# Mental Health Support and its Relationship to Linguistic Accommodation in Online Communities

**Eva Sharma**
Georgia Tech
Atlanta, GA USA
evasharma@gatech.edu

**Munmun De Choudhury**
Georgia Tech
Atlanta, GA USA
munmund@gatech.edu

## ABSTRACT
Many online communities cater to the critical and unmet needs of individuals challenged with mental illnesses. Generally, communities engender characteristic linguistic practices, known as norms. Conformance to these norms, or linguistic accommodation, encourages social approval and acceptance. This paper investigates whether linguistic accommodation impacts a specific social feedback: the support received by an individual in an online mental health community. We first quantitatively derive two measures for each post in these communities: 1) the linguistic accommodation it exhibits, and 2) the level of support it receives. Thereafter, we build a statistical framework to examine the relationship between these measures. Although the extent to which accommodation is associated with support varies, we find a positive link between the two, consistent across 55 Reddit communities serving various psychological needs. We discuss how our work surfaces a tension in the functioning of these sensitive communities, and present design implications for improving their support provisioning mechanisms.

## ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

## Author Keywords
online communities; mental health; mental illness; social support; linguistic accommodation

## INTRODUCTION

> *I haven't been able to share this before... I'm doing okay, but I cry a couple of times a week. I remember that I was treated cruelly and I hate that all of that happened to me [...] I feel like my life is a lie.* – Paraphrased post excerpt shared in a Reddit mental health support community.

Social support is critical for attaining improved mental wellbeing [19, 55, 62]. The presence of support helps individuals deal with uncontrollable and emotionally crippling life events by providing a 'buffer' against the potentially adverse effects of stressful or difficult situations [20, 57]. However, outside of therapeutic contexts, vulnerable individuals often have limited ability to access adequate social support [50, 51, 78].

*Online mental health communities* (OMHCs), in recent years, have emerged as prominent resources for mental health support [84]. In fact, support derived from these communities has been found to causally improve mental wellbeing like reduced likelihood of suicidal thoughts [31]. Such support can range from *emotional support* (ES) to *informational support* (IS), often taking the form of empathy, acknowledgment, advice, or situational appraisal around diverse issues like mental illness, crisis, addiction, and abuse [17, 30, 31]. Moreover, due to the high quality of support provided by these OMHCs, they are also considered as a "safe haven": they enable individuals to express disinhibiting emotions, engage in self-disclosures of stigmatized experiences, and develop trusted relationships with peers [1, 6, 17, 30, 31]. Broadly speaking, OMHCs provide a support mechanism to cater to the timely and situationally critical needs of vulnerable individuals, as illustrated by the paraphrased quote above.

The identity of any community, including OMHCs, is defined by a set of conventions and traditions, referred to as *norms*. Conformance to these norms indicates an individual's commitment to a community, in which case, they are rewarded with social approval and acceptance [69]. Per the Speech Accommodation theory by Howard Giles [42], people often achieve normative conformance with a community through linguistic means: "*when people interact they adjust their speech, their vocal patterns and their gestures, to accommodate with others*". More specifically, a community's conventions are usually carried over time and manifest themselves via communication rules or linguistic styles established by its members [8, 37, 54, 67]. Such *linguistic accommodation* is linked to improved social feedback, increased solidarity, better social exchanges, and reciprocated feelings of intimacy [39].

Although these insights have been validated in general purpose online communities [25, 28, 45, 64], there is a lack of similar empirical evidence indicating that in OMHCs, linguistic accommodation translates to social support: a specific but important form of social feedback. Since an ability to seek and provide social support is at the crux of the overall goals of OMHCs, examining the link between support and accommodation can be beneficial to understand the communities' underlying social processes—that is, how, in the light of their



*CHI 2018,* April 21–26, 2018, Montreal, QC, Canada



DOI: https://doi.org/10.1145/3173574.3174215

ingrained linguistic norms, OMHCs respond to the critical psychological needs and requests of vulnerable individuals.

**Our Work.** In this paper, we address the question: *how do members' linguistic accommodation practices impact the support provisions in OMHCs?* Towards this research goal, we study a large corpus of publicly shared posts and comments from 55 OMHCs on Reddit that provide help and advice on issues like depression, anxiety, addiction, and trauma. First, we develop a supervised learning technique to assess the levels of ES and IS received by the posts. Then, we adapt a well-validated method, the Linguistic Style Matching [45], to quantify linguistic accommodation expressed by the posts' authors. Finally, using multinomial logistic regression, we examine the relationship between the measures of linguistic accommodation in posts and the amount of support they receive from the community.

Although in the Reddit communities we study, support is intended to be geared towards the most urgent and critical requests they receive [5, 61, 65], find that not only the type of support offered by them depends on the issues they deal with but also its level of support depends on the extent of linguistic accommodation demonstrated by the support seekers in their posts. In fact, across these diverse OMHCs, we *consistently* observe that with an increase in the linguistic accommodation in a post, there is an increase in the ES or IS it receives.

Our results introduce new insights into the functioning of OMHCs, specifically, in the manner in which ES and IS are influenced by support seekers' conformance to the community's norms. In fact, we note an apparent tension between the importance of compliance to norms by support seekers and the intention of the support providers to direct timely help in these OMHCs. We discuss the implications of our work for online support relating to mental health challenges. We conclude by outlining design suggestions that OMHCs can incorporate to direct and improve support around the unique, critical, or unmet needs of distressed individuals.

**Privacy, Ethics and Disclosure.** Given the sensitivities of the online communities examined in this paper, we have adopted some cautionary steps. Although we work with public data, we have not included any personally identifiable information, and have paraphrased the reported quotes to protect the privacy of the users. Some of the these quotes contain language that may be perceived to be emotionally triggering.

## RELATED WORK

### Social Support in Online Mental Health Communities

*Role of Social Support in Health and Well-Being*

Since the 1970s, there has been significant interest in understanding the role of social support as a coping resource and in aiding psychological adjustment and illness recovery [6, 20, 58]. Kaplan defines social support as "*the degree to which an individual's needs for affection, approval, belonging, and security are met by significant others*"[49]. A number of studies, notably of Cohen and Willis, have demonstrated that the adequacy of social support is directly related to the severity of psychological symptoms and/or acts as a buffer between distressful events and stress [20]. Importantly, Goffman noted that individuals with emotional distress in particular, benefit from interactions with and support from "sympathetic others" who share the same social stigma and experiences [43].

Recognizing these benefits, Cutrona and Suhr [23] developed a helpful categorization schema "Social Support Behavioral Code" for understanding and assessing support. Two categories of support proposed in this schema have received the most theoretical and empirical attention: *emotional support* and *informational support* [71, 80], a classification we adopt in our work. Other research has noted that the need for the amount and type of support depends on the nature of the distressful experience and the criticality of psychological needs [22, 29, 58, 74, 79]. We situate our study within the context of this prior research.

*Studies of Online Mental Health Communities*

Over the years, researchers have repeatedly observed that online communities may serve as sources of peer-to-peer social support around diverse health challenges [36]. A key aspect of these communities is that they provide members with access to other people with similar challenging conditions, including those with more experience dealing with relevant health issues [63]. Members of OMHCs receive ES either directly, through empathetic messages, or indirectly, by being exposed to others having similar experiences [1, 4, 33]. They also gain IS by receiving helpful information and advice related to treatment and medication, identifying possible explanations to their problems, and building social capital [14, 46].

Prior work has also exclusively explored the characteristics and dynamics of OMHCs, and how they enable support seeking and offering mechanisms, especially in populations with stigmatized experiences, and vulnerable emotional and mental health status [30]. Andalibi et al. [1] studied how individuals with experience of sexual abuse history seek support in relevant online communities on Reddit. Another study found that these communities provide critical help in addressing clinical questions and building a constructive information sharing environment [47]. More recently, De Choudhury et al. quantified the extent to which support obtained from OMHCs effectively reduces depression, and improves mental well-being [31]. Other studies, by differentiating between the two forms of support: ES and IS, have analyzed their efficacy in these communities and how support impacts member retention [12, 79, 83, 87]. Our work is situated centrally in this body of research, wherein we first propose a content-driven approach to automatically detect levels of ES and IS in posts shared in OMHCs, and then examine how they vary across different types of communities.

### Community Norms and Linguistic Accommodation

Community members adopt common set of values, known as norms, to demonstrate their commitment to the community and gain a sense of belonging [10, 13]. In the absence of reliable community-specific "assessment signals" [32], norms, which are central tenets of online communities [69, 84, 86], often manifest themselves as stable linguistic practices, enabling the establishment of a common context for the community members [8, 37, 54, 67]. Therefore, we adopt the language of content (posts) shared in OMHCs as a mechanism to identify their underlying norms.

Further, not only do linguistic practices of members signal community norms, as noted above, the Speech Accommodation theory suggests that people adjust their linguistic style with respect to that of another's during social interactions [41]. Such convergence induces similar linguistic behavior in the community members [9, 15]. Drawing from this theory, in this paper, we investigate whether and how linguistic accommodation presents itself in different OMHCs.

In a complementary line of work around community norms, it has been found that the social exchanges between people become more effective when individuals assess their course of action in terms of the costs and perceived social rewards [21, 35]. This suggests that an individual can seek more favorable evaluation of themselves from another person by reducing dissimilarities between them [39, 35]. For communities, this cost translates to the effort required to align oneself linguistically with the rest of the group, rewarding them with increased feedback, acceptance, inclusion, social capital, and status [42].

Several studies have explored the aforementioned social rewards elicited by a member's linguistic alignment with the community in different online settings [26, 27, 40, 64]. Ireland et al. found that linguistic style matching reflects implicit interpersonal processes central to romantic relationships [48]. Gonzales et al. [45], who also proposed one of the psycholinguistically grounded measures of linguistic accommodation known as Linguistic Style Matching (LSM), found that, in the context of smaller online groups and dyadic conversation pairs, linguistic accommodation was a predictor of underlying social dynamics, specifically group cohesiveness, performance and trust between the members. In studying chronic disease-related online communities, Park et al. observed that vocabulary alignment with the rest of the community resulted in increased number of comments, thus prolonging that individual's participation in the community [66]. Similarly, correlating comments' linguistic alignment with support in a cancer support community, Wang et al. analyzed the influence of a post's first commenter on other commenters' behavior in threaded conversations [81]. While prior work has largely studied social affinity or rapport as a common form of social feedback, we look at a more nuanced form of social feedback appropriate in the context of OMHCs – social support which in turn comes from community acceptance and rapport. Thus, our work advances prior investigations by examining how the extent of linguistic accommodation in OMHCs influences the levels of ES or IS received by the support seekers.

## DATA

### Data Acquisition

We used publicly accessible data from Reddit which is a widely used online forum. On the platform, registered users share content in the form of text, links and images. Users can create a new post or comment on existing posts. These posts are organized by their topic of discussion into a variety of communities known as "subreddits".

In recent research, Reddit has been known to facilitate mental health discourse [30] through various subreddits: such as depression (*r/depression*), anxiety (*r/anxiety*) and suicidal ideation (*r/SuicideWatch*). Such communities also provide ES and IS [2] to individuals coping with psychological distress (e.g., *r/helpmecope*, *r/rapecounseling*, *r/traumatoolbox*).

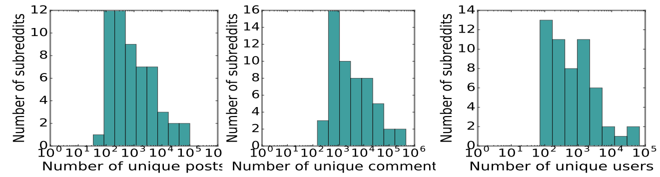



Figure 1: In order from left: Distribution of: subreddits over total number of posts, subreddits over total number of comments, and subreddits over total number of users.

To compile a comprehensive list of Reddit OMHCs for our study, we employed snowball sampling. First, drawing from prior work [30], we used the subreddit search functionality to craft various search queries using mental health issues as keywords like *support, counseling* and *mental health*. We iteratively augmented them based on related keywords that frequently co-appeared in the community descriptions. Our final set of keywords were: *support, counseling, mental health, mental, trauma, abuse, depression, suicide, therapy* and *coping*. Using these keywords, we compiled a list of 55 OMHCs.

We then leveraged the Reddit data archive available on Google's BigQuery [11]. BigQuery is a cloud based data warehouse, that allows third parties to access large publicly available datasets through simple SQL-like queries [38]. For all the 55 communities, we extracted all the posts and comments made between January 2014 and August 2016. This provided us with 358,409 posts and 1,832,702 comments across those 55 communities, with a mean of 6,516.53 posts ($\sigma$ = 18,663.22) and a mean of 33,321.85 comments ($\sigma$ = 87,638.06) per community. Our dataset included 245,527 unique users. Figure 1 presents the distribution of posts, comments and users. To collect data on each community's usage statistics, we crawled RedditMetrics [60] during our period of analysis (Jan 2014-Aug 2016) and found that these communities spanned various sizes, with mean subscriber count ranging between 132-189,922 ($\mu = 100,151.58; \sigma = 67,470.30$).

### Categories of Mental Health Communities

Given the large number of communities in our dataset, to simplify our ensuing analysis, we categorized them based on the broader topics and mental health issues they focus on. To do this categorization, we used a two-step unsupervised learning based machine-human framework. Our approach was motivated by two observations: 1) human labeling can help extract semantically meaningful and contextually relevant community categories, but is difficult to scale, and 2) clustering techniques are scalable, but, on their own, may not provide meaningful groupings of the OMHCs.

(1) First, we used $k$-means clustering algorithm to perform initial clustering on the $n$-grams ($n = 3$) of the posts shared in the 55 OMHCs. We applied a parameter sweep for $k$, the number of clusters, between 2 and 8, and determined the optimal $k$ to be 5 based on the largest Silhouette coefficient: a measure of how similar an object is to its own cluster (cohesion) compared to other clusters (separation).

| Community Category | Communities | #Posts | #Comments | #Users |
|---|---|---|---|---|
| Trauma & Abuse ($C_1$) | *r/abuse, r/adultsurvivors, r/afterthesilence, r/Anger, r/bullying, r/CPTSD, r/domesticviolence, r/emotionalabuse, r/ptsd, r/PTSDCombat, r/rapecounseling, r/StopSelfHarm, r/survivorsofabuse, r/SurvivorsUnited, r/traumatoolbox* | 11,213 | 44,486 | 8,661 |
| Psychosis & Anxiety ($C_2$) | *r/Agoraphobia, r/Anxiety, r/BipolarReddit, r/BipolarSOs, r/BPD, r/dpdr, r/psychoticreddit, r/MaladaptiveDreaming, r/Psychosis, r/PanicParty, r/schizophrenia, r/socialanxiety* | 70,696 | 374,072 | 41,580 |
| Compulsive Disorders ($C_3$) | *r/calmhands, r/CompulsiveSkinPicking, r/OCD, r/Trichsters* | 8,032 | 35,948 | 5,340 |
| Coping & Therapy ($C_4$) | *r/7CupsofTea, r/BackOnYourFeet, r/Existential_crisis, r/getting_over_it, r/GriefSupport, r/helpmecope, r/hardshipmates, r/HereToHelp, r/itgetsbetter, r/LostALovedOne, r/offmychest, r/MMFB, r/Miscarriage, r/reasonstolive, r/SuicideBereavement, r/therapy* | 100,248 | 510,168 | 77,031 |
| Mood Disorders ($C_5$) | *r/depression, r/depressed, r/ForeverAlone, r/GFD, r/lonely, r/mentalhealth, r/Radical_Mental_Health, r/SuicideWatch* | 168,220 | 868,028 | 112,915 |

Table 1: Five Reddit OMHC categories and their associated subreddits used in this paper, obtained through $k$-means clustering following human annotation. Descriptive statistics of each category are also shown.

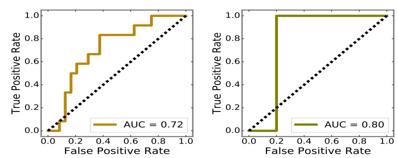
(2) Next, two annotators, familiar with OMHCs on Reddit, independently refined these machine-labeled clusters and assigned them suitable descriptive labels using a semi-open coding approach. For the purpose, they actively referred to the textual descriptions given in the subreddit landing pages, borrowing from the conceptualizations of different mental health conditions in DSM-5 [3]. Finally, they identify five categories of mental health communities: Trauma & Abuse, Psychosis & Anxiety, Compulsive Disorders, Coping & Therapy, and Mood Disorders (see Table 1 for their descriptive statistics).

## METHODS

### Support Classification

We now present a machine learning framework to quantify support in OMHCs. We adopt the two-class characterization of online social support—emotional (ES), and informational (IS) [12, 23, 83]. We infer the levels of ES and IS from the comments made on the posts in our OMHC dataset [31].

Starting with our entire corpus of 1,832,702 comments (excluding ones by authors of the corresponding posts), we selected a random sample of 400 comments (three removed due to duplication). Then two annotators, familiar with OMHCs on Reddit, separately coded each of them for the level of ES and IS, using a three-point Likert scale (1=least supportive, 3=most supportive) [59]. Per prior work [83], in case of IS, the annotators looked for information or advice about treatment, whereas for ES, they looked for empathy, encouragement and kindness. At the end of the annotation task, the annotators met to resolve disagreements, and the inter-rater agreement metric Cohen's $\kappa$ was found to be 0.879 and 0.876 for IS and ES respectively. For each coded comment, we computed a final score for both ES and IS; we used a balanced bucketing approach to get integer scores over the average of the rater codes. In the entire coded sample of 400 posts, there were 185, 131, and 81 posts with IS scores 1, 2, and 3, while 135, 134, and 128 posts with ES scores 1, 2, and 3 respectively. An example paraphrased comment from the Psychosis & Anxiety community category that received 1 and 3 for IS and ES said: "*You're just like the rest of us [...] kindred spirits here! We can all be borderlines together*".




| | $\mu_{ES}$ | $\sigma_{ES}$ | $\mu_{IS}$ | $\sigma_{IS}$ |
|---|---|---|---|---|
| A | 0.59 | 0.04 | 0.72 | 0.03 |
| P | 0.66 | 0.06 | 0.68 | 0.03 |
| R | 0.59 | 0.04 | 0.72 | 0.03 |
| F1 | 0.57 | 0.04 | 0.68 | 0.02 |

Figure 2: ROC curve and performance metrics (A: Accuracy, P: Precision, R: Recall, F1: F1-score) for ES (left) and IS (right) classification. The metrics are reported based on $k$-fold cross-validation ($k = 10$).

Using these annotated comments as training data, we then built two tertiary classification models (support score classes: 1, 2, and 3) corresponding to the two forms of support. Borrowing from prior work, we employed the well-validated psycholinguistic lexicon, Linguistic Inquiry and Word Count or LIWC [75] to calculate classification features in comment text that could be predictive of ES or IS. We specifically focused on a set of 50 LIWC categories which have been found to be characteristic of mental health related social media content [75]. IS classification model included an extra binary indicator feature corresponding to the presence of URLs, as the comments providing IS often include links to information regarding treatment and medication.

We trained several classification models corresponding to each form of support. A multinomial logistic regression model was observed to yield the best performance (in terms of accuracy and F-1 score) for predicting the ES of a comment; and a Support Vector Machine (SVM) classifier with a polynomial kernel of degree 3 performed the best for predicting IS. The performance of both the classification models was evaluated using $k$-fold cross-validation ($k = 10$). Figure 2 gives different performance metrics of the classifiers. ES classifier achieved a mean accuracy of 59%, which is better than the baseline accuracy of 35%. Similarly, for IS classifier, the mean accuracy was 72%, which we again found to be an improvement over baseline accuracy of 60%. These are further bolstered by the results of [83] that presents models with correlations(0.76-0.80) between human and machine labels for ES and IS, which are not significantly different from our case.

Next, we used these two classifiers to machine label the held-out comments from our dataset with ES and IS scores (1, 2, or 3). One expert annotator manually cross-verified a random sample of 100 comments from this machine labeled dataset. This activity yielded IS and ES accuracies of 66% and 73% respectively, which is consistent with the performance of the classifier, indicating its robust performance. Finally, for each post, we calculated the averages of ES and IS for all of its comments, followed by balanced bucketing, to obtain aggregate ES and IS scores received by that post.

### Modeling the Link between Support and Accommodation

*Measuring Linguistic Accommodation*

Next, we present our method for measuring linguistic accommodation expressed in the posts of our dataset. Our work adopts from one of the approaches to measure of linguistic accommodation [45, 77]. We employ technique, known as the Linguistic Style Matching (LSM), that assesses the alignment between the linguistic style of an individual and that of the community's. It considers the rate of use of function (e.g., prepositions, conjunctions, articles, and other content-free parts of speech [75]) words in an individual's speech (content) to be a proxy for stylistic alignment as they help identify relationships between language and social psychological states [16, 18]. Since, they are subconsciously produced and are difficult to manipulate in one's speech pattern, they are appropriate for this study. We adapted the LSM algorithm via the following pipeline of steps (summarized in Figure 3):

(1) Linguistic accommodation is established over time [28, 76] and is quantified in terms of how closely an individual's linguistic style follows that used by the community in the past. Thus, we split posts in each community into two parts, $P_1$ and $P_2$, based on the median of their timestamps. Here, $P_1$ and $P_2$ are two set of posts made before and after the median timestamp for the respective community. We expect that the linguistic accommodation of a post in $P_2$ would depend on the style manifested in posts in $P_1$ along with any post in $P_2$ with a smaller timestamp.

(2) We begin by calculating the proportion of function words in $P_1$ and $P_2$ posts using the 12 categories in LIWC [75].

(3) Then, corresponding to each post $p$ in $P_2$ in a community, we construct a sequence of historical posts in that community—i.e., all $P_1$ posts, along with those in $P_2$ with a timestamp smaller than that of $p$. Using the mean proportion of function words in all historical posts of $p$, we obtain $p$'s LSM score by calculating the ratio of the absolute difference between $p$'s function word usage and that of its prior posts, averaged across all function word categories.

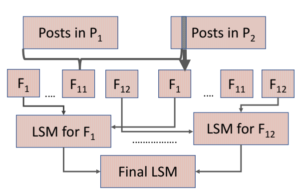



Figure 3: Steps to calculate a post's LSM score; $F_i$ is the $i$-th function word.

*Statistical Models*

Finally, we present a set of statistical models that we employ to assess the relationship between linguistic accommodation (LSM score) of a post and the support it receives within a community. We develop two sets of nested multinomial logistic regression models, one for ES and the other for IS. Our response variable for each set of models is the support score of a post, which is a 3-class outcome variable exhibiting the values 1 (low), 2 (medium) and 3 (high) for ES and IS. Each set of models, in turn, consists of: (a) a "control model" meant to quantify the relationship between confounding predictor variables and the support scores, and (b) "LSM model" which includes the LSM score of a post as an *additional* predictor variable. Drawing from prior work [1, 82], we include the following confounding predictor variables—post specific variables (items 1, 2 below), post author specific ones (items 3, 4), and community-oriented variables (item 5):

(1) *Content*: We capture a post's topical content as confounding predictor, as it affects the extent of support elicited by that post. For this, we used all the non-function word categories in the LIWC dictionary [75].

(2) *Post Length*: Since, lengthy posts receive greater support in online communities [2], we included the number of whitespaced words (post length) as a control variable.

(3) *Self-Disclosure*: The amount of self-disclosure in a post also affects community feedback, including social support [1, 2, 82]. We used the technique from prior work [82] to obtain ground truth labels for the amount of self-disclosure in the posts in our dataset and included it in our models[1].

(4) *Author Tenure and Familiarity*: An individual's knowledge of the community's norms also impacts the support that they receive from the community [28, 64]. We quantify this "community knowledge" in two ways. First, we determine the tenure of a post's author in a community, calculated as the time difference between the timestamp of their first post (in our dataset) and that of the current post. Second, we quantify a post author's familiarity with a community by identifying the number of posts made by them in that community before the current post; a higher number is likely to indicate greater familiarity with the community's norms.

(5) *Throwaway Account*: The anonymity of a post's author impacts levels of social feedback and support [2]. On Reddit, anonymity is established by creating temporary accounts, also known as throwaway accounts [56]. We consider a post to be shared from a throwaway account if the word "throwaway" or a lexically similar variant is used either in the username, in the title, or body of the post as used in prior work [56].

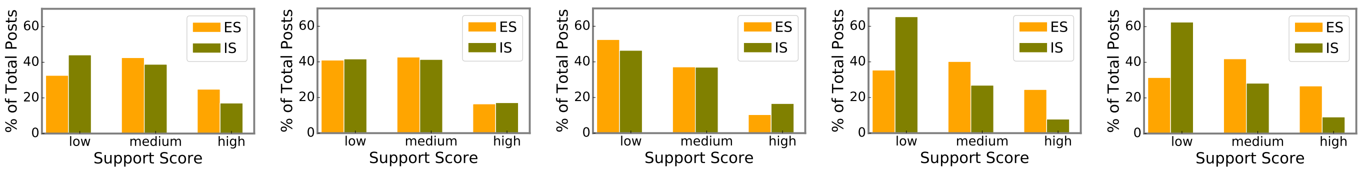
(6) *Size of the Community*: We include the number of community subscribers as a variable in our models as a large community invites more comments on its posts, which in turn, can inflate the observed ES or IS.

## RESULTS

### Support Classification

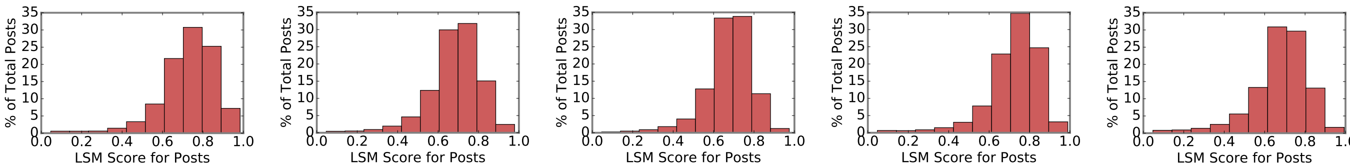
To begin, we analyze the outcomes of the support classifiers that machine labeled all the comments in our dataset.

[1] With the help of two human annotators and following the guidelines in [82], we first coded 400 randomly sampled posts with an inter-rater reliability score of 0.76 (Cohen's $\kappa$). Then using the labeled posts as training data, we developed a 3-class disclosure classifier, a logistic regression model (1 being low and 3 being high). The classifier used normalized n-grams(n=1,2,3) as features. The model had 68% accuracy and 65% F1 score.




(a) Trauma & Abuse (b) Psychosis & Anxiety (c) Compulsive Disorders (d) Coping & Therapy (e) Mood Disorders

Figure 4: Proportion of posts with low (1), medium (2) and high (3) ES and IS per community category.




(a) Trauma & Abuse (b) Psychosis & Anxiety (c) Compulsive Disorders (d) Coping & Therapy (e) Mood Disorders

Figure 5: Distribution of LSM score for total percentage of posts per community category.

We present the distribution of ES and IS aggregated by community categories. Figure 4 shows the distribution of support scores over percentage of total posts that received the respective type of support. Communities related to Trauma & Abuse, Coping & Therapy and Mood Disorders received greater amount of high ES (=3) than that of IS by 8%, 16% and 17% of posts respectively. Additionally, they received greater amount of low IS (=1) as compared to that of ES by 11%, 30% and 31% of posts respectively. This could imply that these communities predominantly provide more ES as compared to IS, which could be due to the needs of individuals approaching these communities or the goal of the communities itself. For instance, one of the communities that we consider to be a Trauma & Abuse community describes itself as "*a place for survivors of all abuse to come together to share their stories, vent, and to assist one another in healing*".

Next, for Compulsive Disorder related communities, 6% of the posts received greater amount of high IS than ES. One of the Compulsive Disorder communities describes itself to be a place for people suffering from OCD to come together and exchange information about treatment. In one of its posts titled as (paraphrased) "*Surrender to the Urge?*", the poster describes their struggle with the need to count things repeatedly and requests for advice. This post scored a high mean IS across all comments, indicating that its comments provided the type of support the poster was seeking for.

We also observe that the distributions of IS and ES scores are roughly equal for communities related to Psychosis & Anxiety. This presumably indicates that the communities from this category equally provide both types of support in terms of advice as well as encouragement and a sense of belonging. For example, in a Psychosis & Anxiety community, one of the post (paraphrased) is about the poster being anxious: "*...I can't do normal things and I have jury duty next month, any tips?...*" received 2 for both IS and ES. Out of the 13 comments on this post, one comment that got IS of 2 says: "*I just checked social anxiety disability box of the form and was allowed to go home [...] you can try this if you want.*". Another comment which got an ES of 2 says, "*It probably won't be as bad as you imagine. I hope it helps. Best of luck*".

Summarily, we find that across all five community categories, ES is higher in case of Trauma & Abuse, Coping & Therapy and Mood Disorders, whereas communities related to Compulsive Disorders tend to provide more IS. Psychosis & Anxiety related communities provide both types of support equally.

### Examining Support and LSM

First, we examine the patterns of linguistic accommodation manifested in the posts belonging to the five community categories, as measured via the LSM measure (ref. Methods). Figure 5 presents the distribution of LSM scores for all the community categories. The histogram in each case shows that most of the posts have higher LSM scores with the mean and the standard deviation of about 0.65 and 0.1 respectively. This observation indicates that people tend to conform with the language conventions of a community and mimic the style of writing of the rest of the community. Moreover, we find that the mean LSM for all community categories lies in the range of 0.63 to 0.68. This further indicates that although there are some discrepancies in the extent of linguistic accommodation in different communities with varied goals and purposes, posts in these communities consistently demonstrate a tendency of aligning their linguistic styles with what the community, values and identifies with. Similar high LSM in smaller groups has also been reported in prior work [44].

Now we analyze the relationship between LSM exhibited in posts and the IS/ES they receive—the primary goal of our work. We examine the distribution of LSM scores over IS and ES in the community categories, as shown in the box and whisker plots of Figure 6. We observe that the boxes within a community category are of nearly the same height which means that majority of posts for a given value of support (1, 2, or 3) agree on the LSM score. We also notice that with the increase in both IS and ES in any community category, there is a small but consistent increase in linguistic accommodation.

However, we do notice that this increase is more pronounced in case of community categories Trauma & Abuse and Coping & Therapy for ES. A similar increase is evident in case of the Trauma & Abuse and Mood Disorders communities for IS. Since Coping & Therapy category includes communities where individuals seek psychotherapy and counseling, it is more likely that higher linguistic alignment in posts will result in greater expression of empathy, kindness, and emotional acknowledgment from the community. On the other hand, in Mood Disorder communities, higher linguistic accommodation may indicate that posters are seeking concrete, direct ways of help and advice, which the community seems to provide.

We further investigate the linguistic attributes of posts in various communities where the relationship between LSM and IS/ES is most pronounced. For this, we consider posts having both LSM and support either high or low. We divide the posts in a community category into two groups – one with posts having high support and high LSM and the other with the posts eliciting low support and low LSM. In case of high LSM, we include posts with LSM score higher than one standard deviation over mean, and similarly for low LSM, we consider those with LSM score lower than one standard deviation below mean. And, the high support posts are the posts that received support score of 3 whereas the low support posts include the ones with support score 1.

| | |
|---|---|
| **Informational Support (IS)** | |
| *High LSM, High Support* | |
| $C_1$ | depressive, music, insight, episode, classes |
| $C_2$ | pains, blood, nausea, xanax, prozac |
| $C_3$ | insulin, stab, luvox, abnormal, scalp |
| $C_4$ | medication, motivation, doctor, treatment, insurance |
| $C_5$ | psychiatrist, symptoms, medication, prescribed, disorder |
| *Low LSM, Low Support* | |
| $C_1$ | black, scars, name, write, girls |
| $C_2$ | victories, sharing, app, dpdr, message |
| $C_3$ | simply, gift, sides, learn, pull |
| $C_4$ | moon, fuck, you, your, shame |
| $C_5$ | bills, women, loans, birthday, girls |
| **Emotional Support (ES)** | |
| *High LSM, High Support* | |
| $C_1$ | awkward, disgusted, wonderful, trusted, hated, |
| $C_2$ | relationship, distance, dating, failure, hurts, deserve |
| $C_3$ | extremely, ruminations, fruit, plagiarism, curiosity |
| $C_4$ | suicidal, blame, dead, feelings, childhood |
| $C_5$ | together, killing, hurting, husband, loved |
| *Low LSM, Low Support* | |
| $C_1$ | treatments, ice, plan, book, bear, |
| $C_2$ | coffee, caffeine, disease, diet, asleep, |
| $C_3$ | math, amp, contamination, posted, driving |
| $C_4$ | moon, fuck, dollars, water, teeth |
| $C_5$ | loans, steam, bills, information, curious |

Table 2: Results from Sage Analysis showing frequent and infrequent words used in posts with high LSM and high support score, and low LSM and low support score.

For each community category, we then create two language models using posts from the respective groups described above. We also create another model of all the posts in the same community, and then use this as our base model for extracting the frequent vocabulary used in the two groups. To do so, we use SAGE [34] which is a generative model of text where each (latent) class label is endowed with a model of the deviation in log-frequency from a background distribution. Table 2 shows the frequently used words in the posts in the aforementioned sets of posts—high LSM, high ES/IS and low LSM, low ES/IS. Most frequent words in all the sets are the ones that are found to be more relevant to its community category, that is, more aligned with the community's goals and norms.

For example, in case of IS, the set of posts from Psychosis & Anxiety related communities with high LSM and high IS show frequent usage of medical condition related words like 'blood', 'nausea' and 'pain', along with medications like 'xanax'. E.g., consider a (paraphrased) post excerpt which talks about the use of 'xanax': "*I still have Xanax prescription, it makes me feel better. [...] it's good to feel normal again.*". This post received an LSM score of 0.75 and IS of 3. One of its comments advised the poster to use the medication moderately, "*Xanax are great if you're sensible with them...*". A post from the same community which got a low LSM score of 0.45 talks about an Apple Watch that the poster got after getting a medical procedure done. This post received an IS of 1.

Similarly, in a Mood Disorders related community, a post described an individual's struggle with depression after their spouse left them, *"...I'm still in love with him and am having a hard time coping with him leaving..."*. This post also received an LSM score of 0.79 and had 55 comments on it. The mean ES for this post was 3 and most of the comments were sympathetic and were targeted to provide encouragement to the poster. E.g., one of the high ES comments on this post said *"...I'm very sorry to hear you're going through this! Please try to not feel lonely...."*. Thus, the posts that have high linguistic alignment and high ES use the words that are found to be relevant to the community in which the post was made.

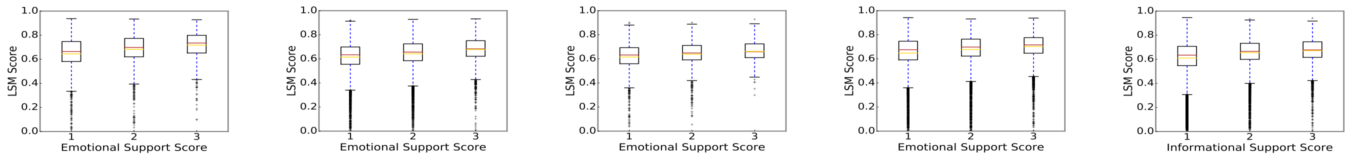
### Nested Multinomial Logistic Regression

From our analyses so far, we see the evidence of a positive relationship between both ES as well as IS and a post's LSM score. Can we quantify this relationship in a principled, statistical manner, and is this relationship significant? To do so, we now present the results of our set of nested multinomial regression models (ref. Methods), that aim to utilize LSM to predict the response variable, ES or IS of a post.

Note that, to ensure that the predictor variables of our regression models are linearly independent of each other, we perform multicollinearity test. We obtain correlation matrix by performing Pearson's correlation on all the variables in LSM Model and then compute the Eigen vector of this correlation matrix. We find that the individual eigenvalues in that matrix are closer to 1 than to 0, which means that our predictor variables are indeed linearly independent of each other.

**Goodness of Fit.** In Table 3, we report several goodness of fit measures for these sets of models – one each for IS and ES – that estimate the support received by the posts in our dataset for each of the community categories. The first model, or the Control Model, includes all the control variables explained earlier in Methods. In addition to these variables, in a second model, referred to as the LSM Model, we additionally use LSM score as an extra variable.

Using Table 3, we first evaluate the goodness of fits for the Control and LSM multinomial logistic regression models based on the deviance metric. As deviance is the measure of lack of fit to the data, lower values are better. We observe that, in all community categories, compared to the Null Model, both of our models (Control, LSM) provide considerable explanatory power with significant improvements in deviances.




(a) Trauma & Abuse (b) Psychosis & Anxiety (c) Compulsive Disorders (d) Coping & Therapy (e) Mood Disorders

(f) Trauma & Abuse (g) Psychosis & Anxiety (h) Compulsive Disorders (i) Coping & Therapy (j) Mood Disorders

Figure 6: Box plots showing the relationship between LSM and ES (top) and IS (bottom) in all community categories.

Next, comparing the Control Model and the LSM model mutually, we examine whether the addition of the LSM variable in the latter resulted in any improvement in model fit. We find that the difference between the deviance of the Control model and the LSM model approximately follows a $\chi^2$ distribution with degrees of freedom (*df*) equal to twice the total number of additional parameters in the LSM model. For example, in case of IS in the Trauma & Abuse community category, comparing the deviance of the LSM model with that of Control model, we see that the additional information provided by LSM score has significant explanatory power: $\chi^2(2, N = 5602)$ = 11097.4 - 11044.8 = 52.6, $p < 10^{-12}$. Similarly, in the case of ES in the same (Trauma & Abuse) community category, the difference between the deviance of the Control and LSM models also approximates a $\chi^2$ distribution: $\chi^2(2, N = 5602)$ = 11447 - 11404.8 = 42.2, $p < 10^{-12}$. We observe similar results in case of other community categories for both support types.

Finally, note that goodness of fit ($\chi^2$) is the highest (=717, $p < 10^{-15}$) for Mood Disorders related communities, whereas the lowest for Compulsive Disorders related community category (=11.6, $p < 10^{-3}$), both in the case of IS. This indicates, that in the Mood Disorder communities LSM is highly predictive of IS; in the Compulsive Disorder ones, it is the least.

**Quantifying the Role of LSM.** Finally, we present the logistic regression coefficients for the LSM model for both IS and ES: see Table 4. Here, each coefficient is the estimated change in the *log* of odds ratios of having support score as 2 or 3 for a unit increase in LSM score holding the variables from the Control Model constant at a certain value. For example, the *log* odds for one unit increase in LSM score for IS = 1, in case of the community category Trauma & Abuse ($C_1$), is 1.75. This means that the multinomial *log* odds of IS being 2 over 1, is expected to increase by 1.75 with a unit increase in LSM score, while holding the control variables in LSM model constant. Similar patterns are noted in case of other community categories for both ES as well as IS LSM models.

The corresponding odds ratios are also shown in Table 4. Odds ratios are exponentiation of the coefficients. Here, the odds ratio of coefficient value of LSM indicates how the risk of the support score having value 2 or 3 compared to the risk of the support score being 1, changes with the change in LSM score.

| Community category (# obs.) | | *Null* | *Control* | *LSM* |
|---|---|---|---|---|
| *Informational Support (IS)* | | | | |
| | df | 0 | 88 | 90 |
| | deviance | 11543 | 11097.4 | 11044.8 |
| $C_1$ (5,606) | $\chi^2$ | | 445.6*** | 52.6*** |
| | deviance | 72552 | 69186 | 68880 |
| $C_2$ (35,348) | $\chi^2$ | | 3366*** | 306*** |
| | deviance | 8207.6 | 7811 | 7799.4 |
| $C_3$ (4,016) | $\chi^2$ | | 396.6*** | 11.6*** |
| | deviance | 83290 | 79622 | 79358 |
| $C_4$ (50,124) | $\chi^2$ | | 3668*** | 264*** |
| | deviance | 146544 | 140386 | 139672 |
| $C_5$ (84,110) | $\chi^2$ | | 6158*** | 714*** |
| *Emotional Support (ES)* | | | | |
| | df | 0 | 88 | 90 |
| | deviance | 12043.8 | 11457.2 | 11415.8 |
| $C_1$ (5,606) | $\chi^2$ | | 586.6*** | 41.4*** |
| | deviance | 72098 | 68456 | 68334 |
| $C_2$ (35,348) | $\chi^2$ | | 3642*** | 122*** |
| | deviance | 7556.6 | 7260.8 | 7242.8 |
| $C_3$ (4,016) | $\chi^2$ | | 295.8*** | 18*** |
| | deviance | 108102 | 104486 | 104218 |
| $C_4$ (50,124) | $\chi^2$ | | 3616*** | 268*** |
| | deviance | 181788 | 175434 | 175028 |
| $C_5$ (84,110) | $\chi^2$ | | 6354*** | 406*** |

Table 3: Goodness of fit statistics comparing the Null, Control and LSM models for IS and ES along with their statistical significance (*** $p < 0.001$) for all five community categories.

Thus, odds ratio greater than 1 indicates that the risk of the support score being 2 or 3 relative to the risk of the support score being 1 increases as LSM increases. As shown in Table 4, the odds ratios for support scores = 2 and 3 for both IS as well as ES are greater than 1, which means that with increase in LSM score, the support received by a post increases.

We also note that across the five community categories, the odds ratios in the case of IS are higher than in case of ES which means that the chances of getting more support with the increase in LSM score is higher in case of IS than in ES. For example, we find that in case of Trauma & Abuse ($C_1$) communities, the relative odds ratios for support score = 3 are higher in case of IS than ES, with a difference of 3.71 (19.72 - 16.01). Whereas, in case of Compulsive Disorders ($C_3$) related communities, the relative odds ratios for support score = 3 are

| | Informational Support (IS) | | | | Emotional Support (ES) | | | |
|---|---|---|---|---|---|---|---|---|
| | $\beta$ Weights | | Odds Ratios | | $\beta$ Weights | | Odds Ratios | |
| | $y=2$ | $y=3$ | $y=2$ | $y=3$ | $y=2$ | $y=3$ | $y=2$ | $y=3$ |
| $C_1$ | 1.75 | 2.98 | 5.74 | 19.72 | 1.29 | 2.77 | 3.64 | 16.01 |
| $C_2$ | 1.642 | 2.85 | 5.15 | 17.33 | 0.83 | 1.99 | 2.28 | 7.29 |
| $C_3$ | 0.71 | 1.82 | 2.04 | 6.18 | 0.89 | 2.79 | 2.45 | 16.35 |
| $C_4$ | 1.68 | 2.17 | 5.39 | 8.75 | 0.93 | 2.26 | 2.52 | 9.53 |
| $C_5$ | 1.80 | 2.92 | 6.07 | 18.48 | 0.73 | 2.01 | 2.08 | 7.43 |

Table 4: Multinomial logistic regression coefficients ($\beta$ weights) of predictor variable LSM score for LSM model. Low support score (=1) is the referent group for both IS and ES regression models.

lower in case of IS by a value of 10.17 (16.35 - 6.18). This means that with an increase in linguistic accommodation, it is easier to get higher IS in Trauma & Abuse communities, which predominantly provides higher ES. On the other hand, it is easier to get higher ES in case of Compulsive Disorders communities, which provides higher IS (Figure 4).

## DISCUSSION

Our work has provided some of the first insights examining social support and linguistic accommodation in Reddit OMHCs serving a variety of psychological needs. These findings allow us to identify several implications for these communities, from a theoretical and a design perspective. We also reflect on what our results mean for the operation and functioning of OMHCs.

### Theoretical Implications

**Support Matching and Provisioning.** A notable finding of our work is that although all community categories offer considerable ES and IS, certain communities direct more ES (e.g., Mood Disorder communities), while some others provide more IS (e.g., Compulsive Disorder communities). This indicates that in spite of all these communities being related to mental health they cater to different social support goals. Our results find validation in the Optimal Matching Theory [22], which argues that social support is a multi-dimensional construct and certain types of support may be more effective when matched with certain requests. In fact, Cutrona and Russell [22] identified that coping with a mental illness like depression may need more ES which explains why Mood Disorder communities, that span subreddits like *r/depression*, offer more ES. Whereas, since the posters in the Compulsive Disorder communities like *r/OCD* tend to seek advice around their condition, IS serves as a better match to their posts. In short, this indicates that support provisions in the different OMHCs aligns with the particular topics they focus on.

**Conformance and Support.** The most notable finding of our work is that social support in OMHCs, whether ES or IS, is positively associated with linguistic accommodation. Recall that the sociolinguistic theories suggest the existence of a positive link between conformance to linguistic norms and social feedback in general purpose offline/online groups [41]. We find this to hold true for online communities catering to mental health topics as well. Our results thus enable us to refine our sociolinguistic understanding of the role played by the deeply ingrained social processes in these specialized communities. They illustrate that, given the stigmatized and sensitive nature of issues dealt by these communities, its members develop linguistic conventions so that the community as a whole can serve as a safe place for candid disclosures, while also providing support to vulnerable individuals. The necessity of linguistic alignment to receive support in these communities may also be attributed to the fact that they intend to promote the establishment and maintenance of socially cohesive relationships grounded in trust, rapport and empathy. Prior work has noted these qualities to be fundamental to OMHCs [2], and normative compliance vital to achieving them [52, 68].

### The Tension Between Accommodation and Support

Despite the above theoretical implications, the finding that both ES and IS have a positive relationship with linguistic accommodation in OMHCs also surfaces a *tension*: whether to have support seekers adhere to the community norms to maintain a coherent identity, or to maintain the communities' goals of directing timely support to individuals expressing vulnerability. Multiple media reports have acclaimed Reddit and similar OMHCs for their ability to help distressed individuals expressing unique and urgent needs [61, 65, 5], as also captured in the quote in the beginning of the paper. One of them notes: "[..] *online communities taking the place of IRL therapists for helping people deal with their mental health issues. Even the darkest places, like Reddit, can surprise you with its displays of humanity*" [65]. However, to take an example, we note two posts in *r/SuicideWatch*, from our dataset, that expressed similar immediacy and criticality of needs, but received unequal levels of support: "*I think I'm gonna tell the people who gave up on me goodbye and then end it there*" (ES=2; IS=1), and "*I don't even know why I'm posting this, I just don't think I can do it anymore. I have no desire to live*" (ES=1; IS=1). Based on our analysis, we found that the former post that received greater ES exhibited higher accommodation than the latter (LSM=0.73 versus LSM=0.48).

Our observations regarding these OMHCs, therefore, raise a few questions: Should linguistic alignment be an absolute necessity to receive help, even if somebody expresses critical needs, such as risk of self-injury? Should less conforming posters not receive sufficient support even if they genuinely require immediate attention? We note that prior literature has acknowledged the challenges individuals with mental illnesses face during social exchanges [20]. Detachment from the social realm is a known attribute of distressed individuals [72]. Feeling of lack of autonomy and control over one's life and a "paralysis" in decision-making around word and linguistic choices are also established to be associated with many mental illnesses like anxiety [24, 73]. Consequently, socially determined norms and responsibilities place these individuals in situations where they have little control over decisions concerning their lives [85]. Aaron Beck's cognitive model of emotional disorders further states that individuals with mental illnesses can allocate limited cognitive and attentional resources towards the norms of their social context [7].

Nevertheless, our results *do* show that many support seekers are indeed managing to linguistically align with the norms of the OMHCs. As noted above, we also recognize that linguistic accommodation can have its own benefits for developing a holistic community identity and building trust and rapport. However, to democratize support provisions for all and to navigate the tension noted above, we posit that making linguistic normative compliance a *requirement*, implicitly or explicitly, may contradict the broader purposes of OMHCs. Instead, it would serve OMHC members better if they make their platforms more conducive by allowing non-conforming support seekers gain explicit awareness of the underling norms, or equipping support providers with better tools to easily direct help to these individuals with critical psychological needs.

### Design Implications

Continuing the above discussion, we propose two design solutions that target the question: How can OMHCs cater to the needs of the non-linguistically aligning support seekers who typically end up harnessing limited support benefits?

**Influencing Community Design.** Currently many social media platforms, including Reddit, include design affordances to help communities (subreddits) highlight their rules and guidelines. For example, the sidebar pane of a subreddit's landing page is often used in OMHCs to inform members about the purpose of that community along with the type of content that is permissible to be shared in that community. However, these features do not provide a simple way for support seekers, often already challenged by critical and distressful situations, to identify the specific normative behaviors that are typically associated with quality ES or IS in a specific community.

We suggest that community designers can make an OMHC's norms more apparent and built into the interface design, specifically to facilitate the subconscious learning of linguistic style of the community,thereby improving chances to receive more support. We propose implementing a tool that will first employ our statistical models to identify historical posts with both high linguistic accommodation and high ES or IS. Thereafter, within community guidelines or Frequently Asked Questions (FAQs), such as the side pane on a subreddit page, the tool can include (paraphrased and de-identified) examples of these identified posts. An alternate design can also promote such examples at the top of the landing page of the communities—on Reddit this is enabled by a feature known as "sticky comments" [70]. Due to the computational nature of how the examples are identified, the displayed information on the landing/FAQ pages can be periodically updated to reflect the evolving norms of the OMHCs.

**Assisting Support Providers.** Our second set of design suggestions focus on the providers of support in OMHCs. Recall that our findings imply that posters whose linguistic style does not match with the rest of the community, are likely to receive less support. However, as noted above, there could be situations where the poster is in urgent need of support and attention, such as in communities like *r/SuicideWatch*. We believe in these cases, OMHCs need to include design capabilities that can "override" the implicit or explicit need for greater linguistic accommodation in receiving adequate social support. New tools are also needed to help support providers efficiently and quickly navigate the stream of incoming requests, noting their conformance to the community's norms, while also being mindful of the criticality of the requests.

We suggest the development of intervention tools as a solution, that can issue timely alerts to active support providers and moderators about posts with low linguistic accommodation (assessed with our LSM approach) and which have received little support so far (assessed by our support classifier). The support providers can then reach out to these posters in a prompt fashion for directing and allocating appropriate resources. Further, to prevent a deluge of such alerts from overwhelming the support providers, these intervention tools can also incorporate visual summaries and interactive interfaces with the content of the posts [53], where posts are categorized based on the issues they discuss and the urgency they manifest. This can help support providers quickly spot the pressing issues the poster seeks to discuss, in spite of any stylistic discrepancies; making it easier for them to provide timely and helpful feedback.

### Limitations and Future Work

We recognize some limitations to our work. First, we recognize that the three support classes for ES and IS are likely to have subtle differences among them, and while our support classifier showed satisfactory performance and agreed well with expert annotations, classifier performance may improve with greater availability of labeled data. Next, given our focus on mental health, we adopted a specific measure linguistic accommodation: the LSM measure. Although psycholinguistically motivated, future work could examine how our findings hold under alternative measures of linguistic conformance.

Importantly, there could also be multiple factors, beyond linguistic accommodation that could impact whether a post shared in an OMHC receives adequate support. We identified and controlled for a number of such factors in our statistical models. However, we were limited by those which can be directly or indirectly measured from our data. On a related note, we suggest caution in deriving causal interpretations from our results. Although we found a positive link between linguistic accommodation and support in the communities we studied, we cannot be certain if this is causal in nature.

## CONCLUSION

In this paper, we presented a comprehensive study examining the relationship between linguistic accommodation and social support in online mental health communities. Employing a large dataset of 55 Reddit communities that focus on a variety of mental health topics, we first quantitatively derived measures for two kinds of social support, emotional and informational, as received by posts shared in these communities. Then we measured linguistic accommodation exhibited in them, based on a psycholinguistic measure. Our results showed that there is a significant positive association between linguistic accommodation and both the types of support, consistent across the communities we studied. Based on these findings, we note a tension between the vitality of conformance to a community's norms and the goals of support providers. Our work bears implications for the design tools that can help improve online support provisioning mechanisms.

## REFERENCES

1. Nazanin Andalibi, Oliver L Haimson, Munmun De Choudhury, and Andrea Forte. 2016. Understanding social media disclosures of sexual abuse through the lenses of support seeking and anonymity. In *CHI*. ACM, 3906–3918.

2. Nazanin Andalibi, Pinar Öztürk, and Andrea Forte. 2017. Sensitive Self-disclosures, Responses, and Social Support on Instagram: The Case of# Depression.. In *CSCW*. 1485–1500.
3. American Psychiatric Association and others. 2013. *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.
4. Antonina Bambina. 2007. *Online social support: the interplay of social networks and computer-mediated communication*. Cambria press.
5. Tanya Basu. 2016. What it's like to moderate Reddit's Suicide Watch. `http://www.cnn.com/2016/02/25/health/reddit-suicide-watch/index.html`. (2016). [Online; accessed 16-September-2017].
6. Christopher E Beaudoin and Chen-Chao Tao. 2007. Benefiting from social capital in online support groups: An empirical study of cancer patients. *CyberPsychology & Behavior* 10, 4 (2007), 587–590.
7. Aaron T Beck. 1979. *Cognitive therapy and the emotional disorders*. Penguin.
8. Barbara Becker and Gloria Mark. 2002. Social conventions in computermediated communication: A comparison of three online shared virtual environments. In *The social life of avatars*. Springer, 19–39.
9. E Berscheid and EH Walster. 1969. Rewards others provide: Similarity. *Interpersonal attraction* (1969), 69–91.
10. Michael Biggs and Daniela Büchler. 2011. Communities, values, conventions and actions. *The Routledge companion to research in the arts* (2011), 82–98.
11. Google BigQuery. 2015. Google BigQuery Reddit Data Archive. `https://bigquery.cloud.google.com/table/fh-bigquery:redditposts.fullcorpus201512?pli=1`. (2015). [Online; accessed 14-September-2017].
12. Prakhar Biyani, Cornelia Caragea, Prasenjit Mitra, and John Yen. 2014. Identifying Emotional and Informational Support in Online Health Communities.. In *COLING*. 827–836.
13. Pierre Bourdieu. 1990. *The logic of practice*. Stanford University Press.
14. Brant R Burleson, Erina L MacGeorge, ML Knapp, and JA Daly. 2002. Supportive communication. *Handbook of interpersonal communication* 3 (2002), 374–424.
15. Donn Erwin Byrne. 1971. *The attraction paradigm*. Vol. 11. Academic Pr.
16. R Sherlock Campbell and James W Pennebaker. 2003. The secret life of pronouns flexibility in writing style and physical health. *Psychological science* 14, 1 (2003), 60–65.
17. Scott E Caplan and Jacob S Turner. 2007. Bringing theory to research on computer-mediated comforting communication. *Computers in human behavior* 23, 2 (2007), 985–998.
18. Cindy Chung and James W Pennebaker. 2007. The psychological functions of function words. *Social communication* (2007), 343–359.
19. Jae Eun Chung. 2013. Social interaction in online support groups: Preference for online social interaction over offline social interaction. *Computers in Human Behavior* 29, 4 (2013), 1408–1414.
20. Sheldon Cohen and Thomas A Wills. 1985. Stress, social support, and the buffering hypothesis. *Psychological bulletin* 98, 2 (1985), 310.
21. Karen S Cook, Coye Cheshire, Eric RW Rice, and Sandra Nakagawa. 2013. Social exchange theory. In *Handbook of social psychology*. Springer, 61–88.
22. Carolyn E Cutrona and Daniel W Russell. 1990. Type of social support and specific stress: Toward a theory of optimal matching. *Social support: An interactional view* (1990), 319–366.
23. Carolyn E Cutrona and Julie A Suhr. 1994. Social support communication in the context of marriage: an analysis of couples' supportive interactions. *Communication of social support: Messages, interactions, relationships, and community* (1994), 113–135.
24. Odd Steffen Dalgard, K Tambs, and others. 1995. Social support, negative life events and mental health. *The British Journal of Psychiatry* 166, 1 (1995), 29–34.
25. Cristian Danescu-Niculescu-Mizil, Michael Gamon, and Susan Dumais. 2011. Mark my words!: linguistic style accommodation in social media. In *WWW*. ACM, 745–754.
26. Cristian Danescu-Niculescu-Mizil and Lillian Lee. 2011. Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs. In *Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics*. Association for Computational Linguistics, 76–87.
27. Cristian Danescu-Niculescu-Mizil, Lillian Lee, Bo Pang, and Jon Kleinberg. 2012. Echoes of power: Language effects and power differences in social interaction. In *WWW*. ACM, 699–708.
28. Cristian Danescu-Niculescu-Mizil, Robert West, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013. No country for old members: User lifecycle and linguistic change in online communities. In *WWW*. ACM, 307–318.
29. Kathryn P Davison, James W Pennebaker, and Sally S Dickerson. 2000. Who talks? The social psychology of illness support groups. *American Psychologist* 55, 2 (2000), 205.
30. Munmun De Choudhury and Sushovan De. 2014. Mental Health Discourse on reddit: Self-Disclosure, Social Support, and Anonymity. In *ICWSM*.

31. Munmun De Choudhury and Emre Kıcıman. 2017. The language of social support in social media and its effect on suicidal ideation risk. In *ICWSM*, Vol. 2017. NIH Public Access, 32.
32. Judith S Donath and others. 1999. Identity and deception in the virtual community. *Communities in cyberspace* 1996 (1999), 29–59.
33. Christine Dunkel-Schetter. 1984. Social support and cancer: Findings based on patient interviews and their implications. *Journal of Social issues* 40, 4 (1984), 77–98.
34. Jacob Eisenstein, Amr Ahmed, and Eric P Xing. 2011. Sparse Additive Generative Models of Text. In *ICML*. 1041–1048.
35. Richard M Emerson. 1976. Social exchange theory. *Annual review of sociology* 2, 1 (1976), 335–362.
36. Gunther Eysenbach, John Powell, Marina Englesakis, Carlos Rizo, and Anita Stern. 2004. Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. *Bmj* 328, 7449 (2004), 1166.
37. Daniel C Feldman. 1984. The development and enforcement of group norms. *Academy of management review* 9, 1 (1984), 47–53.
38. Sérgio Fernandes and Jorge Bernardino. 2015. What is bigquery?. In *Proceedings of the 19th International Database Engineering & Applications Symposium*. ACM, 202–203.
39. Kathleen Ferrara. 1991. Accommodation in therapy. *Contexts of accommodation: Developments in applied sociolinguistics* (1991), 187–222.
40. Matt Garley and Julia Hockenmaier. 2012. Beefmoves: dissemination, diversity, and dynamics of English borrowings in a German hip hop forum. In *Proceedings of the 50th Annual Meeting of the ACL*. Association for Computational Linguistics, 135–139.
41. Howard Giles and Peter Powesland. 1997. Accommodation theory. In *Sociolinguistics*. Springer, 232–239.
42. Howard Giles, Donald M Taylor, and Richard Bourhis. 1973. Towards a theory of interpersonal accommodation through language: Some Canadian data. *Language in society* 2, 02 (1973), 177–192.
43. Erving Goffman. 2009. *Stigma: Notes on the management of spoiled identity*. Simon and Schuster.
44. Amy L Gonzales, Jeffrey T Hancock, and James W Pennebaker. 2009. Language style matching as a predictor of social dynamics in small groups. *Communication Research* (2009).
45. Amy L Gonzales, Jeffrey T Hancock, and James W Pennebaker. 2010. Language style matching as a predictor of social dynamics in small groups. *Communication Research* 37, 1 (2010), 3–19.
46. Vicki S Helgeson, Sheldon Cohen, Richard Schulz, and Joyce Yasko. 2001. Long-term effects of educational and peer discussion group interventions on adjustment to breast cancer. *Health Psychology* 20, 5 (2001), 387.
47. Jina Huh. 2015. Clinical questions in online health communities: the case of see your doctor threads. In *Proceedings of the 18th ACM Conference on CSCW & Social Computing*. ACM, 1488–1499.
48. Molly E Ireland, Richard B Slatcher, Paul W Eastwick, Lauren E Scissors, Eli J Finkel, and James W Pennebaker. 2011. Language style matching predicts relationship initiation and stability. *Psychological science* 22, 1 (2011), 39–44.
49. Berton H Kaplan, John C Cassel, and Susan Gore. 1977. Social support and health. *Medical care* 15, 5 (1977), 47–58.
50. Ichiro Kawachi and Lisa F Berkman. 2001. Social ties and mental health. *Journal of Urban health* 78, 3 (2001), 458–467.
51. Ronald C Kessler and Jane D McLeod. 1985. *Social support and mental health in community samples.* Academic Press.
52. Sara Kiesler, Robert Kraut, Paul Resnick, and Aniket Kittur. 2012. Regulating behavior in online communities. *Building Successful Online Communities: Evidence-Based Social Design. MIT Press, Cambridge, MA* (2012).
53. Bum Chul Kwon, Sung-Hee Kim, Sukwon Lee, Jaegul Choo, Jina Huh, and Ji Soo Yi. 2016. VisOHC: Designing visual analytics for online health communities. *IEEE transactions on visualization and computer graphics* 22, 1 (2016), 71–80.
54. Wan Shun Eva Lam. 2008. Language socialization in online communities. In *Encyclopedia of language and education*. Springer, 2859–2869.
55. Aideen Lawlor and Jurek Kirakowski. 2014. Online support groups for mental health: A space for challenging self-stigma or a means of social avoidance? *Computers in Human Behavior* 32 (2014), 152–161.
56. Alex Leavitt. 2015. This is a throwaway account: Temporary technical identities and perceptions of anonymity in a massive online community. In *CSCW*. ACM, 317–327.
57. Nan Lin, Mary W Woelfel, and Stephen C Light. 1985. The buffering effect of social support subsequent to an important life event. *JHSB* (1985), 247–263.
58. Marifran Mattson and Jennifer Gibb Hall. 2011. *Health as communication nexus: A service-learning approach*.
59. John McIver and Edward G Carmines. 1981. *Unidimensional scaling*. Vol. 24. Sage.
60. Reddit Metrics. 2015. Reddit Metrics. `http://redditmetrics.com/`. (2015). [Online; accessed 15-September-2016].

61. Felicity Morse. 2016. Social media and suicide: What it's like being a moderator on r/SuicideWatch. http://www.bbc.co.uk/newsbeat/article/35577626/social-media-and-suicide-what-its-like-being-a-\moderator-on-rsuicidewatch. (2016). [Online; accessed 16-September-2017].

62. JA Naslund, KA Aschbrenner, LA Marsch, and SJ Bartels. 2016. The future of mental health care: peer-to-peer support and social media. *Epidemiology and psychiatric sciences* 25, 02 (2016), 113–122.

63. Mark W Newman, Debra Lauterbach, Sean A Munson, Paul Resnick, and Margaret E Morris. 2011. It's not that i don't have problems, i'm just not putting them on facebook: challenges and opportunities in using online social networks for health. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*. ACM, 341–350.

64. Dong Nguyen and Carolyn P Rosé. 2011. Language use as a reflection of socialization in online communities. In *Proceedings of the Workshop on Languages in Social Media*. ACLs, 76–85.

65. Arielle Pardes. 2016. How Reddit Became The Internet's Therapist. https://health.good.is/features/internet-therapy-reddit. (2016). [Online; accessed 16-September-2017].

66. Albert Park, Andrea L Hartzler, Jina Huh, David W McDonald, and Wanda Pratt. 2015. Homophily of vocabulary usage: beneficial effects of vocabulary similarity on online health communities participation. In *AMIA Annual Symposium Proceedings*, Vol. 2015. American Medical Informatics Association, 1024.

67. Tom Postmes, Russell Spears, and Martin Lea. 2000. The formation of group norms in computer-mediated communication. *Human communication research* 26, 3 (2000), 341–371.

68. Jenny Preece. 2000. *Online communities: Designing usability and supporting socialbilty*. John Wiley & Sons, Inc.

69. Jenny Preece, Diane Maloney-Krichmar, and Chadia Abras. 2003. History of emergence of online communities. *Encyclopedia of Community. Berkshire Publishing Group, Sage* (2003), 1–11.

70. r/modnews. 2016. Moderators: Sticky Comments is now available to all mods. https://www.reddit.com/r/modnews/comments/3x8vzl/moderators_sticky_comments_is_now_available_to/?st=j7qv8sla&sh=b5bd21b6. (2016). [Online; accessed 16-September-2017].

71. Catherine Schaefer, James C Coyne, and Richard S Lazarus. 1981. The health-related functions of social support. *Journal of behavioral medicine* 4, 4 (1981), 381–406.

72. Thomas J Scheff. 1967. *Mental illness and social processes*. Harper & Row.

73. Hannah R Snyder, Natalie Hutchison, Erika Nyhus, Tim Curran, Marie T Banich, Randall C O'Reilly, and Yuko Munakata. 2010. Neural inhibition enables selection during language processing. *Proceedings of the National Academy of Sciences* 107, 38 (2010), 16483–16488.

74. Claire F Sullivan. 1996. Recipients' perceptions of support attempts across various stressful life events. *Communication Research Reports* 13, 2 (1996), 183–190.

75. Yla R Tausczik and James W Pennebaker. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *JLS* 29, 1 (2010), 24–54.

76. Paul J Taylor and Sally Thomas. 2008. Linguistic style matching and negotiation outcome. *NCMR* 1, 3 (2008), 263–281.

77. Catalina L Toma. 2014. Towards conceptual convergence: An examination of interpersonal adaptation. *Communication Quarterly* 62, 2 (2014), 155–178.

78. R Jay Turner, B Gail Frankel, and Deborah M Levin. 1983. Social support: Conceptualization, measurement, and implications for mental health. *Research in community & mental health* (1983).

79. Tatiana A Vlahovic, Yi-Chia Wang, Robert E Kraut, and John M Levine. 2014. Support matching and satisfaction in an online breast cancer support community. In *CHI*. ACM, 1625–1634.

80. Joseph B Walther and Shawn Boyd. 2002. Attraction to computer-mediated social support. *Communication technology and society: Audience adoption and uses* 153188 (2002).

81. Yafei Wang, John Yen, and David Reitter. 2015. Pragmatic alignment on social support type in health forum conversations. *Proc. of CMCL* (2015), 9–18.

82. Yi-Chia Wang, Moira Burke, and Robert Kraut. 2016. Modeling self-disclosure in social networking sites. In *CSCW*. ACM, 74–85.

83. Yi-Chia Wang, Robert Kraut, and John M Levine. 2012. To stay or leave?: the relationship of emotional and informational support to commitment in online health support groups. In *CSCW*. ACM, 833–842.

84. Steve Whittaker, Ellen Isaacs, and Vicki O'Day. 1997. Widening the net: workshop report on the theory and practice of physical and network communities. *SIGCHI BULLETIN* 29 (1997), 27–30.

85. J Mark G Williams, Fraser N Watts, Colin MacLeod, and Andrew Mathews. 1988. *Cognitive psychology and emotional disorders.* John Wiley & Sons.

86. Samuel M Wilson and Leighton C Peterson. 2002. The anthropology of online communities. *Annual review of anthropology* 31, 1 (2002), 449–467.

87. Diyi Yang, Robert Kraut, and John M Levine. 2017. Commitment of Newcomers and Old-timers to Online Health Support Communities. In *Proceedings of the 2017 CHI*. ACM, 6363–6375.