"I Am 30F and Need Advice!": A Mixed-Method Analysis of the Effects of Advice-Seekers' Self-Disclosure on Received Replies

Yixin Chen, Scott A. Hale, Bernie Hogan
Oxford Internet Institute, University of Oxford
{yixin.chen, scott.hale, bernie.hogan}@oii.ox.ac.uk

Abstract

In community question answering sites, users can easily make a post to ask questions or seek advice. Others volunteer replies to these posts to provide answers of varying quality, detail, and helpfulness. In the advice-seeking process, self-disclosure enables posters to provide a relatable context for their requests but comes at a cost of greater identifiability. We focus on the "r/Advice" Reddit community and present a mixed-method study on how self-disclosure of advice-seekers shapes the prevalence and detail of the feedback received. We focus particularly on age and gender disclosure as both are reliably detected and normatively considered in the context of giving advice. We use both hurdle negative binomial regression models and discourse analysis to examine the relationship between self-disclosure and the replies received and explore themes related to disclosure. The results show that advice-seekers' age or gender disclosure correlates with more replies and more helpful replies, but the effects of age and gender disclosure are not additive. We also find both reciprocity and homophily effects in disclosure as reply-givers are more likely to self-disclose when the advice-seeker does so. The lack of additive effects alongside the thematic analysis suggests disclosure practices are used to elicit sufficient credibility or basis for empathy, whereas too much or too little disclosure creates uncertainty or inhibits the applicability of the received advice.

Introduction

Many online platforms can be characterised as Community Question Answering (CQA) websites or facilitate it as a practice. This practice tends to involve anonymous or pseudonymous exchanges where people ask questions, receive replies, and often react to the replies in some manner. These reactions vary from votes, to badges, to requests for clarification (Liu et al. 2014). The benefit of these sites comes from the presumed collective intelligence of the community (El Adlouni et al. 2019). The costs come from the potential for identifiability, particularly on sensitive topics (Vitak and Kim 2014). People asking questions on CQA sites sometimes first offer some personal information as context. Such self-disclosure can be necessary because it enables posters to introduce their situations and helps others better understand their requests (Fu, Chang, and Danescu-Niculescu-Mizil 2019).

Online self-disclosure has been found to have both benefits and risks. Benefits include helping increase intimacy and facilitate user participation (Li-Barber 2012), while risks include triggering social rejection and privacy leakage (Vitak and Kim 2014). Therefore, it is meaningful to examine how advice-seekers' self-disclosure affects received replies on CQA sites. Regarding the content of disclosure, age and gender are basic categories which are commonly disclosed in discussions of various topics; so, they are often studied (Balani and De Choudhury 2015; Lankton, McKnight, and Tripp 2017; Umar, Squicciarini, and Rajtmajer 2019). However, their effects on advice-seeking process are also not clear. Moreover, prior research revealed a reciprocity effect of general self-disclosure (Barak and Gluck-Ofri 2007) and proposed several theories to explain it (Wetzel and Wright-Buckley 1988; Rubin 1975; Worthy, Gary, and Kahn 1969), but it is not clear whether the reciprocity effect also exists on advice-seeking communities, and if so, how it is triggered.

This research fills the gap by examining the relationship between advice-seekers' self-disclosure (regarding age and gender) and the quantity and quality of the replies they received. We focus on a typical CQA forum "r/Advice" on Reddit and conduct a mixed-method analysis. Specifically, we use hurdle negative binomial regression to investigate the statistical relationship between self-disclosure in the open post and its replies, and use discourse analysis to explore how advice-seekers' self-disclosure relates to the nature or content of their requests and how that self-disclosure shapes the associated replies. Our discourse analysis indicates that advice-seekers disclose information about themselves to follow social norms, provide background information, increase credibility, and seek targeted advice. Additionally, we found that advice-seeking posts with age or gender disclosure receive more replies and more helpful replies. However, the results do not support the additive effects of these two types of self-disclosure. Moreover, we observed that the reciprocity effects exist for age and gender disclosure, which is used to signal credibility or communicate empathy. Overall, this study contributes to the understanding of the role of self-disclosure in the advice-seeking and advice-providing process, and provides insights into how to foster the development of CQA sites.

Related Work

Community Question Answering

Community Question Answering (CQA) has emerged as a popular tool to seek information online. On CQA sites such as Stack Overflow and some communities on Reddit, users can ask questions or seek advice conveniently by posting personalized requests, and other users can answer these questions or provide advice by replying to these posts and continuing the threads. In this way, the original poster can make use of the collective intelligence of other Internet users (El Adlouni et al. 2019).

Compared with the traditional offline question- or advice-asking approaches which are usually limited to one's personal network including people they know in real life, an advantage of CQA is anonymity or pseudonymity. Users often use pseudonymous or anonymous accounts instead of real names to make posts or replies in online communities to avoid being identified, and the anonymity or pseudonymity provides an open and disinhibiting platform for communication, especially when talking about sensitive or stigmatized experiences (Pavalanathan and De Choudhury 2015). This disinhibition could further make users more comfortable with self-disclosing (Tidwell and Walther 2002; Birnholtz, Merola, and Paul 2015). We will further discuss research on online self-disclosure in the next few sections.

Self-Disclosure in Online Communities

As users usually do not know each other on online CQA sites, self-disclosure, which is defined as a "process of making the self known to others" (Jourard and Lasakow 1958), is necessary for support/advice-seekers to provide contextual information and may trigger more appropriate and empathetic responses (Pfeil and Zaphiris 2007). Traditional offline self-disclosure is commonly known to increase mutual understanding and trust (Laurenceau, Barrett, and Pietromonaco 1998; Galegher, Sproull, and Kiesler 1998). Moving to online spaces, self-disclosure can similarly benefit both individuals and communities by helping establish personal connections (Taylor et al. 2019), increase intimacy, clarify identity, and obtain social approval (Vitak and Kim 2014). On the other hand, users face risks when disclosing about themselves, especially in online contexts, as self-disclosure may trigger social rejection, reduction of integrity, and concerns over self-presentation and privacy (Vitak and Kim 2014). Research has also involved specific categories of self-disclosure, of which age and gender are two popular and essential categories regardless of the discussion topic (Lankton, McKnight, and Tripp 2017; Umar, Squicciarini, and Rajtmajer 2019; Balani and De Choudhury 2015). Studies on these two disclosure categories also showed their context-specific benefits such as a higher level of person-centeredness and politeness in responses to support-seeking posts (Pan et al. 2020) and perceived helpfulness of product reviews (Sikdar et al. 2021), and risks such as stigma or stereotyping (Han, Huang, and Wang 2019).

Considering the importance of online self-disclosure, researchers have proposed various ways to measure it. Some researchers viewed self-disclosure as a single ordinal or continuous variable and used supervised learning to measure it (Balani and De Choudhury 2015; Wang, Burke, and Kraut 2016). However, different categories of self-disclosure may have different effects on online communication, so researchers further measured fine-grained self-disclosure categories. For example, Lankton, McKnight, and Tripp (2017) measured various types of self-disclosure through surveys. For large-scale datasets, studies considered popular disclosure categories such as age and gender in various platforms (e.g., Twitter (Emmery, Chrupała, and Daelemans 2017), mental health communities (Jagfeld et al. 2021), and news websites (Umar, Squicciarini, and Rajtmajer 2019)), and extracted disclosure from multiple fields such as accounts' preference setting and self-reported sentences.

In addition to the exploration of self-disclosure itself, research has found reciprocity in self-disclosure, which means one's self-disclosure elicits their communicating partner's self-disclosure (Barak and Gluck-Ofri 2007). There are various theories to explain that reciprocity, including trust theory, modeling theory, and social exchange theory. The trust theory posits that one's self-disclosure makes their interaction partner feel liked and trusted, and therefore reciprocate the disclosure as a sign of liking and trust (Wetzel and Wright-Buckley 1988). The modeling theory views reciprocal self-disclosure as a type of human mimicry where the initial self-disclosure serves as a cue for expected response (Rubin 1975). The social exchange theory claims that one's self-disclosure makes the communicating partner think they received something of value and feel an obligation to return disclosure as a reward (Worthy, Gary, and Kahn 1969). Moreover, research has found various benefits of reciprocal self-disclosure, such as facilitating trust and intimacy in offline therapy (Hanson 2005) and helping the original discloser feel supported in online communities (Andalibi et al. 2018).

Responses to Posts in Online Communities

As this study focuses on the effects of advice-seekers' self-disclosure on their received replies, we finally turn to studies investigating the dependent variable. "No response" is a critical challenge that users face on CQA sites (Li and King 2010). Whether or not a post can receive any reply and the total number of replies show the level of user engagement and whether a CQA site is successful (Kayes et al. 2015). Thus, a rich body of research has examined the factors that may influence the probability or the number of replies to posts, e.g., readability (Risch and Krestel 2020), topic (Park et al. 2020), and involvement of psychological processes (Maity, Kharb, and Mukherjee 2017). Some studies also found the effects of questioner-based features such as user reputation (Movshovitz-Attias et al. 2013), user experience (Liu and Jansen 2018), and the use of throwaway accounts (Ammari, Schoenebeck, and Romero 2019).

Receiving replies is not the end of the story. Replies can vary greatly in their content and quality (Welser et al. 2007) and not all of them are helpful or informative. CQA sites often allow users to show their opinions on received replies by voting for high-quality content, accepting a reply as the best answer, or replying to helpful replies (Liu et al. 2014). Therefore, in addition to the number of replies, researchers have also explored which posts would finally succeed in receiving satisfactory or acceptable replies, and similarly found the effects of both question-based factors and questioner-based factors (Anderson et al. 2012; Peng et al. 2021). However, the original posters' disclosure has not been considered.

Research Questions and Hypotheses

Although the studies above proposed models to measure the level of overall self-disclosure, it is not clear how self-disclosure affects online communication, especially in community question answering. There is also a notable gap in research concerning the examination of specific categories of self-disclosure and their aggregate impact, such as the prevalent and near-universal age and gender disclosure. Therefore, this study aims to understand how individuals' self-disclosure shapes the feedback they receive in advice-seeking communities. We structure these aims into broad research questions, and where appropriate, falsifiable hypotheses:

- ***RQ1***: How does advice-seekers' self-disclosure relate to the nature or content of their request?
- ***RQ2***: How does self-disclosure in advice-seeking posts relate to the quantity of replies (including helpful replies)?
  - ***H1***: Self-disclosure of either age or gender will elicit more replies.
  - ***H2***: Self-disclosure of either age or gender will elicit more replies marked as helpful.
  - ***H3***: Self-disclosure of both age *and* gender will elicit more replies as well as more replies marked as helpful.
- ***RQ3***: How does self-disclosure in advice-seeking posts relate to the quality of the replies?
  - ***H4***: Self-disclosure of age or gender from the advice seeker will elicit more self-disclosure from the advice giver.
  - ***H5***: Self-disclosure of age or gender will elicit advice preferentially from those of similar age or gender.

Data and Methods

To answer the research questions, we apply a mixed-method approach including quantitative hurdle negative binomial regression analysis (*RQ2* and *RQ3*) and qualitative discourse analysis (*RQ1* and *RQ3*). In this section, we will introduce the data source, specify the measurement of variables in regression analysis, and discuss regression models and discourse analysis steps.

Data

Reddit is a popular online forum where posts are organized into a variety of communities called "subreddits". This study focuses on a popular advice-seeking subreddit "r/Advice" which was created in 2008 and has over 680k members as of June 2022. Specifically, we use the Python Pushshift.io API Wrapper to collect posts published in r/Advice in 2021 and their associated replies, and use the Python Reddit API Wrapper to get updated data and user features. We dropped duplicated, removed, and deleted posts, and posts made by deleted, suspended, or banned accounts; we also dropped deleted replies under the remaining posts. This finally provided us with 135,398 posts and 853,976 attached replies. We performed preprocessing including expanding the contractions (e.g., "I'm" to "I am") and stemming (e.g., "disclose" to "disclos").

Measures

Self-Disclosure Identification of self-disclosure is the basis of both quantitative and qualitative analysis. As mentioned previously, this study focuses on two main categories of self-disclosure: age and gender, due to their popularity and importance on CQA sites. We first randomly sampled and coded 100 posts and 630 replies (according to the ratio of posts to replies in the whole dataset) to explore the common patterns of age and gender disclosure. Based on the exploratory results and prior studies (Umar, Squicciarini, and Rajtmajer 2019; Jagfeld et al. 2021), we extracted two patterns of age/gender disclosure using the following rule-based-matching approaches.

- **General Age and Gender Disclosure**: This refers to the usual ways that users disclose themselves in online communication. We modified the method in (Umar, Squicciarini, and Rajtmajer 2019) to identify general age and gender disclosure:
  - Part-of-Speech (POS) Tagging: We first used a pre-trained POS tagging model [1] to obtain the POS tag (e.g., noun, verb) for each token.
  - Named Entity Recognition (NER): Both age and gender disclosure involve named entities. For age disclosure, we used an existing NER model [2] to detect DATE (absolute or relative dates or periods) and CARDINAL (numerals that do not fall under another type) entities. For gender disclosure, we extracted the terms of binary genders (e.g., male, female) (Emmery, Chrupała, and Daelemans 2017) and non-binary genders (e.g., genderfluid) (Blake, Godwin, and Whyte 2020).
  - Rule-Based Matching: Our self-disclosure detection is based on the presence of corresponding named entities and the POS tags of these entities and adjacent tokens. Specifically, we first extracted first-person sub-sentences including age or gender entities, and then limited the tokens between the subject and entity to several POS types (e.g., adverb, determiner, adjective, and noun) to reduce false positives. We also identified self-disclosure sentences that start with age or gender entity without a subject, considering the popular informal expression in online communities.
- **Reddit-Specific Age and Gender Disclosure**: In addition to general self-disclosure, Redditors also develop a simple and convenient way to share their age and gender information, that is, reporting their age and gender

[1] https://spacy.io/api/tagger
[2] https://spacy.io/api/entityrecognizer

together in brackets (Jagfeld et al. 2021) (e.g., *“ I (22m) work from home...”*, *“I (21 nb) need advice...”*). This pattern is widely accepted and used by members in r/Advice. We used regular expression matching to extract it.

To verify our extraction of self-disclosure, we randomly sampled another 100 posts and 630 replies and manually coded them in terms of age and gender disclosure. For age disclosure, our approach achieved a macro-F1 score of 98.8% for the posts and 95.4% for the replies; for gender disclosure, the macro-F1 scores are 97.5% for the posts and 94.4% for the replies.

Replies to Posts We measured the dependent variable, the replies to posts, in multiple dimensions. We counted the number of total replies (*RQ2–H1*, *H3*) each post received. For the number of helpful replies (*RQ2–H2*, *H3*), our calculation used a feature of r/Advice. As Figure 1 shows, in r/Advice, the original poster can show their satisfaction by replying “helped” to the replies they received and valued, and then the moderator bot will confirm it and the original post will be marked as “advice received”. Therefore, we utilized moderators’ confirmation messages to identify “helpful replies” and counted the number of such replies received by each post. We also counted the number of replies with age or gender disclosure (*RQ3*) based on the results of self-disclosure identification.

Figure 1: Structure of threads in the r/Advice community. Some elements are blurred to protect privacy.

Control Variables As discussed in related work, prior research has identified a set of factors that can affect replies or helpful replies that a post receive. Therefore, we include these factors as control variables in our regression analysis.

- **User Features**: Users’ experience and reputation affect not only others’ replies to their posts (Movshovitz-Attias et al. 2013; Liu and Jansen 2018) but also their evaluation of these replies (Peng et al. 2021). Thus, we consider the following user features: 1) tenure: the length of time a user has stayed on Reddit, 2) post / reply karma: the total scores a user obtained through history posts / replies, and 3) helper rank: a user’s ranks in the flair system of r/Advice (e.g., 1: helper, 5: super helper).
- **Readability** Readability is an important factor in online communication (Risch and Krestel 2020). Following prior work (Park et al. 2017), we generated a composite readability score by averaging two widely used readability indices—the Simple Measure Of Gobbledygook (SMOG) index and the Coleman and Liau index. Both of these assess the average grade level a reader requires to understand a text. We also incorporated the text length (lexicon count) since it shows the linguistic complexity of a text.
- **Psychological Processes** The psycholinguistic features of a question are shown to affect whether it will be answered (Maity, Kharb, and Mukherjee 2017). Thus, we adopted a well-validated psycholinguistic lexicon LIWC (Linguistic Inquiry and Word Count) (Pennebaker et al. 2015), which captures the involvement of psychological processes in texts through multiple dictionary categories. We considered the LIWC categories including positive emotion, negative emotion, cognitive processes, social processes, perceptual processes, and biological processes.
- **Throwaway Account** The use of throwaway accounts is viewed as a good proxy for anonymity (Pavalanathan and De Choudhury 2015), and therefore may affect others’ feedback to online posts. Following prior research on Reddit (Ammari, Schoenebeck, and Romero 2019), we identified both throwaway usernames (e.g., *thrw*, *throwaway*, *throw*, *thraway*) and throwaway statements (e.g., “This is a throwaway account”).
- **Topic** As the necessity of self-disclosure for different information needs may be different, we adopted the Latent Dirichlet Allocation (LDA) model to discover the topics of posts. We built multiple LDA models with different numbers of topics and chose the one with 6 topics which has the highest coherence score. We checked the keywords that contribute the highest weights to each topic and labeled the topics (shown in Table 1). As the sum of the six topic variables is constant (equals one), we took the sixth topic “daily life” as our reference category and only fed the other five topics into the regression models.

Topic Name	Topic Words
Relationship	girl, ask, relationship, date, love
School	school, colleg, class, studi, life
Family	live, famili, hous, parent, home
Work	money, compani, manag, week, month
Mental health	mental, depress, anxieti, love, problem
Daily life	sleep, come, night, drink, get

Table 1: Extracted LDA topics and their top five representative stems.

Hurdle Negative Binomial Regression Models

The hurdle negative binomial regression model predicts outcome variables using two separate parts (Welsh et al. 1996). The first part uses a logit regression to predict the probability of zeros, and the second part applies a negative binomial regression to predict non-zero values. The split of two parts makes the model suitable for dealing with the excess of zeros; so, it is widely adopted in empirical studies (Arens et al. 2014; Hofstetter et al. 2016).

In this study, our dependent variables are over-dispersed (number of replies: mean=4.52, std=17.32; number of helpful replies: mean=0.22, std=1.27; number of age disclosure replies: mean=0.05, std=1.23; number of gender disclosure replies: mean=0.03, std=0.36), so the negative binomial regression is a good fit for them. Additionally, these dependent variables have excessive zeros because advice-seeking posts might not receive any reply of a specific type; so, we performed hurdle negative binomial regression to deal with the excessive zeros. We checked the correlation table (Figure 2) and calculated the variance inflation factors (VIF) scores of all predictors. As none of the VIF scores is greater than five, multicollinearity is not a problem in this study.

For each dependent variable, we built a set of nested models to unfold the relationship. This includes a parsimonious model that only incorporates control variables, models that add age or gender disclosure separately, and a full model that incorporates both age and gender and an interaction term between them.

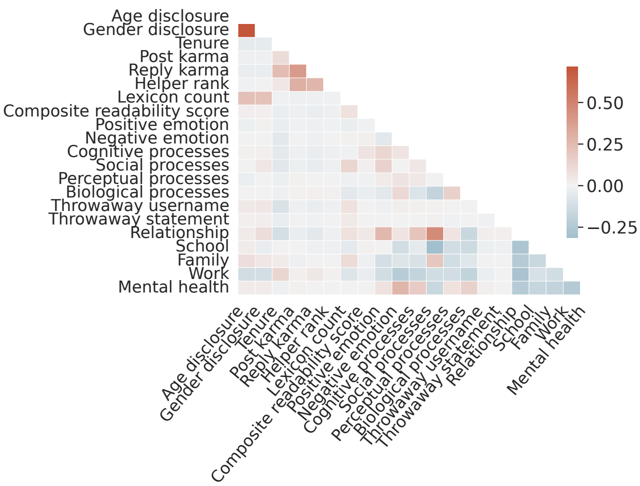

Figure 2: Correlation of independent variables and control variables.

Discourse Analysis

Discourse analysis is an approach focusing on "how social relations, identities, knowledge, and power are constructed in texts" (Crowe 2005). It examines the "language beyond the sentence" and relates it to its social context (Widdowson 1995). In this study, discourse analysis enables us to understand how self-disclosure is situated in the whole conversational thread. Due to the large data size, we randomly sampled 100 posts including age disclosure and 100 posts including gender disclosure as the basis of the discourse analysis. We collected all the replies associated with these posts. This finally provided us with 200 discussion threads including 200 posts and 1,780 replies. To answer *RQ1* and *RQ3*, we paid attention to both advice-seekers' motivations for self-disclosure and the effects of that self-disclosure on advice-providing replies.

The lead author read the posts and replies to obtain an impression and thematically assigned initial codes. In the second round of coding, we employed discourse analysis to understand how users manipulate "language and ways of acting, interacting, thinking, believing, valuing, and feeling to enact particular social identities and engage in social activities" (Gee 2010). In other words, our coding considered not only the content of the posts and replies, but also how they were situated in the whole conversational thread and related to others' utterances. In this way, we analyzed why original posters chose to disclose themselves to seek advice, and how replies were affected by original posters' self-disclosure. We iteratively engaged in coding and interpretation to ensure coherency.

Results

Descriptive Statistics of Posting and Self-Disclosure

Of all 135,398 posts, 33,052 (24.41%) posts include age disclosure and 25,719 (19.00%) posts include gender disclosure; of all 853,976 replies, 8,592 (1.01%) replies include age disclosure and 4,599 (0.54%) replies include gender disclosure. Table 2 shows the frequency of posts and replies categorized by disclosure of age, gender, both, or neither. We further classified age disclosure into five groups (<18, 18−24, 25−44, 45−59, and ≥60) (Hannestad et al. 2000), and classified gender disclosure into three groups (male, female, and non-binary).

	Non-gender	Gender	Total
Non-age	99,172	3,174	102,346
Age	10,507	22,545	33,052
Total	109,679	25,719	135,398

(a) Post

	Non-gender	Gender	Total
Non-age	842,742	2,642	845,384
Age	6,635	1,957	8,592
Total	849,377	4,599	853,976

(b) Reply

Table 2: Cross table of age disclosure and gender disclosure.

Predicting Reply Frequency and Self-Disclosure

To address *RQ2* and *RQ3*, we utilize hurdle negative binomial regression models to investigate the statistical relationship between self-disclosure and both the quantity and quality of replies. These hurdle models disaggregate into two models, one that assesses the odds of the outcome or not (the zero hurdle model) and one that assesses the count of the outcome (the count model).

The Relationship Between Self-Disclosure and the Number of (Helpful) Replies Received To answer *RQ2*, we analyze how advice-seekers' self-disclosure is associated with the number of replies and the number of helpful replies. The results are shown in Table 3.

Any Reply Model 1.2 and Model 1.3 (Table 3) illustrate the main effect of age and gender disclosure on received replies. Holding control variables constant, adding age disclosure (coeff.=0.32) or gender disclosure (coeff.=0.28)

		Baseline (controls only)		+ Age Disclosure		+ Gender Disclosure		+ Both Disclosure	
		Model 1.1		Model 1.2		Model 1.3		Model 1.4	
		Zero hurdle	Count	Zero hurdle	Count	Zero hurdle	Count	Zero hurdle	Count
All replies (*RQ2*)	Age disclosure			0.32***	0.28***			0.33***	0.36***
	Gender disclosure					0.28***	0.21***	0.24***	0.24***
	Age × Gender							-0.24***	-0.35***
	Intercept	2.13***	0.32***	2.06***	0.28***	2.08***	0.29***	2.06***	0.27***
	AIC	669,114		668,376		668,739		668,292	
	BIC	669,517		668,798		669,161		668,753	
		Model 2.1		Model 2.2		Model 2.3		Model 2.4	
		Zero hurdle	Count	Zero hurdle	Count	Zero hurdle	Count	Zero hurdle	Count
Helpful replies (*RQ2*)	Age disclosure			0.34***	0.27***			0.07*	0.24***
	Gender disclosure					0.45***	0.29***	0.35***	0.38***
	Age × Gender							0.05	-0.31**
	Intercept	-1.98***	-10.11	-2.07***	-10.6	-2.07***	-11.05	-2.08***	-10.67
	AIC	139,053		138,690		138,520		138,503	
	BIC	139,456		139,112		138,943		138,965	

Table 3: Hurdle negative binomial regression coefficients of advice-seekers' self-disclosure for the number of all replies and helpful replies received (*RQ2*). $^{*}p < 0.050$, $^{**}p < 0.010$, $^{***}p < 0.001$. N = 135,398. The reference topic for LDA topic variables is "daily life". AIC and BIC refer to the Akaike information criterion and Bayesian information criterion respectively.

separately correlates with a significantly higher probability of receiving at least one reply. This implies that self-disclosure may be helpful to avoid posts receiving no replies—a common problem in online advice seeking. Additionally, including age disclosure (coeff.=0.28) or gender disclosure (coeff.=0.21) also predicts a significantly larger number of replies. Model 1.4 considers both age and gender disclosure, and adds an interaction term. The results show that the main effects of age disclosure and gender disclosure on the probability (age coeff.=0.33; gender coeff.=0.24) and count (age coeff.=0.36; gender coeff.=0.24) are still significant, indicating that either type of disclosure increases both the probability and quantity of replies. However, the interaction term age×gender predicts a significantly lower probability of receiving any replies (coeff.=–0.24) and less replies (coeff.=–0.35). This suggests the limited marginal benefit of self-disclosure: that is the benefit of disclosing gender *and* age is less than the sum of disclosing each separately, but the overall effect of disclosure is still positive.

Helpful Replies Receiving replies is not the end of an advice-seeking process. We also pay attention to whether a problem is finally marked as solved, and if so, the number of helpful replies. As Table 3 shows, Model 2.2 and Model 2.3 add age and gender disclosure to the baseline model respectively. The results demonstrate that advice-seekers' age disclosure or gender disclosure significantly increases the probability of receiving a helpful reply (age coeff.=0.34; gender coeff.=0.45) and the number of replies (age coeff.=0.27; gender coeff.=0.29). When considering both disclosure types and adding the interaction term age×gender (Model 2.4), the main effects of both types of disclosure on the probability (age coeff.=0.07; gender coeff.=0.35) and the count (age coeff.=0.24; gender coeff.=0.38) remain significant. That is, age and gender disclosure predict a higher probability that the original post would be marked as "advice received" and a larger number of helpful replies. The interaction term age×gender is significant in the count model (coeff.=–0.31), revealing that the effect of disclosing age and gender is less than the sum of the effects of disclosing each separately.

The Relationship Between Self-Disclosure in Posts and Self-Disclosure in Replies To address *RQ3*, we examine the reciprocity of self-disclosure. As Table 4 shows, we build regression models that utilize advice-seekers' age/gender disclosure in posts to predict the number of age/gender-disclosure replies they would receive. We further explore the content of reciprocal disclosure.

Age Disclosure in Replies According to Model 3.2 and Model 3.3 (Table 4), when only one type of disclosure is considered, advice-seekers' age disclosure (coeff.=1.22) or gender disclosure (coeff.=0.81) is positively associated with the probability of receiving an age-disclosure reply. In Model 3.4, which considers both types of disclosure and their interaction effect, both age (coeff.=1.39) and gender disclosure (coeff.=0.43) still predict a higher probability of receiving any age-disclosure reply. The effect size of age disclosure is larger than that of gender disclosure. These results suggest the reciprocal effects of self-disclosure. Moreover, the significance in the interaction term age×gender (coeff.=–0.65) indicates that the effects of age disclosure and gender disclosure on the probability of receiving an age-disclosure reply are not additive.

Gender Disclosure in Replies The results regarding gender-disclosure replies also show the reciprocity of self-disclosure. Model 4.2 and Model 4.3 (Table 4) exhibit that

		Baseline (controls only)		+ Age Disclosure		+ Gender Disclosure		+ Both Disclosure	
		Model 3.1		Model 3.2		Model 3.3		Model 3.4	
		Zero hurdle	Count	Zero hurdle	Count	Zero hurdle	Count	Zero hurdle	Count
Age-disclosure replies (*RQ3*)	Age disclosure			1.22***	0.01			1.39***	-0.08
	Gender disclosure					0.81***	0.03	0.43***	-0.37
	Age × Gender							-0.65***	0.48
	Intercept	-3.68***	-15.63	-4.11***	-15.61	-3.88***	-14.44	-4.13***	-12.87
	AIC	40,380		39,161		39,903		39,132	
	BIC	40,782		39,583		40,325		39,593	
		Model 4.1		Model 4.2		Model 4.3		Model 4.4	
		Zero hurdle	Count	Zero hurdle	Count	Zero hurdle	Count	Zero hurdle	Count
Gender-disclosure replies (*RQ3*)	Age disclosure			1.01***	0.30**			0.78***	0.31
	Gender disclosure					1.17***	0.21*	1.43***	0.08
	Age × Gender							-0.95***	-0.07
	Intercept	-4.14***	-11.69	-4.48***	-18.75	-4.47***	-19.52	-4.57***	-11.89
	AIC	27,995		27,446		27,297		27,199	
	BIC	28,397		27,868		27,719		27,661	

Table 4: Hurdle negative binomial regression coefficients of advice-seekers' self-disclosure for the number of age/gender-disclosure replies received (*RQ3*). $^{*}p < 0.050$, $^{**}p < 0.010$, $^{***}p < 0.001$. N = 135,398. The reference topic for LDA topic variables is "daily life". AIC and BIC refer to the Akaike information criterion and Bayesian information criterion respectively.

including age or gender disclosure in advice-seeking posts is separately associated with a higher probability of receiving any gender-disclosure reply (age coeff.=1.01; gender coeff.=1.17) and a larger number of gender-disclosure replies (age coeff.=0.30; gender coeff.=0.21). Model 4.4, which incorporates both disclosure types and adds their interaction term, suggests that the main effects of age disclosure (coeff.=0.78) and gender disclosure (coeff.=1.43) remain significant. Nevertheless, the negative effect of the interaction term (coeff.=–0.95) indicates that the effects of advice-seekers' age disclosure and gender disclosure on the probability of receiving gender-disclosure replies are also non-additive, consistent with the relationship observed for age-disclosure replies. The results of the count model do not show significant effect of age or gender disclosure, or their interaction, on the number of replies received.

Content of Reciprocal Disclosure To better understand the propensity to disclose in a reply we examine reciprocal self-disclosure. First, we selected all "post–reply" pairs where both the post and the attached reply contain the same type of disclosure (age or gender). This provided us with 3,720 age disclosure pairs and 1,826 gender disclosure pairs. Then we looked at the content of age/gender disclosure in these disclosure pairs.

Regarding age disclosure (shown in Table 5), the most common cells are usually on the diagonal. The notable exception is among the youngest group (disclosing being under 18) which received slightly more replies among the 18-24 group than within their group. The difference is modest and may reflect simply a regression to the mean of age on Reddit. That said, given Reddit's age skew, there was limited evidence for posts from people older than 45 generally.

Post \ Reply	**<18**	**18−24**	**25−44**	**45−59**	**≥60**	**Sum**
<18	531	562	372	26	17	1,508
18−24	146	915	554	46	29	1,690
25−44	41	100	306	10	24	481
45−59	1	3	3	2	1	10
≥60	4	6	7	9	5	31

Table 5: Distributions of age disclosure in replies of the reciprocal "post-reply" age-disclosure pairs.

In terms of gender, we also found the homophily phenomenon. As Table 6 illustrates, for posts in any gender group (female, male, or non-binary), more than half of the replies in the gender-disclosure pairs come from the same gender group.

Post \ Reply	**Female**	**Male**	**Non-binary**	**Sum**
Female	528	484	14	1,026
Male	336	401	17	754
Non-binary	8	12	26	46

Table 6: Distributions of gender disclosure in replies of the reciprocal "post-reply" gender-disclosure pairs.

Interpreting Self-Disclosure in Posts and Replies

In addition to statistical analysis, we conduct discourse analysis to gain a deeper understanding of the content of posts and replies. The results enable us to comprehend seekers' expressed reasons for self-disclosure (*RQ1*) and how that disclosure shapes the content of replies (*RQ3*). We explain the results using representative examples, which are paraphrased slightly to avoid full text search.

Understanding the Use of Self-Disclosure among Advice-Seekers Below we identify four themes indicating how advice seekers strategically deploy personal information in relation to the content of the reply and the advice sought (*RQ1*).

Following Social Norms Personal information can be provided as a sort of convention. That is, it is not highlighted as a key part of the story. Rather, it is appended or included in an almost incidental way. For instance, a user asked for advice on taxes: *"So I (37m) rent my house through a third party but they did not pay taxes... Any advice?"*. Another example is a post seeking advice on managing emotions: *"How can I separate emotions and being a host of myself at work? (26 y/o, F)"*. In these examples, there was not an obvious reason for the users to include age or gender information in the advice-seeking process, but they still mentioned them in a well-accepted way on r/Advice. Other users remark on this convention explicitly. For example, one user clearly expressed their willingness to obey the rules and disclose themselves in an appropriate way: *"This is my first post here so I hope I'm following the rules. I'm 23, and my best friend is 22..."*

Providing Background Information Advice seekers will sometimes deploy personal information as a sort of scene setting to provide a context of themselves as they tell the story of why they are seeking advice. These posts often placed the age or gender information where it could be easily seen (e.g., at the beginning or the end of the post), together with other personal information, if any. For example, a user who wanted advice on making better use of time used self-disclosure as the start of the post: *"My situation: I'm a 23-year-old female and work a full-time job..."* Similarly, a user began with self-disclosure to ask for relationship advice: *"Let me give you the backstory: I (45M) was married to my ex-wife (43F) for 22 years..."* These users considered it relevant to their context without explaining precisely how and leaving that to the reader. For example, a user wrote: *"Ok, this is a complex situation and some context may be helpful. I'm 17 and live in..."* This use of self-disclosure differs from "following social norms" in that users disclosed their gender or age as part of the background information in a more explicit and conscious way.

Increasing Credibility In some cases, the information provided is not done as scene setting in an indirect way but is meant to highlight a specific personal challenge or constraint related to the advice and help increase the credibility of advice-seekers' needs. For instance, a user who sought relationship advice mentioned: *"I'm 22 and I've never had a best friend like him before, so I feel very excited about it."* In this example, the user's age is a reason for the user's current feelings and needs, so her age disclosure makes the story flow logically and look reasonable. Similarly, a user explained her worry about a friend using age-disclosure: *"I can't let him move in with me because I am only 19 and live with my parents."* The user mentioned her age to illustrate why some easy solutions are not practical and made others more willing to continue reading. Another user disclosed her age to explain the probable difficulty of her plan: *"I'm already 22(F). I want to go back to school...By 27, I should be at a good place...Does this seem so out of my reach?"* In these examples, users disclosed themselves to explain their thoughts or behaviors which might be confusing otherwise. As a result, the disclosure helped foster understanding to hopefully elicit more trustworthiness and appropriate advice.

Seeking Targeted Advice Whereas increasing credibility meant a seeker would want the audience to understand their situation, sometimes the seeker wants targeted advice that considers their personal information. It is not that their age or gender necessarily provides the constraint, but instead it provides a basis for establishing shared experience among others who may have had the same experience. For example, a user indicated his needs with age disclosure: *"How can I [21] act younger and more like my peers."* Here it is not merely to contextualise their situation but explicitly seek others with such experience. In another instance, a user showed a preference for advice from similar people when seeking guidance on coping with anxiety: *"I'm a 22-year-old female looking for anyone who has a similar experience that's positive to make me feel not alone and to remind me that everything will be okay."*. In summary, self-disclosure can be used as a tool to clarify desired advice and filter out potential useless or unhelpful replies.

Understanding the Use of Self-Disclosure among Reply-Givers Many reply-givers include personal information from the original advice-seeker and some of them also include personal information of their own. Below we identified three approaches to replying to advice-seekers' posts using their personal information (*RQ3*).

Basis of Reasoning Advice-seekers' self-disclosure can help reply-givers interpret their situation and reason. Sometimes reply-givers are not able to provide specific advice, but advice-seekers' self-disclosure still helps them understand the problem and make progress in supporting advice-seekers through reasoning. For instance, a user asked for suggestions on moving to another area, and a reply-giver wrote: *"Because you're only 16 years old, you may be a target for child trafficking and sexual assault."* In this example, the reply-giver noticed the age disclosure in the original post and mentioned it again to stress some points. Reply-givers can also go further and provide targeted advice after reasoning. Under a post wondering how to deal with anxiety after making mistakes, a reply-giver offered advice: *"You're only 27 and you have many many years to live. This thing will be a distant memory soon so you don't need to make such a permanent decision."* This reply-giver referred to the original poster's self-disclosure and provided support accordingly. Moreover, a user took into account the advice-seeker's self-disclosure not only in the reasoning process, but also in the way of expression: *"You seem young so I'll be gentle..."*

Enabling Self-Evaluation of Helpfulness Advice-seekers' self-disclosure can also serve as a reference for reply-givers to evaluate the usefulness and feasibility of their advice. On the one hand, some reply-givers are confident in their replies because of homophily or authority. Homophily means the similarity between the reply-giver and the original poster. For example, a reply-giver replied to the original poster with great certainty because of their similarity in age: *"Choosing whether or not to get cochlear implants is a personal choice! I also am young and had hearing loss at age 21...I COMPLETELY understand your intentions..."* Similarly, a reply-giver started with gender disclosure to support his advice under a gender-disclosure post: *"Hi! Trans dude here. It's great that you are exploring your identity...That takes a lot of courage and work..."* As for authority, it may happen when the reply-giver is older or more experienced than the original poster. For instance, a reply-giver believed his advice to a 17-year-old poster is convincing because he had experienced the illness before: *"I'm 41f. Although I'm very healthy and happy now, I have been through some severe mental illness in the past...so I'm in a good position to give you some advice."* The authority is also reflected when the reply-giver is in a different position whose opinions the original poster does not know but cares about. For a man who needed advice on making an appealing bio to find a girlfriend, a reply-giver knew her opinions mattered because of gender and replied with gender disclosure: *"Woman here. Including your interests and personality are the big ones. A blank bio won't attract many matches..."*

On the other hand, some reply-givers disclose to acknowledge the potential bias of their advice. For example, a user identified how gender may bias his opinion in the following example replying to a post about body scent: *"I'd believe that would come down to poor hygiene...But like you said, I'm not a woman."* In another example, a user replied to a post about a relationship in a similar way: *"I personally wouldn't want my boyfriend to hang out with a girl he was hooking up...But that's just one gal's opinion."*

Increasing Empathy Self-disclosure offers personal information about advice-seekers that allows reply-givers to better understand their feelings, and those reply-givers may in turn communicate that understanding in replies. In other words, reply-givers may express empathy in response to advice-seekers' self-disclosure, e.g., *"I feel you...I would suggest..."*, *"I was in the same sort of situation...I realized..."* Sometimes reply-givers also disclose themselves to better show their empathy. For instance, a 31-year-old male asked for life advice, and a user replied with empathic words: *"I am in the same boat, just wrapping up nursing school, 34 years old and it's really bumming me out thinking of working until die..."* Under a post regarding driving anxiety, a reply-giver at a comparable age shared similar feelings to make the poster feel not alone: *"I'm 22 and I'm scared of being in any vehicle but driving myself is even worse..."* Reciprocal self-disclosure here serves as a way to express reply-givers' understanding through similar experiences, which is different from the other effect (i.e., "enabling self-evaluation of helpfulness") where reciprocal self-disclosure was used to show reply-givers' evaluation of the usefulness and feasibility of their advice.

Discussion

This study examines how self-disclosure is used to support requests in an advice community. Speaking to *RQ1* on how disclosure relates to the nature and content of the request, our results show that self-disclosure can be used to provide context for requests (Pfeil and Zaphiris 2007) and enhance credibility (Laurenceau, Barrett, and Pietromonaco 1998; Galegher, Sproull, and Kiesler 1998), which is aligned with previous research. Moreover, the results add to existing knowledge by revealing that self-disclosure can serve the additional purpose of identifying individuals with similar experiences in the community and soliciting targeted advice, or less intentionally, simply conforming to community norms.

We further explored the topic by looking at both the presence (*RQ2*) and quality (*RQ3*) of replies to requests that varied in their self-disclosure. Our hypotheses generally assumed that more disclosure would lead to positive outcomes (more and higher quality replies, more self-disclosure, etc). We separated out these hypotheses by type of outcome, but all relate to the presence (or combination) of advice seekers' age and gender disclosure.

Hypothesis 1 asked whether these test variables would lead to more replies. We explore this while accounting for previously established factors such as topic (Park et al. 2020) and readability (Risch and Krestel 2020). Our quantitative results support this hypothesis as age and gender are associated with a greater likelihood of any reply (with significant Zero hurdle coefficients) as well as greater numbers of replies (with significant Count coefficients).

Replies vary a lot in quality (Welser et al. 2007) and the helpfulness of replies is also important to advice-seekers on CQA sites (Liu et al. 2014). *Hypothesis 2* went further to ask whether the replies offered were more likely to be considered helpful. That is, self-disclosure might simply be a form of signalling for interest or it might be a means to establish a context that can improve the quality of the results. In this case, we opted to employ Reddit's convention of advice-seekers labelling posts as 'helpful'. This limits the generalisability of this work since it renders helpfulness as binary. However, it does allow us to consider a large pool of naturally occurring labels of helpfulness which are difficult to replicate in controlled experiments. Similar to *Hypothesis 1*, our results support *Hypothesis 2* as both age and gender show significant independent effects on the probability and the count of replies labelled as helpful.

Previous research has established the reciprocity of general self-disclosure (Barak and Gluck-Ofri 2007). This study further enhances our understanding of reciprocity by examining specific types of self-disclosure, thereby providing support for *Hypothesis 4*. We find that advice-seeking posts with age or gender disclosure were significantly more likely to receive a reply with age or gender disclosure. The coefficient for age disclosure in a reply is largest for the age-only model (assuming the same controls for all models). The same pattern holds for gender.

Before turning to how qualitative analysis may inform these effects, it is worth considering that we do not find support for *Hypothesis 3*. While we hypothesized that both age and gender will have a positive effect on replies together, we were surprised to find that the interaction terms overshadowed any additive effects. In virtually all the full (*.4) models, the interaction parameter is significant and negative. That is to say, disclosing age or gender helps, but disclosing both does not increase the outcome and may come at a cost.

The overall picture suggests that self-disclosure thus appears to have positive effects, such as alleviating the common "no response" problem in community question answering (CQA) sites (Li and King 2010), eliciting reciprocity of replies, and increasing user engagement, which can further benefit the development of online communities (Kayes et al. 2015). If this is the case, it suggests that there are rationales for why individuals might feel encouraged to disclose information as well as encouraged to withhold information. Many of these rationales were introduced in the qualitative results section.

How can disclosure be useful or not in eliciting advice? The qualitative results augmented these model interpretations by identifying some of the reasoning that posters bring to bear on whether and why to self-disclose.

We first were able to establish that there is a clear normative basis to self-disclosure. Both advice-seekers and reply-givers remark on their disclosure as well as consider it common within this forum. We find that advice-seekers' often have deliberate expectations; they disclose details of themselves as a reference for others or to request targeted advice suitable for their case. Reply-givers tend to be more helpful when taking that into account, aligned with prior research indicating how self-disclosure can bring about positive social reactions (Vitak and Kim 2014; Li-Barber 2012).

In line with previous research that demonstrates the impact of self-disclosure on eliciting person-centeredness (Pan et al. 2020) and empathy (Pfeil and Zaphiris 2007), this study shows that reply-givers may utilize advice-seekers' self-disclosure to reason and interpret their needs, and express empathy in response. Furthermore, our results suggest an unexplored function of self-disclosure: advice-seekers' self-disclosure gives others a clear sense of how convincing or feasible their advice could be, and reply-givers may reflect that self-evaluation in replies (i.e. "my reply is good because it comes from a specific, generally similar, position"). To the best of our knowledge, the mechanism of "enabling self-evaluation" has not been explored in prior self-disclosure studies.

However, there may still be drawbacks to disclosure. Our quantitative results suggest the non-additive effects of self-disclosure, demonstrating that more self-disclosure does not necessarily lead to improved outcomes. Previous research identified such drawbacks as social rejection (Vitak and Kim 2014) and stereotyping (Han, Huang, and Wang 2019).

How do reply-givers reciprocate self-disclosure? Our quantitative results indicate the reciprocity of self-disclosure. In Tables 5 and 6 we show the mixing matrix of ages and genders, highlighting how many more replies are on the diagonal (homophilous disclosure) to off diagonal (heterophilous disclosure). Thus, a little signal appears to be useful both garnering more replies and well as replies from more similar persons, in line with *Hypothesis 5*.

Our discourse analysis reveals that one reason evinced for reciprocal self-disclosure appears to be to bolster credibility through an assertion of authority or shared experience, which aligns with the trust theory for the reciprocity of self-disclosure (Wetzel and Wright-Buckley 1988). This shared experience can further signal empathy as reply-givers express their understanding and trust.

Arbitrating between trust-based theories and others, such as the modeling theory (Rubin 1975) and social exchange theory (Worthy, Gary, and Kahn 1969) may require additional data about participants or a different research design. This is because mimicry or the sense of obligation may occur unconsciously or by convention and not be explicitly expressed in the text.

Broader Perspective

All online communication involves the construction of an identity, however slight. In an era where profiles are often encouraged to share increasing amounts of personal information, it is important to understand how such information can be useful or not in eliciting advice from others. Rather than viewing this work as promoting self-disclosure, we view this as helping to optimise what is sufficient self-disclosure while appreciating the utility of spaces that allow for some anonymity or pseudonymity. This helps us understand how crowds can effectively dispense advice on highly personal topics in an open and safe manner. The hurdle models indicated that some self-disclosure appears to elicit more replies and more helpful replies. However, the effect of age or gender appears as useful as the two combined. Further work may want to explore the relative benefits of more or less information disclosure. There may be an optimal amount of disclosure to create empathy and validation without undermining the safety and openness of relatively anonymous CQA sites.

We focused on age and gender disclosure in a general advice-seeking community (r/Advice) given the consistency of their use and the simplicity of data capture. We believe this work can be fruitfully extended by considering shared disclosure of other sensitive topics in niche forums, such as the limited revelation of health issues in health forums or of personal preferences and identities in forums for LGBTQ+ support. This can be seen as a counterweight to the assertion that one can simply disclose any information in any context or use a single persistent identity when seeking information on a sensitive topic.

Future work may look at variations in identity disclosure across other platforms such as Twitter or Facebook where many more identity signals are already pre-given, as well as forums where people would have only forum-specific identities. Also, as we were focusing on signals in context, we did not look through past replies by reply-givers to establish their likely age or gender. This might help understand not only the propensity for an advice-seeker to get a reply, but the propensity of a reply-giver to give a reply and give specific details in return.

Finally, in order to consider the merit of this work in relation to our obligations to ethical research practice, we maintain a dehydrated version of our data with ids and scores but not content or specific age and gender. We kept the Reddit ids in the rows so that future researchers can work with the current permissible state of the original data while allowing individual Reddit users the assurance that deleting their data on Reddit would not leave residual identifying details in our dataset.

Conclusion

This study employs a mixed-method approach to examine age and gender disclosure in an online advice-seeking community. Our results show that advice-seeking posts with age or gender disclosure would be more likely to receive any (helpful) reply and receive more (helpful) replies, while including both types of disclosure together does not have additive benefits in terms of either probability or quantity. We also find that reciprocity exists in self-disclosure, and explore the reasons behind the reciprocal self-disclosure, e.g., reply-givers' intention to enhance credibility or communicate empathy. This study provides insights into the impact of the level of self-disclosure on soliciting advice. Our findings also have implications for the development of CQA sites in all stages of asking for advice, getting replies, and evaluating those replies.

Acknowledgments

This work was supported by the Economic and Social Research Council [grant number ES/P000649/1].

References

Ammari, T.; Schoenebeck, S.; and Romero, D. 2019. Self-declared throwaway accounts on Reddit: How platform affordances and shared norms enable parenting disclosure and support. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW): 1–30.

Andalibi, N.; Haimson, O. L.; Choudhury, M. D.; and Forte, A. 2018. Social support, reciprocity, and anonymity in responses to sexual abuse disclosures on social media. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 25(5): 1–35.

Anderson, A.; Huttenlocher, D.; Kleinberg, J.; and Leskovec, J. 2012. Discovering value from community activity on focused question answering sites: a case study of stack overflow. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 850–858.

Arens, A. M.; Gaher, R. M.; Simons, J. S.; and Dvorak, R. D. 2014. Child maltreatment and deliberate self-harm: A negative binomial hurdle model for explanatory constructs. *Child maltreatment*, 19(3-4): 168–177.

Balani, S.; and De Choudhury, M. 2015. Detecting and characterizing mental health related self-disclosure in social media. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, 1373–1378.

Barak, A.; and Gluck-Ofri, O. 2007. Degree and reciprocity of self-disclosure in online forums. *CyberPsychology & Behavior*, 10(3): 407–417.

Birnholtz, J.; Merola, N. A. R.; and Paul, A. 2015. " 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*, 2613–2622.

Blake, K.; Godwin, M.; and Whyte, S. 2020. "I sexually identify as an Attack Helicopter": Incels, trolls, and non-binary gender politics online. *First Monday*.

Crowe, M. 2005. Discourse analysis: towards an understanding of its place in nursing. *Journal of advanced nursing*, 51(1): 55–63.

El Adlouni, Y.; Rodríguez, H.; Meknassi, M.; El Alaoui, S. O.; and En-nahnahi, N. 2019. A multi-approach to community question answering. *Expert Systems with Applications*, 137: 432–442.

Emmery, C.; Chrupała, G.; and Daelemans, W. 2017. Simple queries as distant labels for predicting gender on twitter. In *Proceedings of the 3rd Workshop on Noisy User-generated Text*, 50–55.

Fu, L.; Chang, J. P.; and Danescu-Niculescu-Mizil, C. 2019. Asking the right question: Inferring advice-seeking intentions from personal narratives. *arXiv preprint arXiv:1904.01587*.

Galegher, J.; Sproull, L.; and Kiesler, S. 1998. Legitimacy, authority, and community in electronic support groups. *Written communication*, 15(4): 493–530.

Gee, J. P. 2010. *How to do discourse analysis: A toolkit: A toolkit*. Routledge.

Han, X.; Huang, H.; and Wang, L. 2019. F-PAD: Private attribute disclosure risk estimation in online social networks. *IEEE Transactions on Dependable and Secure Computing*, 16(6): 1054–1069.

Hannestad, Y. S.; Rortveit, G.; Sandvik, H.; and Hunskaar, S. 2000. A community-based epidemiological survey of female urinary incontinence:: The Norwegian EPINCONT Study. *Journal of clinical epidemiology*, 53(11): 1150–1157.

Hanson, J. 2005. Should your lips be zipped? How therapist self-disclosure and non-disclosure affects clients. *Counselling and Psychotherapy Research*, 5(2): 96–104.

Hofstetter, H.; Dusseldorp, E.; Zeileis, A.; and Schuller, A. A. 2016. Modeling caries experience: advantages of the use of the hurdle model. *Caries research*, 50(6): 517–526.

Jagfeld, G.; Lobban, F.; Rayson, P.; and Jones, S. H. 2021. Understanding who uses Reddit: Profiling individuals with a self-reported bipolar disorder diagnosis. *NAACL HLT 2021*, 1.

Jourard, S. M.; and Lasakow, P. 1958. Some factors in self-disclosure. *The Journal of Abnormal and Social Psychology*, 56(1): 91.

Kayes, I.; Kourtellis, N.; Quercia, D.; Iamnitchi, A.; and Bonchi, F. 2015. The social world of content abusers in community question answering. In *Proceedings of the 24th international conference on world wide web*, 570–580.

Lankton, N. K.; McKnight, D. H.; and Tripp, J. F. 2017. Facebook privacy management strategies: A cluster analysis of user privacy behaviors. *Computers in Human Behavior*, 76: 149–163.

Laurenceau, J.-P.; Barrett, L. F.; and Pietromonaco, P. R. 1998. Intimacy as an interpersonal process: the importance of self-disclosure, partner disclosure, and perceived partner responsiveness in interpersonal exchanges. *Journal of personality and social psychology*, 74(5): 1238.

Li, B.; and King, I. 2010. Routing questions to appropriate answerers in community question answering services. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, 1585–1588.

Li-Barber, K. T. 2012. Self-disclosure and student satisfaction with Facebook. *Computers in Human behavior*, 28(2): 624–630.

Liu, T.; Zhang, W.-N.; Cao, L.; and Zhang, Y. 2014. Question popularity analysis and prediction in community question answering services. *PloS one*, 9(5): e85236.

Liu, Z.; and Jansen, B. J. 2018. Questioner or question: Predicting the response rate in social question and answering on Sina Weibo. *Information Processing & Management*, 54(2): 159–174.

Maity, S.; Kharb, A.; and Mukherjee, A. 2017. Language use matters: Analysis of the linguistic structure of question texts can characterize answerability in quora. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 11, 612–615.

Movshovitz-Attias, D.; Movshovitz-Attias, Y.; Steenkiste, P.; and Faloutsos, C. 2013. Analysis of the reputation system and user contributions on a question answering website: Stackoverflow. In *2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013)*, 886–893. IEEE.

Pan, W.; Feng, B.; Wingate, V. S.; and Li, S. 2020. What to say when seeking support online: A comparison among different levels of self-disclosure. *Frontiers in psychology*, 11: 978.

Park, A.; Zhu, S.-H.; Conway, M.; et al. 2017. The readability of electronic cigarette health information and advice: a quantitative analysis of web-based information. *JMIR public health and surveillance*, 3(1): e6687.

Park, K.; Kwak, H.; Song, H.; and Cha, M. 2020. “Trust Me, I Have a Ph. D.”: A Propensity Score Analysis on the Halo Effect of Disclosing One’s Offline Social Status in Online Communities. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, 534–544.

Pavalanathan, U.; and De Choudhury, M. 2015. Identity management and mental health discourse in social media. In *Proceedings of the 24th international conference on world wide web*, 315–321.

Peng, Z.; Ma, X.; Yang, D.; Tsang, K. W.; and Guo, Q. 2021. Effects of Support-Seekers’ Community Knowledge on Their Expressed Satisfaction with the Received Comments in Mental Health Communities. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–12.

Pennebaker, J. W.; Boyd, R. L.; Jordan, K.; and Blackburn, K. 2015. The development and psychometric properties of LIWC2015. Technical report.

Pfeil, U.; and Zaphiris, P. 2007. Patterns of empathy in online communication. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 919–928.

Risch, J.; and Krestel, R. 2020. Top comment or flop comment? predicting and explaining user engagement in online news discussions. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, 579–589.

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

Sikdar, S.; Sachdeva, R.; Wachs, J.; Lemmerich, F.; and Strohmaier, M. 2021. The effects of gender signals and performance in online product reviews. *Frontiers in big Data*, 4.

Taylor, S. H.; DiFranzo, D.; Choi, Y. H.; Sannon, S.; and Bazarova, N. N. 2019. Accountability and empathy by design: Encouraging bystander intervention to cyberbullying on social media. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW): 1–26.

Tidwell, L. C.; and Walther, J. B. 2002. Computer-mediated communication effects on disclosure, impressions, and interpersonal evaluations: Getting to know one another a bit at a time. *Human communication research*, 28(3): 317–348.

Umar, P.; Squicciarini, A.; and Rajtmajer, S. 2019. Detection and analysis of self-disclosure in online news commentaries. In *The World Wide Web Conference*, 3272–3278.

Vitak, J.; and Kim, J. 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*, 461–474.

Wang, Y.-C.; Burke, M.; and Kraut, R. 2016. Modeling self-disclosure in social networking sites. In *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing*, 74–85.

Welser, H. T.; Gleave, E.; Fisher, D.; and Smith, M. 2007. Visualizing the signatures of social roles in online discussion groups. *Journal of social structure*, 8(2): 1–32.

Welsh, A. H.; Cunningham, R. B.; Donnelly, C. F.; and Lindenmayer, D. B. 1996. Modelling the abundance of rare species: statistical models for counts with extra zeros. *Ecological Modelling*, 88(1-3): 297–308.

Wetzel, C. G.; and Wright-Buckley, C. 1988. Reciprocity of self-disclosure: Breakdowns of trust in cross-racial dyads. *Basic and Applied Social Psychology*, 9(4): 277–288.

Widdowson, H. G. 1995. Discourse analysis: a critical view. *Language and literature*, 4(3): 157–172.

Worthy, M.; Gary, A. L.; and Kahn, G. M. 1969. Self-disclosure as an exchange process. *Journal of personality and social psychology*, 13(1): 59.

Paper Checklist

1. For most authors...

(a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes
(b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes
(c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes
(d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes
(e) Did you describe the limitations of your work? Yes
(f) Did you discuss any potential negative societal impacts of your work? Yes
(g) Did you discuss any potential misuse of your work? Yes
(h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes
(i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes

2. Additionally, if your study involves hypotheses testing...
(a) Did you clearly state the assumptions underlying all theoretical results? Yes
(b) Have you provided justifications for all theoretical results? Yes
(c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? Yes
(d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? Yes
(e) Did you address potential biases or limitations in your theoretical framework? Yes
(f) Have you related your theoretical results to the existing literature in social science? Yes
(g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? Yes

3. Additionally, if you are including theoretical proofs...
(a) Did you state the full set of assumptions of all theoretical results? NA
(b) Did you include complete proofs of all theoretical results? NA

4. Additionally, if you ran machine learning experiments...
(a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? NA
(b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? NA
(c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? NA
(d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? NA
(e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? NA
(f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? NA

5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
(a) If your work uses existing assets, did you cite the creators? NA
(b) Did you mention the license of the assets? Yes
(c) Did you include any new assets in the supplemental material or as a URL? Yes
(d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? Yes
(e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? Yes
(f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR? Yes
(g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset? Yes

6. Additionally, if you used crowdsourcing or conducted research with human subjects...
(a) Did you include the full text of instructions given to participants and screenshots? NA
(b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? NA
(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? NA
(d) Did you discuss how data is stored, shared, and deidentified? NA