# Identity Management and Mental Health Discourse in Social Media

Umashanthi Pavalanathan
School of Interactive Computing
Georgia Institute of Technology
Atlanta, GA 30308
umashanthi@gatech.edu

Munmun De Choudhury
School of Interactive Computing
Georgia Institute of Technology
Atlanta, GA 30308
munmund@gatech.edu

## ABSTRACT

Social media is increasingly being adopted in health discourse. We examine the role played by identity in supporting discourse on socially stigmatized conditions. Specifically, we focus on mental health communities on reddit. We investigate the characteristics of mental health discourse manifested through reddit's characteristic 'throwaway' accounts, which are used as proxies of anonymity. For the purpose, we propose affective, cognitive, social, and linguistic style measures, drawing from literature in psychology. We observe that mental health discourse from throwaways is considerably disinhibiting and exhibits increased negativity, cognitive bias and self-attentional focus, and lowered self-esteem. Throwaways also seem to be six times more prevalent as an identity choice on mental health forums, compared to other reddit communities. We discuss the implications of our work in guiding mental health interventions, and in the design of online communities that can better cater to the needs of vulnerable populations. We conclude with thoughts on the role of identity manifestation on social media in behavioral therapy.

## Categories and Subject Descriptors

J.4. [**Social and Behavioral Sciences**]: Psychology

## Keywords

health, identity, mental health, reddit, social media, throwaways.

## 1. INTRODUCTION

Online fora and support groups, including social media have emerged as prominent resources for individuals who are distressed by affective and behavioral challenges [10, 18, 30]. These tools act as a constantly available and conducive source of information, advice, and support [29]. An important motivation behind the use of these online resources for mental health concerns is also that they support open and honest discourse [33]. Self-disclosure can be an important therapeutic ingredient [22] and is linked to improved physical and psychological well-being [32].


*WWW 2015 Companion*, May 18–22, 2015, Florence, Italy.
ACM 978-1-4503-3473-0/15/05.
http://dx.doi.org/10.1145/2740908.2743049

The nature of online mental health discourse, however, may vary depending on the *nature of identity* adopted by an individual. This is likely to be particularly valid in the case of mental illness, since it is considered socially stigmatic [8]. Literature in sociology also supports this observation. In his celebrated book "Stigma" [14], Goffman examined how, individuals with a socially discredited attribute such as mental illness, tend to manage impressions of themselves in social settings—in order to protect their identities. However we note that in online settings, such as on social media, this constraint may be circumvented. This is because individuals may choose to withhold their actual identities allowing themselves to engage in more candid self-disclosure than is possible in offline settings, or through their identified online personas.

Our motivation for this research is also rooted in the rich literature on online identity construction, which has been recognized as a key aspect of online communities [11, 28]. Prior work demonstrates dissociative anonymity (a resistance to attach to offline identity or to their actual account/online persona), for instance, can be the foundation of online disinhibition [24]. Online disinhibition, the ability to avoid being "visible, verifiable, and accountable", leads people to act differently than they would in offline settings [6]. Social media naturally provides us with a rich ecosystem where we can study ways in which individuals manage their identities to engage in discourse on a stigmatized condition like mental illness.

In this light, this paper focuses on a relatively underexplored area of research involving characterization of behavior around a stigmatized condition, mental illness. In doing so, we extend our own prior work on examining mental health support on social media [10]. Here we investigate identity management in social media in the context of mental health. Specifically, we focus on the social media reddit, which, in contrast to other popular social media and networking platforms like Facebook and Twitter, provisions adoption of *'throwaway' accounts* as semi-anonymous identities for posting content. Throwaways are temporary accounts that reddit users create to dissociate from their primary reddit identity [25]. Most throwaway accounts are used exactly once [13]; thus their use disallows user behavior to be tracked historically, or through postings made from primary reddit accounts. Note that throwaways, however, do not adhere to a strict notion of anonymity [11], but research has shown that they are often used as *proxies of anonymity* [13, 34, 25]. We leverage this observation about throwaway accounts in our study.

Our primary contribution lies in characterizing how different forms of identity on mental health subreddits are associated with distinctive affective, cognitive, linguistic style, and social attributes. Leveraging measures derived from literature in psychology, which suggest language to be a reliable way of measuring people's internal thoughts and emotions, we also study the dif-

ferences in the nature of content shared in throwaway posts and posts from regular reddit accounts on these subreddits. Our findings based on a large corpus of reddit posts indicate the presence of almost six times more throwaway posts in mental health subreddits, in contrast to other subreddits. Thus throwaways may be fulfilling a unique need for individuals seeking to use reddit for discourse around a stigmatic health concern. Moreover, we observe that throwaway postings in mental health forums exhibit increased negativity, greater cognitive bias and self-attentional focus, lowered self-esteem and greater disinhibition, even to the extent of revealing vulnerability to self-harm. Through these findings, throwaways are observed to allow individuals to be less inhibited by self-presentation concerns, presumably due to lack of identifiability and accountability.

Our work, on one hand, indicates the potential of using social media for behavioral therapy. Community moderators may encourage semi-anonymous and disinhibiting discourse around stigmatized experiences. On the other hand, since some throwaway posts manifest self harm and depressive tendencies, our findings may guide the design of in-time, privacy-secure interventions which can bring help to vulnerable populations.

# 2. BACKGROUND AND PRIOR WORK

## 2.1 Identity in Online Communities

The work presented in this paper builds on prior research around online identity and its role in the success and dynamics of online communities [11]. McKenna and Bargh [26] found that anonymity on the web leads to lowered self-presentation concerns: "under the protective cloak of anonymity users can express the way they truly feel and think". In CMC literature, Donath [11] argued that anonymity can be "the savior of personal freedom, necessary to ensure liberty in an era of increasingly sophisticated surveillance". In a study on identity signals on 4chan, which has majorly (90%) anonymous posts, Bernstein et al. [2] found the use of alternative mechanisms such community-specific dialect, images and fluency in posts, for establishing status in the community. Schoenebeck [35] studied posts from the anonymous website YouBeMom.com and found that anonymity provides a comfortable environment for moms to converse without constraints of social norms. Recently Leavitt [25] performed an ethnographic study to examine the context in which *'throwaway' accounts* are used in reddit and found that the perception of anonymity shape the increased use of throwaway accounts.

Our work builds on this body of research and extends our initial findings on the unique use of throwaways in mental health communities on reddit [10]. Mental illness is a stigma [8]—individuals are often uncomfortable revealing such sensitive information or seeking help in face-to-face contexts, including avoiding availing counseling for fear of getting ostracized. We examine how proxies of anonymity, such as reddit's throwaways, may be providing users with an open and honest platform of discourse where such anxiety or trepidation are likely to be absent or minimal.

## 2.2 Discourse and Mental Health

Our second line of motivation is rooted in another rich body of work examining the important role of disinhibition in discourse around mental health. Disinhibition may result in increased self-disclosure [21], which is known to promote improved wellbeing and plays a positive role in psychological counseling and therapy. Jourard [22] reported that self-disclosure was a basic element in the attainment of improved mental health. Ellis [12] reported that discourse on emotionally laden traumatic experiences can be a safe way of confronting mental illness. On similar lines seminal work by Pennebaker et al. [31] found that participants assigned to a trauma-writing condition (where they wrote about a traumatic and upsetting experience) showed immune system benefits. Disclosure in this form has also been associated with reduced visits to medical centers and psychological benefits in the form of improved affective states [32].

A thorough treatment of detecting self-disclosure content in social media is not the focus of this paper. However our work relates to the above literature. We intend to examine how social media, through its ability to allow individuals to choose and regulate their online identities, might be providing new ways of candid expression amid a challenging health experience.

# 3. RESEARCH QUESTIONS

We contextualize the goals of this paper within the above body of work. We focus on the highly popular social media, curation, and news site reddit (`http://www.reddit.com/`). reddit allows its users to submit content in the form of links or text posts. Content entries, i.e., posts, are organized by areas of interest called "subreddits", such as politics, programming, or science. Each subreddit share the same platform mechanics, but is essentially an entirely different community with different rules, norms, and members.

Our notion of mental health communities in this paper leverages the presence of a number of mental health focused subcommunities on reddit [10]. Within these communities, for the purpose of our research, we categorize postings along two kinds of identities adopted by the post authors: **the 'throwaway' accounts** and the regular reddit accounts, which we call **the 'identified' accounts**. Throwaways are distinct from the regular reddit accounts as reddit allows an individual to create these accounts without giving out an email address, personally identifiable information or information about their regular reddit accounts. This allows discourse without drawing additional attention to a user's regular activities [25].

While in theory no personally identifiable information is shared by reddit users even through their primary accounts, it is possible that use of reddit over a long period of time may accumulate enough information in a user's postings. This may reveal aspects of their real lives [27], or of their behavior, personality, and attitudes. Hence for sensitive information disclosure, individuals have been known to prefer to use throwaways instead of their primary reddit accounts [13, 25]. Essentially, use of throwaways would hinder tracking behavior across postings of a user in multiple communities and prevent negative repercussions. Urbanski [34] also reported that the idea of anonymity through throwaways seems embraced in the reddit culture. Since regular reddit accounts may accrue rich context and information about an individual over long-term use of the account, we refer to them as 'identified' accounts.

We address the following inter-related research questions:

**RQ 1** *Are behavioral characteristics of discourse from throwaway accounts different in mental health forums versus other reddit communities?*

**RQ 2** *What are the behavioral characteristics of discourse from throwaway and identified accounts on mental health forums on reddit?*

**RQ 3** *Does the use of throwaway accounts promote sharing of more disinhibiting content on mental health forums compared to content from identified reddit users?*

# 4. DATA

We used reddit's official API (`http://www.reddit.com/dev/api`) to collect posts, comments, and associated metadata from several mental health focused subreddits. Our data collection

| | | | |
|---|---|---|---|
| depression | mentalhealth | Anger | ptsd |
| traumatoolbox | psychoticreddit | SuicideWatch | MMFB |
| getting_over_it | survivorsofabuse | alcoholism | BPD |
| rapecounseling | bipolarreddit | addiction | DPDR |
| hardshipmates | StopSelfHarm | panicparty | |
| socialanxiety | EatingDisorders | feelgood | |

**Table 1: Mental health subreddits used for crawl. (MMFB: MakeMeFeelBetter; DPDR: depersonalization, derealization; BPD: Borderline Personality Disorder).**

| | | | |
|---|---|---|---|
| thatHappened | philosophy | friendship | 505Nerds |
| MildlyInteresting | notfunny | Askreddit | oddlysatisfying |
| Fitness | funny | technology | lifeprotips |

**Table 2: Sample of subreddits belonging to the control group.**

proceeded using the method used in [10]. We first arrived at a comprehensive list of subreddits to focus on, with the help of reddit's native subreddit search feature (`http://www.reddit.com/reddits`). Specifically, we searched for subreddits on "mental health". Two researchers familiar with reddit employed an initial filtering step on the search results returned, so that we focus on high precision subreddits discussing mental health concerns. Thereafter, we focused on a snowball approach in which starting with a few seed subreddits (mentalhealth, depression), we compiled a second list of "related" or "similar" subreddits that are mentioned in the profile pages of the seed subreddits. The subreddits we crawled are given in Table 1. All of these subreddits host *public* content.

Additionally, for the purposes of statistical comparison, we identified a set of subreddits, sample of which is listed in Table 2 as our *control group*—meaning they are unrelated to mental health topics. To compile this list, we collected set of 1000 posts from the "New" category of reddit's home page and extracted the subreddits they were posted in— this method is likely to yield us a near-random sample of subreddits. On this candidate set of subreddits, to eliminate topical or contextual bias, we filtered a random sample of 25 to be our final control subreddit sample. Note that these control communities spanned a variety of sizes (e.g, r/Askreddit is large, r/thatHappened is smaller), and were used both for emotional (e.g., r/friendship) and objective discourse (e.g., r/505Nerds), as well as spanned multi-faceted topics.

For each of these subreddits (mental health related and control), we obtained daily crawls of their posts in the *New* category[1], similar to [10]. Corresponding to each post we collected information on the title of the post, the body or textual content, id, timestamp when the post was made, author id, comments, and the number of upvotes and downvotes it obtained[2]. The crawl of the subreddits used in this paper spanned between November 2013 and March 2014. We provide some basic descriptive statistics of the crawled dataset in Table 3 for mental health subreddits, and in Table 4 for the control.

We now highlight our method of identifying throwaway posts in our dataset. We used a two step process motivated by prior work [10, 13, 25]. First, one of the authors manually inspected a sample of reddit usernames in our dataset, and recorded the various naming strategies adopted by throwaway account owners. We thus identified a set of patterns typically prevalent in the throwaway accounts: *thrw*, *throwaway*, *throw*,*thrw*,*thraway*. This is motivated from the work of Gagnon [13] and our own work [10],

[1]For high traffic subreddits like Askreddit we obtained a random sample of 100 *New* posts per day.

[2]Users, also known as "redditors", can vote each reddit submission "up" or "down" to rank the post and determine its position or prominence on the site's pages. These two attributes associated with a post are referred to as "upvotes" and "downvotes".

| | |
|---|---|
| Total number of posts | 32509 |
| Total number of comments | 146082 |
| Total number of post authors | 23807 |
| Average posts per user | 1.37 ($\pm$3.41) |
| Total throwaway posts | 2552 (7.85%) |
| Total regular posts (non-throwaway) | 29957 (92.15%) |
| Total users posted throwaway posts | 2220 |
| Total users posted identified posts | 21615 |

**Table 3: Basic descriptive statistics of mental health subreddits.**

| | |
|---|---|
| Total number of posts | 21842 |
| Total number of comments | 831356 |
| Total number of post authors | 19345 |
| Average posts per user | 1.13 ($\pm$0.83) |
| Total throwaway posts | 299 (1.37%) |
| Total regular posts (non-throwaway) | 21543 (98.63%) |
| Total users posted throwaway posts | 282 |
| Total users posted identified posts | 19066 |

**Table 4: Basic descriptive statistics of control subreddits.**

wherein such patterns were adopted for identifying throwaway accounts with success. We labeled all the posts written using this type of usernanames as throwaway posts. Additionally, we looked for mentions of "throwaway" in either post titles or post text in the remaining posts and then one of the authors manually inspected each posts to label it as a throwaway post or not. This second step is motivated by the work of Leavitt [25]. In this way we compiled a high precision dataset on posts from throwaway accounts.

# 5. MEASURES

We propose four categories of attributes to characterize behavior manifested by throwaway accounts on the mental health communities we study. These are: (1) **affective attributes**, (2) **cognitive attributes**, (3) **linguistic style attributes**, and (4) **social attributes**. Measures belonging to all of these attribute categories are largely based on the psycholinguistic lexicon LIWC (`http://www.liwc.net`), and were motivated from prior literature on privacy, language and intimacy, and social spheres and information management [1, 11, 20, 19]. Additionally, we leverage insights from prior literature that examine association between the behavioral expression of individuals and their responses to traumatic context and crises, including vulnerability due to mental illness [7, 10].

We consider two measures of affect: positive affect (PA), and negative affect (NA), and four other measures of emotional expression: *anger*, *anxiety*, *sadness*, and *swear*. These measures are computed using the psycholinguistic lexicon LIWC.

We used LIWC to define the cognitive measures as well: (a) cognition, comprising *cognitive mech*, *discrepancies*, *inhibition*, *negation*, *death*, *causation*, *certainty*, and *tentativeness*; and (b) perception, comprising set of words in LIWC around *see*, *hear*, *feel*, *percept*, *insight*, and *relative*. Next, we consider four measures of linguistic style: (1) **Lexical Density**: consisting of words that are *verbs*, *auxilliary verbs*, *nouns*, *adjectives*[3], and *adverbs*. (2) **Temporal References**: consisting of *past*, *present*, and *future* tenses. (3) **Social/Personal Concerns**: included words belonging to *family*, *friends*, *social*, *work*, *health*, *humans*, *religion*, *bio*, *body*, *money*, *achievement*, *home*, and *sexual*. (4) **Interpersonal Awareness and Focus**: comprised words that are *1st person singular*, *1st person plural*, *2nd person*, and *3rd person* pronouns.

Our final set of measures are the social attributes. We utilized a variety of content sharing, social interaction, and social support

[3]Nouns and adjectives were detected using a standard POS tagger.

| Category | MH | Control | t-stat | $p$ |
|---|---|---|---|---|
| **Affective Attributes** | | | | |
| negative emotion | 0.0413 | 0.0239 | 9.476 | *** |
| anger | 0.0120 | 0.0077 | 4.415 | *** |
| anxiety | 0.0081 | 0.0055 | 4.118 | *** |
| sad | 0.0121 | 0.0037 | 6.289 | *** |
| **Cognitive Attributes** | | | | |
| *Cognition* | | | | |
| negation | 0.0309 | 0.0220 | 8.387 | *** |
| certainty | 0.0171 | 0.0131 | 4.800 | *** |
| death | 0.0043 | 0.0011 | 5.838 | *** |
| *Perception* | | | | |
| see | 0.0048 | 0.0080 | -6.362 | *** |
| feel | 0.0107 | 0.0078 | 4.263 | *** |
| **Linguistic Style Attributes** | | | | |
| *Lexical Density* | | | | |
| nouns | 0.1777 | 0.2146 | -8.808 | *** |
| verbs | 0.1899 | 0.1699 | 4.679 | *** |
| adverbs | 0.0658 | 0.0566 | 3.264 | ** |
| *Temporal References* | | | | |
| present tense | 0.1269 | 0.1033 | 8.293 | *** |
| past tense | 0.0391 | 0.0475 | -3.894 | *** |
| *Social/Personal Concerns* | | | | |
| friends | 0.0018 | 0.0027 | -3.113 | ** |
| social | 0.0839 | 0.1241 | -8.200 | *** |
| work | 0.0183 | 0.0272 | -3.232 | ** |
| health | 0.0138 | 0.0102 | 3.826 | ** |
| humans | 0.0090 | 0.0151 | -4.438 | *** |
| money | 0.0039 | 0.0071 | -5.437 | *** |
| sexual | 0.0049 | 0.0073 | -4.203 | *** |
| *Interpersonal Awareness* | | | | |
| 1st person singular | 0.1180 | 0.0802 | 15.910 | *** |
| 1st person plural | 0.0054 | 0.0112 | -4.429 | *** |
| 2nd person | 0.0045 | 0.0121 | -8.456 | *** |
| 3rd person | 0.0195 | 0.0295 | -6.260 | *** |
| **Social Attributes** | | | | |
| number of comments | 5.2614 | 30.5753 | -14.350 | *** |
| vote difference | 6.0733 | 43.0134 | -8.968 | *** |
| median comment length | 78.5943 | 38.8837 | 7.112 | *** |

**Table 5: Results of independent sample t-test between throwaway mental health posts (MH) and throwaway control posts. We report results at two different $\alpha$ levels: .01, .001. The significance values $p$ above are the adjusted $p$'s after adopting the Holm-Bonferroni method [17] which counteracts the problem of multiple comparisons here (total number of measures $m = 52$). This method is intended to control the Familywise error rate and offers a simple test uniformly more powerful than the Bonferroni correction [17].**

indicators as measures in this category. These are: *post length*, *number of comments*, *vote difference* (difference between upvotes and downvotes, divided by the total number of upvotes and downvotes), *comment arrival rate* (average time difference between any two subsequent comments in a post's comment thread), *time to first comment* (time elapsed between the first comment and the time the corresponding post was shared), and *median comment length*.

# 6. RQ 1: DIFFERENCES WITH CONTROL GROUP

Corresponding to our RQ 1, we begin by examining the discerning characteristics of posts made from throwaways in mental health (MH) subreddits compared to those in other subreddits (control subreddits as listed in Table 2). This will help us validate findings from psycholinguistics literature [7]: whether discourse from throwaway accounts in the mental health subreddits bears language markers known to be associated with psychological challenges such as mental illness.

We find statistically significant differences between the two groups. We observe almost 6 times more throwaway posts in the mental health subreddits than in control subreddits. This could be because of the stigmatic nature of mental illness; likely, people prefer to post under a (semi)-anonymous identity more frequently compared to other more mundane social contexts/topics.

This notable difference in the proportion of throwaway posts leads to exploring whether the throwaway posts in mental health subreddits differ from the control subreddits. We use the measures of social, affective, cognitive, and linguistic processes defined earlier for this purpose. We present statistical hypothesis testing between the cohorts using independent samples $t$-tests.

**Affective Attributes.** From Table 5, we observe notable differences—throwaway MH posts tend to use more attributes of negative affective processes such as "negative emotion" (almost twice of the control group), "anger" (almost twice of the control group), "anxiety" and "sad" (more than 3 times as control group). Prior literature indicates such increased negative affect to be associated with depression symptoms such as mental instability and helplessness, loneliness, and restlessness [9]. The following post excerpts illustrate this: *i know killing myself would hurt my family, but so would telling them that i'm a such a worthless failure.*

This tells us that reddit users find writing throwaway posts in the mental health subreddits as a forum to talk about their negative emotions in order to reduce stress and increase positive self-perceptions [12].

**Cognitive Attributes.** Throwaway posts in MH subreddits use more "negations" (1.5 times as control group) such as *no*, *not*, *never*: usage of which is associated with inhibition [24] i.e. MH posts from throwaway accounts are likely more self-conscious than users writing posts in control subreddits using throwaways: *it's a never ending cycle. therapy is not helping at all and i have tried several therapists. i have not seen my dr. since being put on the zoloft almost 2 months ago.* Throwaway MH posts also use more of "feel" and "insight" words which shows their expressions of feelings. Further, throwaway MH posts use more "death" related words (4 times as control group) which indicate their vulnerability to physical harm.

**Linguistic Style Attributes.** Throwaway MH cohort is found to be focused more on the present and the here and now (present tense) less on the past (past tense). Next, lower use of linguistic process term such as "family", "friends", "home", "social", "work", "humans", and "money" imply the throwaway MH redditors are less socially concerned or bothered. In fact, together with the fact that they also use greater number of first person pronouns shows that throwaway posts in mental health subreddits contain more personal stories and are in general, high in self-preoccupation [3]. Lower use of second person pronouns, first person plural pronouns and third person pronouns in throwaway MH posts implies that these redditors tend to be less socially interactive with the larger reddit audience. Increased use of "health" words in throwaway MH posts reveals that these posts talk more about their health related issues.

**Social Attributes.** Lastly, throwaway MH posts receive lesser social support i.e. lower number of comments (almost 6 times lower than control group). This might be due to the smaller size of the audience in the former—mental health related topics form a relatively more niche community compared to a subreddit like Askreddit. However, this cohort *does* receive comments at a faster rate (comment arrival rate almost 4 times than control group), as well as comments that are typically longer compared to the control group comments. This observation aligns with [10]—it is possible that the reddit audience tends to sympathize more with the throwaway MH posters, and provide more helpful and contributory feedback and opinions because of their honest confessions.

| Category | Throw. | Identified | t-stat | $p$ |
|---|---|---|---|---|
| **Cognitive Attributes** | | | | |
| negation | 0.0309 | 0.0282 | 5.537 | *** |
| certainty | 0.0171 | 0.0160 | 3.561 | ** |
| **Linguistic Style Attributes** | | | | |
| *Lexical Density* | | | | |
| verbs | 0.1899 | 0.1847 | 4.872 | *** |
| nouns | 0.1777 | 0.1893 | -5.396 | *** |
| *Temporal References* | | | | |
| *Social/Personal Concerns* | | | | |
| family | 0.0055 | 0.0046 | 4.948 | *** |
| health | 0.0138 | 0.0155 | -3.448 | ** |
| *Interpersonal Awareness* | | | | |
| 1st person singular | 0.1180 | 0.1123 | 6.141 | *** |
| 2nd person | 0.0045 | 0.0069 | -6.277 | *** |
| 3rd person | 0.0195 | 0.0175 | 3.733 | ** |
| **Social Attributes** | | | | |
| post length | 332.0886 | 253.5124 | 13.361 | *** |
| median comment length | 78.5943 | 64.9318 | 8.0622 | *** |

**Table 6: Results of independent sample t-test between throwaway mental health posts and identified mental health posts. Results reported with similar adjustment as in Table 5.**

# 7. RQ 2: THROWAWAY VS. IDENTIFIED MH REDDITORS

Findings on RQ 1 established the manner in which the throwaway redditors engaging in mental health discourse exhibit distinctive characteristics compared to those conversing about other topics. We now turn our attention to examining how the former cohort differs from those in the mental health subreddits who choose to reveal their identities (RQ 2).

**Cognitive Attributes.** Throwaway posts in MH subreddits use significantly more "negation" words than posts made from identified reddit accounts. We note that often negation exhibits in the context of a negative expression and bears a tone of confession: "*all of this is turning me into the person i don't want to be and i cannot see an easy way passed this"*, "*i dont know what to do now"*.

Moreover, throwaway posts present more cognitive biases through use of more "certainty" (e.g. *always, never*) related words. Although "certainty" words are associated with emotional stability [16], throwaway users use these words in a negative context which indicates that this cohort might be suffering from lowered self-esteem and displaying self-derogatory thoughts: e.g. *"i have always suffered from horrible anxiety"*, *"i'm always so worried"*, *"i have an amazing life. but i'm never happy"*, whereas identified redditors do not do so: e.g."*reddit has always been the place where i feel comfortable"*, *"i never remember doing anything bad"*, "*everyone always gives me warm greetings"*,

**Linguistic Style Attributes.** Throwaway postings use more verbs (which indicate discourse around actions), but less entities, e.g., nouns and adjectives. This reveals lowered interest in objects and things around them [5]. Expressing more about actions is found to be correlated with sensitive disclosure [19]. This implies throwaway posts revealing more sensitive information that could identify a person's routine, location or indented actions.

Throwaway redditors also talk more about the "past". However, their usage of "health" words is lower than posts from identified redditors. We conjecture this could indicate an effort to camouflage their ailment. It could also be because they are finding it difficult to come to terms with the reality that they are experiencing a direct or indirect psychological challenge. While the increased use of first person singular pronouns and lower use of second person pronouns in throwaway posts (similar to the findings in [10]) suggest that the users writing these posts are more self-focused and less interactive, the increased use of third person pronouns suggest they engage in greater discourse about others—likely about their friends and family, as we observed an increases use of "family" words.

**Social Attributes.** Finally, significantly higher length of posts correlates with the throwaway users [16]. This is known to be a sign of verbal fluency and cognitive complexity, which tells us that they might be engaging in more candid discourse given the (semi)-anonymous identity. The comment responses to these throwaway posts are also tend to be lengthier which might be the "reciprocity effect" [22]—when one shares more information, the responders tend to exhibit support by writing lengthier comments.

In summary, increased negative expression, raised cognitive bias and judgement issues, self-attentional focus, more candid discourse and lowered self-esteem in throwaway posts show that these posts bear distinctive behavioral markers in relation to stigmatic topics such as mental illness. Our findings thus are in the same vein as observed in our preliminary analysis in [10].

# 8. RQ 3: CONTENT DIFFERENCES

Our final research question investigates whether the use of throwaway accounts in mental health subreddits allows individuals to engage in sensitive and disinhibiting discourse about their health state and experiences. For this purpose, we look at frequently used uni-, bi-, and trigrams (top 500) in the throwaway posts as well as identified posts (Table 7, 8 list samples). We situate our findings in the light of content characterization around self-disclosure examined in prior work. We use a qualitative coding scheme to characterize n-grams from the throwaway and the identified accounts. In particular, we use a three-layer categorization scheme proposed by Altman and Taylor [1] to guide the content analysis of the depth of self-disclosure. Altman and Taylor suggest that disclosure can be categorized into either peripheral, intermediate, and core layers. The peripheral layer is concerned with disclosure around one's biographic information (e.g. age), the intermediate layer with attitudes, values and opinions, and the core layer spans disclosure of one's personal beliefs, needs, fears, and values. Aside from this characterization, Joinson [20] who characterized sensitive disclosure in terms of the extent of "revealed vulnerability".

**n-gram Usage in Throwaways.** We find considerable self-disclosure from the throwaway accounts along the intermediate and core layers. Corresponding to the intermedia layer of self-disclosure, we find the posters expressing their attitudes and opinion, which generally bear a negative tone ("shitty", "I don't want", "I've lost", "don't know what", "no idea", "the worst"): *my dad would beat the living shit out of me, at times to an inch before death. i've been to the hospital so many times i've lost track*.

Further, several posts from the throwaway cohort hint at a sense of urgency or desire to act: ("just need to", "to do something", "am going to"): *now i'm not crazy, i'm not a danger to any one, i just need to stay busy until i can see a new therapist in the next couple of days*.

Corresponding to the core layer of self-disclosure, we find throwaway posts extensively sharing posters' personal beliefs and fear. This might reveal their vital constructs and private, sensitive informational attributes ("if I could", "part of me", "because I know") [19]. The cohort also expresses a desire to avail help/need from the community ("want to talk", "what do i"): *i think about suicide at least once or twice a day but im not sure if i could go through with it.*

| | | | |
|---|---|---|---|
| i dont want | dont want to | want to kill | failed |
| want to die | i 've lost | because I know | upset |
| want to talk | i hate myself | dont know what | shitty |
| just need to | the worst | to do something | no idea |
| what do i | part of me | what I want | if I did |
| if I could | am going to | bring myself to | |

**Table 7: Uni-, bi-, trigrams from throwaway MH redditors.**

Next, many of the posts from the throwaway cohort indicate their vulnerability to physical, mental, or emotional harm with a self-loating tone; for instance, thoughts about killing themselves ("want to kill"), suicide ("want to die") and hating themselves ("I hate myself"): *i hate this. i hate myself. i don't want to f****** be this person anymore. i'm unmotivated, unfocused [...]*

Finally, we also observe that throwaway posts share a great deal of information confessing posters feelings about an incident, an experience, or a thought, sometimes for self-clarification or personal expression ("what I want", "don't want to", "I've lost", "don't know what", "bring myself to", "if I could", "if I did", "failed"): *i know i should leave, but i can't bring myself to because i want to be happy like before, before i found out about the drugs. so i've been debating ending it all.*

Another observation we draw is that some throwaway posts explicitly state the users' purpose of using a throwaway account, likely to avoid the possibility of being identified by their social circles (e.g. *"my girlfriend knows my regular account"; "friends know my main profile"*). Some also state that they are embarrassed because of the personal nature of the information being shared (e.g. *"as i am sad and embarrassed"; "i'm still not comfortable being open about taking antidepressant"*). In addition, some posts mentioned that using throwaway was "obvious"(e.g. *"clearly this is a throwaway"; "this is definitely a throwaway"*). These observations are consistent with recent work on boundary management and adoption of temporary accounts on reddit [25].

**n-gram Usage in Identified Posts.** redditors using their identified or regular accounts and engaging in mental health discourse, on the other hand, do not display as much intimate personal revelations, disparaging, or confessional thoughts corresponding to the intermediate or core layers of self-disclosure. Rather they focus on deriving constructive support from the reddit community. This is indicated in the usage of the frequent n-grams in Table 8 ("to help me", "how can i", "share", "to find a", "time to"): "*so i really just want to find a different doctor. who isn't rude*. In fact redditors with regular/identified accounts are also found to reveal their attitudes with a positive tone in contrast to the throwaway users: "*it'll be a little more difficult to study, but i'm sure i'll figure it out. i'll also try to seek help [...], so i hope to never have to come here and bother you guys once more.* ".

We thus conclude that throwaways allows reddit users engaging in mental health discourse to exhibit considerable disinhibition. The content of their discourse is distinct from that of identified redditors, in that it relates to the core aspects of the self rather than mundane aspects of social relations typical on social media [19].

# 9. DISCUSSION

**Theoretical and Practical Implications.** The distinctive use of throwaway accounts in the mental health forums indicates that such identities are fulfilling a unique need. They are allowing expressing views and thoughts about a topic that is often considered to be sensitive or unacceptable to the mainstream [8]. Presence of almost six times more throwaway redditors in mental health communities in contrast to other subreddits in our dataset reveals the wide spread use of throwaways in forums where sensitive topics are being discussed. Note that while it is possible that certain other reddit communities of deviant behavior (e.g., pornography) may have similar or higher proportion of throwaways, however compared to communities of non-deviant behavior, we do see a larger presence of throwaway users in mental health subreddits.

| | | | |
|---|---|---|---|
| to be able | so i can | to help me | i will be |
| and i want | i 'm sure | to find a | time to |
| share | how can i | i started to | to try and |

**Table 8: Uni-, bi-, trigrams from identified MH reddit users.**

Broadly, our findings align with prior literature [15, 23, 2, 35], where anonymity has been found to be a *desirable* attribute in certain online communities. Our findings show that, in the context of mental health, semi-anonymity enabled by throwaways allowed greater disinhibition. Psychology literature indicates that such disinhibition (in the form of journaling and discourse) can be an effective healing process [4, 31]. In fact, increased self-awareness and present-orientedness, affective experiencing and cognitive processing we observe in our data, are known to be associated with candid, ingenuous and sensitive discourse [19]. Nevertheless, note that a causal claim cannot be derived from our findings.

Our findings also bear implications for next generation social systems. The distinctive and candid discourse from throwaway redditors indicates that there is a need for social media designers to build appropriate recommendation, and support tools that can make such discourse fruitful for their psychological healing. We propose the following broad design considerations — (1) tools may be built to facilitate or encourage (semi)-anonymous participation as is enabled through throwaways for sensitive health topics; (2) community owners and moderators can enable better support mechanisms for such (semi)-anonymous participation; and (3) they can help direct in-time psychological services to help exceedingly vulnerable tendencies manifested in some throwaway postings, including self-derogatory and self-loathing thoughts and feelings of suicide.

In summary, we note the positive benefits of candid discourse in psychological healing as enabled by the use of reddit throwaways. However, forms of identity around health discourse raise important questions about our ethical responsibilities in identifying particularly vulnerable populations. We hope this work triggers conversations and involvement with the ethics and clinician community to investigate opportunities and caution in this regard.

**Limitations and Future Directions.** Importantly, our paper does not make any claim about attributing mental illness as a health concern to the posters of the reddit posts we study (per clinical DSM criteria). We also caution against drawing generalizations of this work. Our findings indicate the manner in which individuals might be adopting throwaway accounts on reddit to engage in discourse around a stigmatic condition like mental illness. However it is possible that the distinctive affective, stylistic, social, and cognitive attributes evident from our results are characteristic of reddit throwaways exclusively, and may not generalize to other social media that provision use of *true anonymity* (e.g., /b/ on 4chan). [11]. It will be fruitful to contrast reddit's use for mental health against other online health forums, so as to explore what makes reddit a unique platform for these issues.

# 10. CONCLUSION

In this paper, we presented a study on identity choices that pervade mental health discourse on social media, specifically reddit. We investigated behavioral differences in terms of social, affective, cognitive, and linguistic style attributes. We observed statistically distinctive characteristics of throwaway reddit postings around mental health topics. Our findings further indicated that semi-anonymity, as enabled by throwaways, provided a candid and disinhibiting platform of discourse. We believe our work will provide new insights to improving online social systems focused on supporting mental health discourse.

# 11. ACKNOWLEDGEMENTS

We thank Alex Leavitt, Eric Gilbert, and Amy Bruckman for helpful discussions, and the anonymous reviewers for their valuable feedback. This research is partly supported by a National Institutes of Health grant #R01 GM112697-01.

## 12. REFERENCES


[1] Irwin Altman and Dalmas A Taylor. *Social penetration: The development of interpersonal relationships.* Holt, Rinehart & Winston, 1973.

[2] Michael S Bernstein, Andrés Monroy-Hernández, Drew Harry, Paul André, Katrina Panovich, and Gregory G Vargas. 4chan and/b: An analysis of anonymity and ephemerality in a large online community. In *ICWSM*, 2011.

[3] Adriel Boals and Kitty Klein. Word use in emotional narratives about failed romantic relationships and subsequent mental health. *Journal of Language and Social Psychology*, 24(3):252–268, 2005.

[4] Philip A Burke and Rebekah G Bradley. Language use in imagined dialogue and narrative disclosures of trauma. *Journal of traumatic stress*, 19(1):141–146, 2006.

[5] David B Centerbar, Simone Schnall, Gerald L Clore, and Erika D Garvin. Affective incoherence: when affective concepts and embodied reactions clash. *Journal of personality and social psychology*, 94(4):560, 2008.

[6] Andrea Chester and Agi O'Hara. Image, identity and pseudonymity in online discussions. *International Journal of Learning*, 13(12), 2007.

[7] Cindy Chung and James W Pennebaker. The psychological functions of function words. *Social communication*, pages 343–359, 2007.

[8] Patrick Corrigan. How stigma interferes with mental health care. *American Psychologist*, 59(7):614, 2004.

[9] Munmun De Choudhury, Scott Counts, Eric Horvitz, and Aaron Hoff. Characterizing and predicting postpartum depression from facebook data. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 2014.

[10] Munmun De Choudhury and Sushovan De. Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In *Proc. ICWSM*. AAAI, 2014.

[11] Judith S Donath et al. Identity and deception in the virtual community. *Communities in cyberspace*, 1996:29–59, 1999.

[12] Darren Ellis and John Cromby. Emotional inhibition: A discourse analysis of disclosure. *Psychology & health*, 27(5):515–532, 2012.

[13] Tiffany Gagnon. The disinhibition of reddit users. *Adele Richardson's Spring 2013 ENC 1102*, 2013.

[14] Erving Goffman. *Stigma: Notes on the management of spoiled identity*. Prentice-Hall, 1963.

[15] Jonathan Grudin. Group dynamics and ubiquitous computing. *Communications of the ACM*, 45(12):74–78, 2002.

[16] Jeffrey T Hancock, Christopher Landrigan, and Courtney Silver. Expressing emotion in text-based communication. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 929–932. ACM, 2007.

[17] Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, pages 65–70, 1979.

[18] Christopher M Homan, Naiji Lu, Xin Tu, Megan C Lytle, and Vincent Silenzio. Social structure and depression in trevorspace. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pages 615–625. ACM, 2014.

[19] David J Houghton and Adam N Joinson. Linguistic markers of secrets and sensitive self-disclosure in twitter. In *System Science (HICSS), 2012 45th Hawaii International Conference on*, pages 3480–3489. IEEE, 2012.

[20] Adam N Joinson. Self-disclosure in computer-mediated communication: The role of self-awareness and visual anonymity. *European Journal of Social Psychology*, 31(2):177–192, 2001.

[21] Adam N Joinson and Carina B Paine. Self-disclosure, privacy and the internet. *The Oxford handbook of Internet psychology*, page 2374252, 2007.

[22] Sidney M Jourard. Healthy personality and self-disclosure. *Mental Hygiene. New York*, 1959.

[23] Cliff Lampe and Paul Resnick. Slash (dot) and burn: distributed moderation in a large online conversation space. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 543–550. ACM, 2004.

[24] Noam Lapidot-Lefler and Azy Barak. Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition. *Computers in human behavior*, 28(2):434–443, 2012.

[25] Alex Leavitt. "this is a throwaway account": Temporary technical identities and perceptions of anonymity in a massive online community. In *CSCW*, 2015.

[26] Katelyn YA McKenna and John A Bargh. Coming out in the age of the internet: Identity "demarginalization" through virtual group participation. *Journal of personality and social psychology*, 75(3):681, 1998.

[27] Emily Merritt. *An analysis of the discourse of Internet trolling: A case study of Reddit. com*. PhD thesis, 2012.

[28] David R Millen and John F Patterson. Identity disclosure and the creation of social capital. In *CHI'03 extended abstracts on Human factors in computing systems*, pages 720–721. ACM, 2003.

[29] Phoenix KH Mo and Neil S Coulson. Exploring the communication of social support within virtual communities: A content analysis of messages posted to an online hiv/aids support group. *Cyberpsychology & behavior*, 11(3):371–374, 2008.

[30] Minsu Park, David W McDonald, and Meeyoung Cha. Perception differences between the depressed and non-depressed users in twitter. In *Proceedings of ICWSM*, 2013.

[31] James W Pennebaker and Cindy K Chung. Expressive writing, emotional upheavals, and health. *Foundations of health psychology*, pages 263–284, 2007.

[32] Joshua M Smyth. Written emotional expression: effect sizes, outcome types, and moderating variables. *Journal of consulting and clinical psychology*, 66(1):174, 1998.

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

[34] Doug Urbanski. Upvoting the audience: a burkean analysis of reddit. 2013.

[35] Sarita Yardi. The secret life of online moms: Anonymity and disinhibition on youbemom. com. *ICWSM 2013*, 2013.