associated with people

nearby provides better grounding for conversation (e.g. common knowledge of events nearby, local weather) and therefore increases self-disclosure.

Anonymous Audience Applications

Interestingly, several recent mobile applications exemplified the possibility to tease factors of identification and audience apart. Such applications, like Rumr, Wumi, Yik Yak, Whisper, and Secret (now defunct), have two main characteristics. The first characteristic is allowing users to anonymously create short posts, and browse through a feed of posts from others users without any identity marker (e.g. user ID, real name, profile photo, email address). In other words, such platforms are designed to offer practical anonymity. Under such identity design, or rather, lack of identity design, two consequences immediately follow: (1) connections cannot be drawn between posts created by the same user[2]; and (2) the risks of self-disclosure are kept low due to lack of accountability or reputation.

The second characteristic of the applications mentioned above is that users can anonymously disclose to a specific audience, such as their social connections from other platforms (e.g. from Facebook or from their phone's address book), or to people nearby, often using the mobile phone's location-aware features, even as these posters remain anonymous. For example, in the early versions of the Secret app, after granting access to information from their phone address book and Facebook friend list, users could then see posts marked as from a "friend" or a "friend of a friend". Therefore, the audience on this platform is one's social ties despite the fact that one cannot know exactly who posted what. Yik Yak, on the other hand, uses a different audience logic: people nearby. In Yik Yak, posts are geotagged and organized by location. Users can anonymously post "locally" (within 5-mile radius) and "peek" into posts at other locations. It is worth noting that in practice some applications use a mix of both designs thus resulting in a mixed audience. In fact, Secret had transitioned from social ties model into a predominantly nearby model before shutting down in April 2015.

Goals and Functions of Self-Disclosure

There are a variety of reasons that people self-disclose. Self-disclosure decisions can be driven by a "fever" model [51]. Keeping one's thoughts to themselves requires active mental work, which could result in pressure, stress, and preoccupation [34, 61, 60]. Self-disclosure can release this stress, or "fever", potentially bringing many benefits to the individual. Self-disclosure is rewarding [54] and helps connectedness, a primary human need [49]. Researchers have long suggested that there are therapeutic effects for self-disclosure [41, 43, 51]. Various forms of self-disclosure, including talking with friends or relevant groups for social support [58] and confiding in a therapist [25] can be beneficial to one's mental and physical health. In particular, people use social media platforms to self-disclose to "satisfy their instrumental needs and adapt their communication behaviors accordingly" [5].

---

[1]It is worth stressing that we only examined self-disclosure, rather than general negative behaviors such as bullying, trolling, and other aggressive actions that are more likely to occur in anonymous settings [7, 53] due to the lack of accountability and the same mitigation of risk that supports self-disclosure.

[2]It is important to note that the companies operating these platforms *are* often able to connect different posts from the same individual through identifying the device that the platform is running on.

The fever model of self-disclosure does not address the risks involved. Self-disclosure can result in increased vulnerability, loss of privacy [5, 26, 57], negative impact on identity and self-presentation [20], as well as present risks caused by disclosing to unintended audience (context collapse) [22, 36]. Thus, the functional model [5] provides a more complete understanding of self-disclosure decisions by taking both benefits and risks into consideration [5, 11]. Through the lens of the functional model, decisions about self-disclosure, such as to disclose or not, and if yes, when, where, and to whom, are all determined dynamically as people try to maximize benefits (e.g. social validation, self-expression) and minimize risks [19, 39].

Following the functional model of self-disclosure, factors that influence decisions about self-disclosure include: content intimacy, anonymity [35], relationship with the target person [17, 19], norms [16, 44], reciprocity [21], valence, appropriateness [11, 18], identity, and individual personality. Theses factors can interact with each other in modulating the benefit/risk dynamic.

Hypotheses

In this work, we examine the relationship between content intimacy and the likelihood of self-disclosure. We look at how identification and audience moderate such relationship. Previous studies have found a negative correlation between the intimacy level of an item and past self-disclosure behavior about it, meaning that individuals disclose less about more intimate topics [1, 30]. We refer to this negative correlation as the *regulation* effect. Therefore, we propose the following conceptual model that will guide the narration of our hypotheses:

$$\text{self-disclosure} = \text{baseline} - \text{regulation} \times \text{intimacy}$$

Our hypotheses relate to how identification (anonymous or real name), the audience of sharing (social ties or nearby), and the valence of content affect this relationship between intimacy of content and the likelihood of self-disclosure. For all these variables, the effect could be on the *baseline* (i.e. impacting self-disclosure for any level of content intimacy) or the *regulation* of intimacy (i.e. changing the way in which self-disclosure is sensitive to changes in content intimacy). Note that since past behavior does not actually predict self-disclosure [16], we focus in our work on attitudes towards self-disclosure rather than actual behavior. Thus, when we hypothesize about self-disclosure we use the phrase as a shorthand for "reported self-disclosure comfort level".

Anonymity lowers the risk of harming one's personal image, or having the information being disclosed being used against one's interests. By lowering risks, anonymity can increase overall self-disclosure. Suler observed the online disinhibition effect where people self-disclose online more frequently or intensively than they would in person [53]. Under anonymity, people become less inhibited, potentially because they are less constrained by the expectations of others or perceive less associated risks of social sanction [3, 27]. This allows people to explore aspects of their identity that otherwise would have been impossible [55]. Therefore:

H1a Anonymity increases the self-disclosure baseline.

Closer relationship with the target audience is correlated with higher self-disclosure [40]. Naturally, people confide in friends and trust them with higher level of self-disclosure across the board [48]. Social penetration theory [2] describes that as relationship develops, interpersonal communication moves to deeper and more intimate. Therefore:

H1b The audience of social ties increases the self-disclosure baseline compared to an audience of people nearby.

As discussed above, research has established the regulation effect of intimacy on self-disclosure, in which people are less likely to disclose more intimate information. In addition, Rubin's 1978 study on freshman dormitories found that friendship was more highly related to self-disclosure in intimate than non-intimate topic areas, whereas proximity was more highly related to disclosure in non-intimate than in intimate areas [48]. Considering the fact that real name platforms resemble face-to-face disclosures in terms of identification, which is the setup in which previous studies were conducted, we hypothesize that:

H2a When identification is by real name, there is a regulation effect of intimacy on self-disclosure.

While people confide in friends, they still care about self presentation and impression management. Especially in "social awareness streams" settings, self-disclosure does not always take place on a one-on-one basis (the condition under which previous psychological studies were conducted [48]). In online identified platforms, people are less likely to make more intimate disclosures [33, 38], potentially due to the visibility and persistence of information, or due to context collapse [36]. Thus, the intimacy regulation effect should be strongest when one is identified by their real name and when the audience is comprised of social ties.

H2b When identification is by real name, the regulation effect of intimacy on self-disclosure is stronger with an audience of social ties than with people nearby.

Anonymity is likely to moderate the intimacy regulation of self-disclosure in the two different audience conditions. When the audience is social ties, anonymous platforms show no identifiers such as real name or photo. Nevertheless, because of the very nature of self-disclosure, this technological anonymity could more easily be compromised as friends could figure out who posted certain messages based on the information included. The more people self-disclose, the more likely their social ties are to infer the discloser's identity. Thus, disclosing content of high intimacy level still poses risks, while disclosing content of low intimacy level does not. Therefore, intimacy's regulation effect should still exist, but only be weakened with anonymity. In other words:

H2c When the audience is social ties, there is a regulation effect of intimacy on self-disclosure, and the regulation is stronger when identification is by real name compared to anonymous conditions.

When disclosing to people nearby anonymously, on the other hand, anonymity will be less likely to be compromised by self-disclosure, as the audience has less information and context to guess the identity of the original poster. Therefore, self-disclosing content of high or low intimacy level should be of approximately equal risks, lifting intimacy's regulation effect altogether. Therefore, we hypothesize that:

H2d When the audience is people nearby and identification is anonymous, there is no regulation effect of intimacy on self-disclosure.

Finally, we hypothesize about the effect of content valence (positive or negative) on self-disclosure. Following the functional model of self-disclosure [5, 11] that weighs the benefits and risks, positive and negative valence content has different benefit/risk balance. Benefits of self-disclosure can be grouped according to intrinsic (release stress) [25, 42, 43, 51] and social (relationship, e.g. social validation) [4]. On real name platforms, self-disclosing about positive content adds to one's social image, while negative content may pose risks of harming one's social presentation. Therefore,

H3a When identification is by real name, positive content has higher self-disclosure baseline than negative content.

When anonymous, though, there is little social value of self-disclosure, and the intrinsic benefits of self-disclosing about negative valence items may be higher than positive valence ones, as predicted by the fever model [51] (negative items bring more stress so it will be more rewarding to release this stress). Hence,

H3b When identification is anonymous, negative content has higher self-disclosure baseline than positive content.

METHOD

To investigate the hypotheses, we conducted a questionnaire-based survey experiment, collecting responses from participants about comfort levels of self-disclosure for content items of different levels of intimacy. We used a mixed factorial design, with two between-subject and two within-subject factors. The between-subject factors focus on the Identification and Audience features of social media platforms, each with two options, resulting in four conditions in total, as shown in Figure 1: Real Name / Social Ties (abbreviated as *RS*), Real Name / Nearby (*RN*), Anonymous / Social Ties (*AS*), Anonymous / Nearby (*AN*). The within-subject factors focus on features of self-disclosure: Intimacy (low through high) and Valence (negative vs. positive). The questionnaire consists of four parts, in this order: item intimacy-level rating, platform description and verification, item self-disclosure comfort rating, and demographic information. We describe each part in detail below, but first we describe the development of the content items as each item was rated by each participants for both intimacy level, and comfort in self-disclosure.

Generating Items: Varying Intimacy and Valence

To study the effect of intimacy on self-disclosure under the different conditions, we followed two classic studies in using an item-based questionnaire [30, 48]. These questionnaires use items of difference intimacy levels to ask about

self-disclosure. There are two advantages to using the questionnaire-based method to understand self-disclosure. First, a questionnaire allows getting data on items that participants have not disclosed or would not disclose in particular settings, thus complementing other methods that mainly focus on content analysis of disclosed content obtained from actual behavior [15, 59]. Second, the questionnaire allows us to get data on items across different intimacy levels, providing more detailed insight into the relationship between the two variables.

We built upon two measuring instruments previously used in the field of psychology on self-disclosure. The most widely used self-disclosure questionnaire is the 60-item Jourad's Self-Disclosure Questionnaire (JSDQ) [30]. This questionnaire presents participants with different items (e.g. "what it takes to get me worried, anxious, and afraid") from different categories (e.g. "Attitudes and Opinions"). We followed Jourard [30] and created items in six categories: (1) Tastes and Interests; (2) Attitudes and Opinions; (3) Work or Studies; (4) Economic and Social Status ; (5) Self-concept and Interpersonal Relations; (6) Physical Appearance and Sex. The first three categories are exactly the same as Jourard's questionnaire; and the latter three are expansions of the original ones, which were money, personality, and body.

Next, we develop specific items in each of the six categories. In this stage, we again built on existing work, using Jourard's questionnaire as well as a later variation from Rubin's study of friendship, proximity, and self-disclosure [48]. We could not use the items in either study directly, for two reasons. First, some items were outdated (e.g. "my view of the Nixon administration") as these studies were conducted around 1958 and 1978 respectively; second, we wanted to study the effect of valence, which was not part of the original studies and was therefore not consistent in the items they used. Therefore, we used a structured approach to develop items in opposite-valence pairs. For each category, we examined items from Jourard's and Rubin's studies to identify representative sub-topics in this category. For example, for "Attitudes and Opinions", both studies had items about attitudes towards religion. We then developed a pair of items related to this sub-topic, as in "political or religious views that I admire" (positive valence) and "my indifference or dislikes about certain political or religious views" (negative valence). We developed three positive-negative pairs for each of the six categories, resulting in 36 items in total, 18 positive and 18 negative. Two independent raters evaluated the valence of items, and differences were resolved through discussion and re-developing the items until agreement was reached.

We validated the items in several ways before and after the data collection. We were interested not in complete agreement but in items that are consistently rated for intimacy, allowing for personal differences. As a pre-test, we asked 40 AMT workers to rate the intimacy of each item (on a 1–7 scale). We verified that the intra-item scores are not divergent (SD < 2) and that the mean items ratings fit a normal distribution from 1–7, covering a spread of intimacy levels.

The outcome of this process was a set of items in different categories capturing a range of intimacy score and representing both positive and negative valence, as shown in Table 1. For further validation, Table 1 also shows the item intimacy score distributions from the final experiment, including means and standard deviations for each. As we explain below, the participants were also asked to rate how comfortable they would be disclosing about these items. The self-disclosure items, as well as the full source data collected in this work, are available on the ACM Portal page and on Github.[3]

Questionnaire Flow

After agreeing to the consent form, each participant completed the following four stages of the questionnaire.

Item Intimacy-Level Rating

In this phase, the participants were shown the developed items in a random order, and were asked to rate the intimacy of these items on an 7-point scale, with endpoints from "Not intimate at all" to "Extremely intimate", following Rubin [48]. As mentioned above, the distributions of these scores for each item as shown in Table 1.

Platform Description and Verification

Next, participants were randomly assigned to one Identification (Real Name or Anonymous) and one Audience (Social Ties or Nearby) treatment, our between-subject experimental manipulation. Based on the assignment, we introduced the participant to the platform setup through textual description accompanied with a mock-up of the platform, designed to help the participants understand the identification and audience design. The mock-ups and textual descriptions are shown in Figure 1, corresponding to the 2 × 2 experimental manipulation. In *Real Name* conditions, each post is identified with a placeholders for a name and a photo. In *Anonymous* conditions, none of the identifiers are present. In *Social Ties* conditions, the source of the friend is indicated (e.g. mobile contacts or Facebook friends). In *Nearby* conditions, posts are presented in a map view as a reinforcement reminding the participant that their post is visible to people within one mile distance.

After the participant was introduced to the platform, they were asked to answer two verification questions that check their understanding of the platform design: (1) "In this application, do you know the real name of people who created specific posts?" and (2) "In this application, whose posts can you see?" If the participant failed to respond correctly, the system would display error message in red directing the participant to re-read the platform description and try answering again. Only when the participant responded to to the verification questions correctly did the system proceed to the next section of the questionnaire.

Item Self-Disclosure Comfort Rating

In this phase, we collected from each participant self-disclosure scores for each of the 36 items. The items were presented in a random order, different than when we collected the intimacy rating. For each item, the participants were

(a) Real Name / Social Ties (RS)   (b) Anonymous / Social Ties (AS)

(c) Real Name / Nearby (RN)   (d) Anonymous / Nearby (AN)

Identification = Real Name In this application, your real name is shown to other users. Upon registration, we will obtain a verified photo of you, your real name, and your email address. Your photo and name will be displayed together with any trace that you leave on the platform, such as posting, commenting, and liking.

Identification = Anonymous In this application, your identity is completely unknown to others. There is no registration process. Upon downloading the app, you can browse, post, comment, like other user's posts anonymously. There is no username, icon, or any other identifier attached to any posts you create.

Audience = Social Ties: In this application, you can share your thoughts with your friends. The application will obtain your mobile contacts, as well as Facebook friend list. Then, you can write short posts and publish them in a feed. Only your friends and contacts could see and comment on your post.

Audience = Nearby: In this application, you can share your thoughts with people near you. The application will obtain your geolocation. Then, you can write short posts and publish them in a feed. Only people within one-mile of where your post is published could see and comment on your post. Others won't see your actual location.

Figure 1. Mock-ups and textual descriptions used in experiment

asked to respond if they *feel comfortable* posting about this aspect of themselves to the platform we presented to them.

This "comfort" question is an adaptation from the disclosure scale used in Rubin's paper [48], tailored specifically to ask about *attitude* towards disclosing instead of *likelihood of action* or appropriateness of disclosure [4]. We asked about

M	SD	Text		M	SD	Text
Tastes and Interests						
1.91	1.45	My favorite foods, the ways I like food prepared		2.27	1.41	The kind of music, books, movies or TV shows that I cannot bear
1.90	1.36	My favorite singers, movie stars, writers, places, and brands		3.87	1.65	People or organizations that I strongly dislike
2.59	1.42	The kind of party, or social gathering that I like the best		3.01	1.52	The type of social gathering that would bore me, or that I wouldn't enjoy
Attitudes and Opinions						
3.16	1.62	Political or religious views that I admire		4.03	1.68	My indifference or dislikes about certain political or religious views
2.21	1.43	Positive comments about the progressing of society		2.99	1.45	My criticism about persisting social problems such as poverty and injustice
3.93	1.54	Support for certain controversial choices such as dropping out of school		4.74	1.50	Skepticism about the choices that people make, e.g., career choices, marriage
Work or Studies						
2.59	1.50	What I enjoy, and get the most satisfaction from in my present work or study		3.95	1.55	Pressures and strains in my work or study
2.90	1.45	What I feel are my special strong points for my work or study		2.71	1.51	What I find to be the most boring and unenjoyable in work or study
3.14	1.46	Things that I accomplish or achieved in work or study		4.20	1.57	Frustrations about how my work or study not being valued at all
Economic and Social Status						
4.16	1.45	Items or signals about how wealthy I am, e.g., luxury trips, accessories		5.62	1.50	Pressing need for money right now, e.g., outstanding bills, debts
4.12	1.56	Optimism about my future financial worth, e.g., getting job offers		4.90	1.46	Pessimistic views about my own future employment prospects and salaries
3.36	1.43	Connections to people who have high social status		5.30	1.47	Feelings of inferiority economically or socially compared to others around
Interpersonal Relations and Self-Concept						
3.60	1.83	Having good times with my significant other		5.36	1.56	Disappointments or bad experiences I have had in romantic relationship
2.99	1.59	Missing home or old friends		5.40	1.48	Things that I dislike or resent about my friends
3.86	1.54	Things that make me especially proud of myself		5.43	1.66	What I dislike the most about myself
Physical Appearance and Sex						
6.34	1.10	Confidence in my sexual adequacy		6.26	1.32	Disappointments in past sexual activity or fear for the first sexual experience
4.32	1.61	My standards about attractiveness of an ideal partner		4.69	1.66	Things I don't like about my appearance – nose, eyes, hair, skin, etc.
3.39	1.71	Happiness for being fit, healthy, and attractive		4.10	1.58	Efforts to improve physical appearance, such as exercise, diet, surgery

Table 1. Items and their intimacy ratings, organized by category. Positive valence items are on the left.

comfort levels and not likelihood or appropriateness to control for other factors could influence the use of social media, e.g. affordance of the service. In a preliminary study we found that when asked about likelihood to disclose, participants marked items that they are not likely to disclose not just due to self-disclosure risk and concerns, but also (and significantly) due to ideas of whether the content would be interesting or not. We believe that asking about comfort level mitigated some of these limitations; the linear relationship between intimacy and disclosure score in our final data provides evidence that this assumption is correct. Therefore, we asked about comfort level, using a 4-point scale as follows:

- *Very uncomfortable*: you will be extremely concerned about posting anything about this aspect of yourself.

- *Somewhat uncomfortable*: you have some concerns about posting about this aspect of yourself, but you might still do it with very little detail.

- *Somewhat comfortable*: you don't mind posting something about this aspect of yourself, but probably won't do so fully or in great detail.

- *Very comfortable*: you don't mind posting fully or in great detail about this aspect of yourself.

Demographic Information
The questionnaire concluded by collecting the demographic information of the participant including age, gender, geographical area (urban, suburban, rural) as well as their frequency of use of social media (frequently, occasionally, seldom, never) and their exposure to different social media platforms.

Participants
Our participants were recruited via Amazon Mechanical Turk (AMT). We required AMT workers to be in the U.S. (enforced via IP address matching), adult, with HIT approval rate equal or higher than 98% to ensure that participants have a history of quality responses. We paid $1.50 for each HIT, estimated to be 10 minutes of work (for a fair wage of $9/hr). There were no other requirements for participation.

The recruitment process and study protocol were reviewed and approved for the inclusion of human subjects by the Cornell University Institutional Review Board (Protocol ID# 1504005550).

Hypotheses Pre-registration and Data Sharing
The study method and hypotheses were pre-registered[4] to enhance open science and promote transparency. Pre-registration reduces the risk of Type 1 errors caused by over-testing and "fishing" activities [24, 28]. As mentioned above, we also made available the full source data collected in this work, together with the self-disclosure items we developed.

Data Collection and Preparation
We received 307 responses from AMT workers who completed the questionnaire. The average completion time was 11 minutes and 22 seconds. Workers were only allowed to submit one response each, verified by their Turker ID.

We performed some pre-processing to filter out potential "spammers". While Amazon Mechanical Turk is a platform originally designed and built for performing human computation tasks, it was shown to be effective and valid for behavioral research with advantages such as subject pool diversity, low cost, and faster theory/experiment cycle [37]. These advantages come with the potential cost of some low quality responses. Especially for surveys, it was found that some responders submit very low entropy responses (e.g. consistently selecting the same option across all items) or alternate between a small number of options in a regular pattern [62]. In our case, each participant's intimacy ratings of 36 items are

---

[4] http://www.socialscienceregistry.org/trials/840

Condition	Male	Female	N
Real Name / Social Ties (RS)	30	37	67
Real Name / Nearby (RN)	36	32	68
Anonymous / Social Ties (AS)	29	36	65
Anonymous / Nearby (AN)	40	29	69
Total	135	134	269

Table 2. Participants

an indicator for data quality, as the ratings are designed and pre-tested for variation. We performed some careful filtering of the data based on the intimacy scores participants assigned, removing from the dataset individuals whom exhibited little variation of intimacy scores (e.g., always entered the same value) or whose mean intimacy score was extreme. More precisely, we filtered out participants whose mean was more than two standard deviations away from the average standard deviation of all participants, keeping participants whose mean intimacy score is between 2.14 to 5.64 (on a 1-7 intimacy scale). Similarly, we filtered participants whose responses' standard deviation were more than two standard deviations away from the distribution of all participants, keeping those with standard deviation values between 0.83 to 2.64. This procedure filtered out 38 participants (12.3%) and left us with 269 participants. Of course, this procedure may result in removing several legitimate participants. As a robustness test, we ran our models with the full participant data and verified that the results were aligned with those reported below, albeit with a slight decrease in the models' R^2.

The distribution in each condition of the 269 (after filtering) participants is shown in Table 2. To ensure randomization was done correctly, we performed a randomization check and verified there were no significant differences between the participants in each group in terms of the demographics we collected, including gender, age, type of residence area, or social media use.

RESULTS

The focus of our experiment is the relationship between intimacy and willingness to self disclose, and how this relationship is impacted by the various factors of our mixed factorial design. Recall that the data collected from our experiment is a set of 36 item observations from each participant. Each observation is a pair consisting of the intimacy level of the item, and the comfort level to self-disclose about this item. Each participant is assigned one of the four conditions (*RS*, *RN*, *AS*, *AN*) viewing the same 36 items. We had 9,684 such observations in total.

As we assume that the relationship between intimacy and self disclosure is linear, we use three different linear models for the analysis, reporting only on the result of a Multiple Linear Regression for simplicity (results of other models largely agree). The use of more sophisticated models could be called for since for each participant the observations are non-independent (for example, consider a participant with natural tendencies to disclose at a higher rate). The Intraclass Correlation Coefficient (ICC) for our data is 0.23, a level suggesting that there are clustering effects, but the effects are not severe. Indeed, when we ran multilevel regression and re-

Variables	1 Estimate (SE)	2 Estimate (SE)	3 Estimate (SE)	4 Estimate (SE)
(Intercept)	1.16*** (.02)	1.14*** (.02)	.99*** (.02)	.97*** (.02)
Main Effects				
Intimacy	-.30*** (.00)	-.27*** (.01)	-.26*** (.01)	-.26*** (.01)
Identification (Ref: Real Name)	.37*** (.02)	.41*** (.03)	.41*** (.03)	.51*** (.03)
Audience (Ref: Nearby)	.20*** (.02)	.24*** (.03)	.24*** (.02)	.19*** (.03)
Interactions				
Audience × Identification	—	-.09** (.04)	-.09* (.04)	-.09** (.04)
Audience × Intimacy	—	-.08*** (.01)	-.08*** (.01)	-.07*** (.01)
Audience × Identification × Intimacy	—	.04* (.02)	.04* (.02)	.04* (.02)
Identification × Intimacy	—	-.00 (.01)	.00 (.01)	-.01 (.01)
Valence				
Valence (Ref: Negative)	—	—	.33*** (.02)	.37*** (.03)
Valence × Intimacy	—	—	.03* (.01)	.03** (.01)
Valence × Identification	—	—	—	-.19*** (.04)
Valence × Audience	—	—	—	.11** (.04)
Adjusted R^2	.32	.32	.34	.35

Significance codes: $^*p < .05$, $^{**}p < .01$, $^{***}p < .001$

Table 3. Regression coefficients predicting self-disclosure

gression with cluster-robust standard errors, the results were almost identical, with only one minor difference in regards to the outcome of *H2c*.[5] We therefore report the results using a the simple linear regression model. Note that for simplicity of analysis and discussion, the intimacy scores, originally rated from 1 to 7 from "Not intimate at all" to "Extremely intimate", were re-centered such that 0 is "medium intimacy" making the intercept of the regression more meaningful. The dependent variable, self-disclosure, was coded from 0 ("Very uncomfortable") to 3 ("Very comfortable").

The results of four different models of multiple linear regression are shown in Table 3 and discussed in detail below.

Platform and Intimacy Effects
We start with Model 1, examining the main effects of intimacy and the platform features, Identification and Audience, on self-disclosure. As hypothesized, Identification and Audience type both significantly impact the self-disclosure baseline (i.e., higher level of sharing likelihood, independent of the content intimacy). There was a significant positive effect of Identification (reference: Real Name) on self-disclosure level, supporting *H1a* that anonymity *increased* the self-disclosure baseline. There was a significant positive effect of Audience (reference: Nearby) on self-disclosure, suggesting that audience of social ties increased self-disclosure compared to the audience of people nearby, supporting *H1b*. There was also a significant negative effect of intimacy on self-disclosure, suggesting a regulation effect where people feel less comfortable self-disclosing content of high intimacy,

---

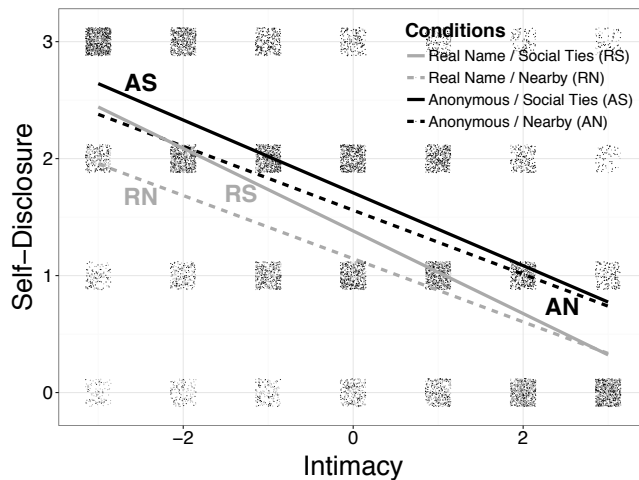[5]As the data is available, interested readers can run their favorite model.

Figure 2. The relationship between self-disclosure and intimacy for different platform conditions

supporting *H2a*. All these results are highly stable across all models in Table 3.

Figure 2 visually summarizes the results in Model 1 (and Model 2). The figure shows the raw data and trend lines for the relationship between intimacy and self-disclosure for the different Audience and Identification conditions. In the figure, the X-axis captures the (re-centered) intimacy score, and the Y-axis represents the disclosure level. The black lines represent the Anonymous conditions, and the gray lines represent the Real Name conditions. The solid lines are the trend lines for the Social Ties audience, and the dashed lines represent the Nearby audience. For example, the top line in the figure is the trend line for the Anonymous / Social Ties (*AS*) condition. The scatter plot dots in the background are individual data points of intimacy and disclosure item ratings (with jitter), showing that the core relationship between these two variables is indeed linear.

The results of Model 1 are visually evident in Figure 2. The regulation effect of intimacy is seen in the negative slope for all trend lines across conditions. The main effect of anonymity can be seen as the trend lines in black (Anonymous conditions) are consistently higher than the corresponding ones in gray (Real Name). Similarly, the increased self-disclosure effect of social ties can be seen by comparing the solid Social Ties lines to the corresponding dashed Nearby lines in Figure 2.

Identification and Audience: Moderation Effects
Next, we examine whether the Identification and Audience treatments moderate intimacy's regulation effect by adding the interaction variables in the regression model (Table 3, Model 2). We first look at the case when the Audience is Social Ties (the reference for the Audience variable is the Nearby condition). There is a negative interaction between intimacy and Audience, suggesting that, in the Social Ties condition, the regulation effect of intimacy is stronger regardless of the Identification condition. The three-way in-

teraction shows that the moderating effect of Social Ties on intimacy's regulation is weaker for the Anonymous condition (i.e. stronger for Real Name condition) as hypothesized in *H2b*.

In Figure 2, the stronger regulation of social ties is visible by the solid lines having steeper slopes than the corresponding dashed lines, the Nearby conditions. We only hypothesized about this effect for the Real Name condition (gray lines); nevertheless, the effect seems to hold for the anonymous condition as well. The model and figure both suggest that intimacy more strongly regulates self-disclosure when the audience is Social Ties than Nearby, regardless of anonymity. All these results are highly stable across all models in Table 3.

Next, the model shows there are no effects of the interaction between Identification and intimacy, suggesting that anonymity does not play a moderating role by itself on the *slope* (recall that anonymity does raise the baseline for disclosure for all intimacy levels). But with the positive three-way interaction term, the model shows a not-as-steep slope of intimacy's effect on disclosure under the Anonymous / Social Ties (*AS*) condition (as the identification variable baseline is Real Name and Audience baseline is Nearby), thus supporting *H2c*. In Figure 2, this effect is visible by looking at the Social Ties conditions, *AS* and *RS* (solid lines); the anonymous (*AS*, in black) slope is more flat than the gray *RS* line. This result is highly stable across all models in Table 3.

Finally, we examine the Nearby condition (reference for the Audience variable). The interactions involving Audience are now zero, and the non-significant term of the Identification interaction with intimacy suggests that anonymity does not alter intimacy's regulation effect in this condition, rejecting *H2d*. This result is also clearly visible in Figure 2 that shows a negative slope for the *AN* condition (black dashed line).

Valence Effects
We now examine how valence of content affects disclosure. As Table 3 (Model 3) shows, positive valence increased self-disclosure baseline in all platform conditions, even as the model controls for intimacy (note that negative valence items tend to have higher intimacy scores). There is a moderating effect of the interaction between valence and anonymity (Model 4), suggesting that in Real Name platforms, participants are even more comfortable sharing positive compared to negative valence content. These results provide support to *H3a*, indicating that anonymity somewhat "closes the gap" between positive and negative valence items, and reject *H3b*, as this trend was not enough to override the overall effect of valence.

The effects of valence could be observed intuitively in Figure 3, illustrating the differences between the negative valence (red) and positive valence (green) for each condition. In the figure, we also show the individual data points, also colored by valence. The figure demonstrates that across the board, disclosure levels for positive content are higher than negative (green trend lines are higher than the red). The moderating effect of anonymity is also seen as the gap between
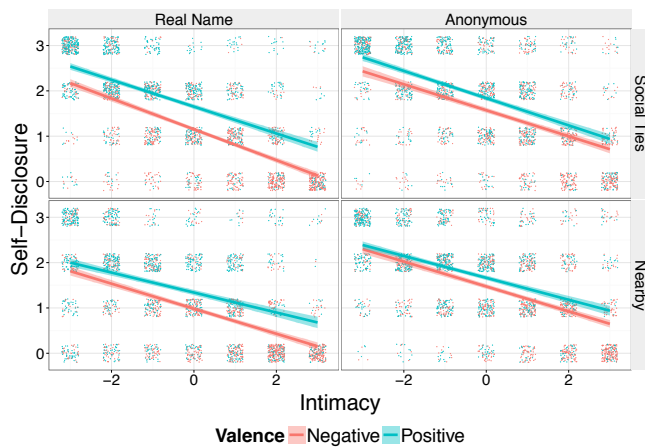
Figure 3. The relationship between self-disclosure and intimacy for positive and negative valence items, in different platform conditions

each two trend lines under the anonymous conditions (on the right) are smaller than the gaps in the real-name conditions.

Our results are summarized in Table 4.

H	Hypothesis Summary	Support?
H1a	Disclosure is greater with anonymity vs. real name conditions	Yes
H1b	Disclosure is greater with social ties vs. nearby conditions	Yes
H2a	Regulation of intimacy in real-name conditions	Yes[a]
H2b	Stronger regulation in *RS* than *RN*	Yes[b]
H2c	Weaker regulation in *AS* than *RS*	Yes
H2d	No regulation in *AN*	No
H3a	Negative valence reduces disclosure in real-name conditions	Yes
H3b	Negative valence increases disclosure in anonymous conditions	No[c]

Table 4. Summary of hypotheses and evidence.

[a]Result holds for anonymous conditions as well
[b]Result holds for *AS* versus *AN* as well
[c]Reverse effect was found

DISCUSSION

Our findings reaffirm previous findings regarding the negative relationship between intimacy of content and self-disclosure [1, 30, 47]. At the same time, the findings shed new light on this relationship in different online social settings, including under different conditions of identification and audience. The findings could help empirically explain many of the observed real-world trends and observations about disclosure and anonymity [15, 59] and highlight outstanding issue and questions regarding self-disclosure in online environments. We discuss the findings in relation to the key independent variables in the study.

Intimacy. Our results indicate that intimacy of content regulates self-disclosure across all conditions and variables in our analysis. While this regulation is moderated in some interesting ways by other variables (more on that below), it is important to note that we did not identify any settings in which this regulation disappears. The regulation effect of intimacy reaffirms classical findings on self-disclosure [1, 30, 47], but this is the first time the relationship is empirically demonstrated

in social media settings. Other social media studies of disclosure studied already-posted content [5] (asking about appropriateness), or compared content on different platforms, introducing many potential confounds such as the platform design.

What this means: there is always inhibition. While we hypothesized that under some conditions there would be complete disinhibition: people's disclosure decision would not be affected by the content intimacy, our results show that such conditions are very unlikely. It would be a very difficult challenge to create systems where people feel completely unconstrained, as the intimacy of content will always regulate disclosure internally.

Audience. Our findings show that in general, people are more comfortable disclosing to social ties than to people nearby, both under anonymous and under real-name settings. This is the first time such a relationship has been directly and empirically measured. Even without any identity marker, the audience still matters for disclosure decisions. An important consideration, though, is that this finding mostly holds for the baseline, i.e. on average across intimacy levels. Critically, our results demonstrate that when sharing with social ties, the disclosure is regulated much more sharply as the intimacy of content grows. This aversion from sharing the more intimate content with social ties seems to hold for both anonymous and non-anonymous settings.

What this means: There are many benefits to self-disclosure, especially when disclosing to friends [11, 48]. As our results make clear, people are willing to share online more with friends, but this tendency weakens when the content becomes more intimate. This regulation is potentially harmful. People are known to get support in social media platforms when they post intimate content, e.g. feelings or expressions of loneliness [10, 31]. Withholding such intimate content means people are less likely to get the support they need, although it is possible that they would seek this support through other channels [4]. Future work could examine how anonymity, or other design elements, can enhance the sharing of intimate content on these services.

Anonymity. Our findings empirically show that anonymity allows for more disclosure, across the board. While disinhibition effects of anonymity were documented in various settings in the past [53, 13], the effect was not measured in a controlled manner. Using our data, we could show that anonymity increases the baseline for self-disclosure, and in addition, that it *sometimes* makes the regulation effect of intimacy on self-disclosure slightly weaker.

What this means: anonymous services make it easier than real-name setups to disclose any kind of content. The overall increase in disclosure is probably due to the reduced risk in disclosing anonymously [3, 17, 35] resulting in a disinhibition effect [53]. Being able to disclose information anonymously to a specific audience could help improve the benefit/risk ratio of online systems and potentially greatly beneficial to groups of people that are vulnerable online. Following Facebook and Google Plus decision to enforce real name poli-

cies in 2012, it has been argued that women benefit a lot from online anonymity [9]. In another example, in the domain of health care, cancer patients distinguish between clinical (sensitive but beneficial to share) and identity information (more risky and less relevant to share) [23], which is critical as patients can gain more when they make disclosures (e.g., social support) [56].

At the same time, our results show a surprisingly small effect of anonymity on the *regulation* of disclosure by intimacy, with one exception. When sharing with social ties, where there intimacy's regulation is strong, anonymity slightly weakens the regulation. Count this finding as "somewhat good news" – especially given the potential benefits of sharing with friends mentioned above.

Valence. Based on our findings, negative valence moderates disclosure in all settings. People are less likely to share items with negative valence than positive, even when controlling for intimacy score. This result is consistent with more recent formulations of functional self-disclosure [5] and with findings showing that people think negative valence is less appropriate to disclose on Facebook and elsewhere [4, 11]. To our knowledge, this is the first study that empirically examines the effect of valence on disclosure in any settings, making our developed items into another contribution of this work. Beyond the main effect of valence, our results also indicate the inhibiting effect of negative valence has *less* impact in anonymous settings. Anonymity allows for more negative valence than Real Name settings, for any type of audience. While consistent with functional self-discourse models [5, 11], to the best of our knowledge, the present study is the first to empirically demonstrate these divergent effects of negative self-disclosure in anonymous and real-name settings.

What this means: anonymous platforms may allow people to self-disclose more intimate negative valence compared to real-name networks, providing an important outlet for such content which is harder to share [6, 10], and allowing the potential benefits of writing about negative aspects of one's life [43]. Such content is harder to share on non-anonymous platforms such as Twitter or Facebook where negative posts are likely to violate self-presentation goals [5].

Recall (or see in Table 1) that our operationalization of valence does not necessarily reflect an attitude towards *others*, and negative valence includes negative statements of self-disclosure, or negative reflections on other topics. Nevertheless, these results may help explain and illustrate the potential and reported toxicity of anonymous apps, where negative content is apparent in anonymous platforms even when they have a context of sharing like people nearby of social ties. As the results show, people are simply more comfortable sharing negative valence in anonymous platforms.

Other factors. Self-disclosure decisions could results from many different factors that were not directly measured in our study. For example, especially in proximity-based social media, physical safety could be a concern, adding to personality-driven risk factors. Another factor that may impact sharing is personality traits such as introversion, which are known to af-

fect how people use social media [46]. It would also be interesting to study the effect of norms, especially in new anonymous services where people rely more on others' behavior as norms in the environment are are not yet well-established or understood [14, 52]. Social identity model of deindividuation (SIDE) predicts that people conform more to the norms when anonymous compared to when they are identified [45]. Relative to Twitter, Facebook, or other known social media, norms on the anonymous platforms are clearly less well defined. An extensive body of research has shown that social norms can guide and constrain behavior [14], including in online communities [12], website commenting [52], and within anonymous groups [44].

Limitations. Our study has a number of limitations that are important to acknowledge. First, we rely in this work on indications of comfort in disclosure, as we could not effectively measure *actual* sharing behaviors – of course, such a study would be hard to do while controlling for intimacy levels in regards to content that is not posted. In addition, we used a limited set of self-disclosure items, in specific categories. These statements and categories do not fully reflect the content that is shared on these platforms. Another limitation is the representativeness of our sample population. All participants were recruited through Amazon Mechanical Turk under specific conditions, resulting in potential sampling bias. However, other research had established the effectiveness of AMT for behavioral studies [37]. Finally, we study hypothetical, though well-defined social media platforms. We believe the general trends in our results are likely to generalize to real-world services like Facebook or Yik Yak.

CONCLUSIONS
With the proliferation of new applications that allow people to share information with others in unidentifiable and anonymous settings, it is important to understand how the affordances of these services could affect disclosure, and how the dynamics of these platforms are different than "traditional" social media. Using data from a mixed-factorial survey experiment, we provided important empirical insights about the relationship between intimacy and self-disclosure in social media, and how identification and audience features of the platform moderate such relationship for positive and negative valence content. Our results, affirming and extending previous work in offline and online settings, help us understand and explain key dynamics in social media platforms and differences between them. Our study sheds more light on how people balance the benefits and risks of various kinds of disclosure online, and is critical for advancing self-disclosure theories.

REFERENCES

1. Irwin Altman and William W Haythorn. 1965. Interpersonal exchange in isolation. *Sociometry* (1965), 411–426.

2. Irwin Altman and Dalmas Taylor. 1973. Social penetration theory. *New York: Holt, Rinehart and Mnston* (1973).

3. John A Bargh, Katelyn YA McKenna, and Grainne M Fitzsimons. 2002. Can you see the real me? Activation and expression of the true self on the Internet. *Journal of social issues* 58, 1 (2002), 33–48.

4. Natalya N Bazarova. 2012. Public intimacy: Disclosure interpretation and social judgments on Facebook. *Journal of Communication* 62, 5 (2012), 815–832.

5. Natalya N Bazarova and Yoon Hyung Choi. 2014. Self-disclosure in social media: Extending the functional approach to disclosure motivations and characteristics on social network sites. *Journal of Communication* 64, 4 (2014), 635–657.

6. Natalya N Bazarova, Yoon Hyung Choi, Victoria Schwanda Sosik, Dan Cosley, and Janis Whitlock. Social sharing of emotions on Facebook: Channel differences, satisfaction, and replies. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15)*.

7. Michael S Bernstein, Andrés Monroy-Hernández, Drew Harry, Paul André, Katrina Panovich, and Gregory G Vargas. 2011. 4chan and /b/: An analysis of anonymity and ephemerality in a large online community. In *Proceedings of the 5th AAAI International Conference on Weblogs and Social Media (ICWSM '11)*.

8. Jeremy Birnholtz, Nicholas Aaron Ross Merola, and Arindam Paul. Is it weird to still be a virgin?: Anonymous, locally targeted questions on Facebook confession boards. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*.

9. danah boyd. 2012. The politics of real names. *Commun. ACM* 55, 8 (2012), 29–31.

10. Moira Burke and Mike Develin. 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 (CSCW '16)*.

11. Alan L Chaikin and Valerian J Derlega. 1974. Variables affecting the appropriateness of self-disclosure. *Journal of Consulting and Clinical Psychology* 42, 4 (1974), 588.

12. Yan Chen, F. Maxwell Harper, Joseph Konstan, and Sherry Xin Li. 2010. Social comparisons and contributions to online communities: A field experiment on MovieLens. *American Economic Review* 100, 4 (2010), 1358–98.

13. Kimberly M Christopherson. 2007. The positive and negative implications of anonymity in Internet social interactions:On the Internet, Nobody Knows Youre a Dog. *Computers in Human Behavior* 23, 6 (2007), 3038–3056.

14. Robert B Cialdini and Melanie R Trost. 1998. Social influence: Social norms, conformity and compliance. In *The Handbook of Social Psychology, Vols. 1 and 2 (4th ed.)*, Daniel T. Gilbert, Susan T. Fiske, and Gardner Lindzey (Eds.). McGraw-Hill, 151–192.

15. Denzil Correa, Leandro Araújo Silva, Mainack Mondal, Fabrício Benevenuto, and Krishna P Gummadi. The Many Shades of Anonymity: Characterizing Anonymous Social Media Content. In *Proceedings of the 9th AAAI International Conference on Weblogs and Social Media (ICWSM '15)*.

16. Paul C Cozby. 1973. Self-disclosure: a literature review. *Psychological bulletin* 79, 2 (1973), 73.

17. Valerian J Derlega and Alan L Chaikin. 1977. Privacy and self-disclosure in social relationships. *Journal of Social Issues* 33, 3 (1977), 102–115.

18. Valerian J Derlega, Janusz Grzelak, and others. 1979. Appropriateness of self-disclosure. *Self-disclosure: Origins, patterns, and implications of openness in interpersonal relationships* (1979), 151–176.

19. Valerian J Derlega, Barbara A Winstead, and Kathryn Greene. 2008. Self-disclosure and starting a close relationship. *Handbook of relationship initiation* (2008), 153–174.

20. Joan Morris DiMicco and David R. Millen. 2007. Identity management: Multiple presentations of self in Facebook. In *Proceedings of the 2007 international ACM conference on Supporting group work (GROUP '07)*.

21. Howard J Ehrlich and David B Graeven. 1971. Reciprocal self-disclosure in a dyad. *Journal of Experimental Social Psychology* 7, 4 (1971), 389–400.

22. Shelly D. Farnham and Elizabeth F. Churchill. 2011. Faceted identity, faceted lives: Social and technical issues with being yourself online. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work (CSCW '11)*.

23. Jeana Frost, Ivar E Vermeulen, and Nienke Beekers. 2014. Anonymity versus privacy: selective information sharing in online cancer communities. *Journal of medical Internet research* 16, 5 (2014).

24. Andrew Gelman. 2013. Preregistration of studies and mock reports. *Political Analysis* 21, 1 (2013), 40–41.

25. Marvin R Goldfried, Lisa A Burckell, and Catherine Eubanks-Carter. 2003. Therapist self-disclosure in cognitive-behavior therapy. *Journal of clinical psychology* 59, 5 (2003), 555–568.

26. Oliver L. Haimson, Jed R. Brubaker, Lynn Dombrowski, and Gillian R. Hayes. 2015. Disclosure, stress, and support during gender transition on Facebook. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15)*.

27. Erin E Hollenbaugh and Marcia K Everett. 2013. The effects of anonymity on self-disclosure in blogs: An application of the online disinhibition effect. *Journal of Computer-Mediated Communication* 18, 3 (2013), 283–302.

28. Macartan Humphreys, Raul Sanchez de la Sierra, and Peter Van der Windt. 2013. Fishing, commitment, and communication: A proposal for comprehensive nonbinding research registration. *Political Analysis* 21, 1 (2013), 1–20.

29. Sidney M Jourard. 1971. Self-disclosure: An experimental analysis of the transparent self. (1971).

30. Sidney M Jourard and Paul Lasakow. 1958. Some factors in self-disclosure. *The Journal of Abnormal and Social Psychology* 56, 1 (1958), 91.

31. Funda Kivran-Swaine, Jeremy Ting, Jed R Brubaker, Rannie Teodoro, and Mor Naaman. 2014. Understanding loneliness in social awareness streams: Expressions and responses. In *Proceedings of the 8th AAAI International Conference on Weblogs and Social Media (ICWSM '14)*.

32. Lee Knuttila. 2011. User unknown: 4chan, anonymity and contingency. *First Monday* 16, 10 (2011).

33. Cliff Lampe, Jessica Vitak, Rebecca Gray, and Nicole Ellison. Perceptions of Facebook's value as an information source. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*.

34. Julie D Lane and Daniel M Wegner. 1995. The cognitive consequences of secrecy. *Journal of Personality and Social Psychology* 69, 2 (1995), 237.

35. Noam Lapidot-Lefler and Azy Barak. 2012. Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition. *Computers in Human Behavior* 28, 2 (2012), 434–443.

36. Alice E Marwick and others. 2011. I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New media & society* 13, 1 (2011), 114–133.

37. Winter Mason and Siddharth Suri. 2012. Conducting behavioral research on Amazons Mechanical Turk. *Behavior research methods* 44, 1 (2012), 1–23.

38. Mark W Newman, Debra Lauterbach, Sean A Munson, Paul Resnick, and Margaret E Morris. It's not that I don't have problems, I'm just not putting them on Facebook: challenges and opportunities in using online social networks for health. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work (CSCW '11)*.

39. Julia Omarzu. 2000. A disclosure decision model: Determining how and when individuals will self-disclose. *Personality and Social Psychology Review* 4, 2 (2000), 174–185.

40. W Babnett Pearce and Stewart M Sharp. 1973. Self-disclosing communication. *Journal of Communication* 23, 4 (1973), 409–425.

41. James W Pennebaker. 1989. Confession, inhibition, and disease. *Advances in experimental social psychology* 22 (1989), 211–244.

42. James W Pennebaker. 1993. Putting stress into words: Health, linguistic, and therapeutic implications. *Behaviour research and therapy* 31, 6 (1993), 539–548.

43. James W Pennebaker. 2012. *Opening up: The healing power of expressing emotions*. Guilford Press.

44. Tom Postmes, Russell Spears, Khaled Sakhel, and Daphne de Groot. 2001. Social influence in computer-mediated communication: The effects of anonymity on group behavior. *Personality and Social Psychology Bulletin* 27, 10 (2001), 1243–1254.

45. Stephen D Reicher, Russell Spears, and Tom Postmes. 1995. A social identity model of deindividuation phenomena. *European review of social psychology* 6, 1 (1995), 161–198.

46. Craig Ross, Emily S Orr, Mia Sisic, Jaime M Arseneault, Mary G Simmering, and R Robert Orr. 2009. Personality and motivations associated with Facebook use. *Computers in human behavior* 25, 2 (2009), 578–586.

47. Zick Rubin. 1975. Disclosing oneself to a stranger: Reciprocity and its limits. *Journal of Experimental Social Psychology* 11, 3 (1975), 233–260.

48. Zick Rubin and Stephen Shenker. 1978. Friendship, proximity, and self-disclosure. *Journal of Personality* 46, 1 (1978), 1–22.

49. Richard M Ryan and Edward L Deci. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist* 55, 1 (2000), 68.

50. Georg Simmel. 1950. The stranger. *The sociology of Georg Simmel* (1950), 402–408.

51. William B Stiles. 1987. I have to talk to somebody. In *Self-disclosure*. Springer, 257–282.

52. Abhay Sukumaran, Stephanie Vezich, Melanie McHugh, and Clifford Nass. 2011. Normative influences on thoughtful online participation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*.

53. John Suler. 2004. The online disinhibition effect. *Cyberpsychology & behavior* 7, 3 (2004), 321–326.

54. Diana I Tamir and Jason P Mitchell. 2012. Disclosing information about the self is intrinsically rewarding. *Proceedings of the National Academy of Sciences* 109, 21 (2012), 8038–8043.

55. Sherry Turkle. 1995. *Life on the screen: Identity in the age of the Internet*. Simon & Schuster Trade.

56. Jeanine Warisse Turner, Jean A Grube, and Jennifer Meyers. 2001. Developing an optimal match within online communities: An exploration of CMC support communities and traditional support. *Journal of Communication* 51, 2 (2001), 231–251.

57. Jessica Vitak and Jinyoung Kim. 2014. "You can't block people offline": Examining how Facebook's affordances shape the disclosure process. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing (CSCW '14)*.

58. Elizabeth Needham Waddell and Peter A Messeri. 2006. Social support, disclosure, and use of antiretroviral therapy. *AIDS and Behavior* 10, 3 (2006), 263–272.

59. Gang Wang, Bolun Wang, Tianyi Wang, Ana Nika, Haitao Zheng, and Ben Y Zhao. Whispers in the dark: analysis of an anonymous social network. In *Proceedings of the 2014 Conference on Internet Measurement Conference (IMC '14)*.

60. Daniel M Wegner. 1989. *White bears and other unwanted thoughts: Suppression, obsession, and the psychology of mental control*. Penguin Press.

61. Daniel M Wegner, David J Schneider, Samuel R Carter, and Teri L White. 1987. Paradoxical effects of thought suppression. *Journal of personality and social psychology* 53, 1 (1987), 5.

62. Dongqing Zhu and Ben Carterette. 2010. An analysis of assessor behavior in crowdsourced preference judgments. In *SIGIR 2010 workshop on crowdsourcing for search evaluation*. 17–20.