Understanding Mechanisms of Emotional Self-Disclosure and Social Support Among Undergraduates in Online Spaces

Ph.D. Proposal

Talie Massachi
Department of Computer Science
Brown University

talie_massachi@brown.edu
website

Version 1.0.0

December 18, 2023

Thesis Committee:

Jeff Huang, Brown University
Ellie Pavlick, Brown University
Nicole Nugent, Brown University

Contents

List of Figures **6**

List of Tables **7**

1 Introduction **1**
1.1 Motivation . . . 1
1.2 Overview of Contributions . . . 2
1.3 Proposed Work and Research Goals . . . 3
1.3.1 Emoji Understanding and Interpretation with Limited Context . . . 3
1.3.2 The Impact of Written Disclosure in Semi-Private Online Spaces on Perceived Social Support . . . 4
1.3.3 The Impact of Social Presence When Disclosing in Online Spaces on Perceived Social Support . . . 4

2 Related Work **5**
2.1 The Benefits of Online Social Support . . . 5
2.1.1 Measures of Social Support . . . 6
2.2 Online Communication Over Private Platforms . . . 6

3 Preliminary Work: Offline to Online Social Support during the COVID-19 Pandemic **7**
3.1 Introduction . . . 8
3.2 Related Work . . . 9
3.2.1 The State of Mental Health During the COVID-19 Pandemic . . . 9
3.2.2 Online Communication as a Coping Mechanism During the COVID-19 Pandemic 9
3.3 Methods . . . 11
3.3.1 Survey Design . . . 12
3.3.2 Survey Questions . . . 12
3.3.3 Participant Recruitment . . . 14
3.4 Results . . . 15
3.4.1 Impact of Physical Isolation on College Students . . . 16
3.4.2 Changing Trends in Online Platform Usage . . . 18
3.4.3 Patterns in Self-Disclosure in Private Online Interactions . . . 18
3.4.4 Main Takeaways . . . 20
3.5 Discussion . . . 21
3.5.1 Negative Relationship Between Public Posting and Well-Being . . . 21
3.5.2 Benefits of Private Communication . . . 22

3.5.3 Divergent Effects of Online Communication on Loneliness and Perceived Social Support . . . 24
3.6 Limitations . . . 24
3.7 Conclusion . . . 25

4 Preliminary Work: Signals of Affect in Messaging Data 27
4.1 Introduction . . . 28
4.2 Related Work . . . 29
4.2.1 Self-Report Measurements of Affect . . . 29
4.2.2 Sentiment Analysis Techniques for Social Media Text . . . 29
4.2.3 Predicting PANAS using Automated Sentiment Analysis . . . 30
4.2.4 Personal Disclosure Across Multiple Social Media Platforms . . . 31
4.3 Sochiatrist System . . . 32
4.3.1 Data Extraction Methods . . . 32
4.3.2 Privacy and Consent . . . 33
4.4 Methods . . . 34
4.4.1 Study Procedure . . . 35
4.4.2 Data Processing and Session Generation . . . 36
4.4.3 Automated Sentiment Analysis . . . 38
4.4.4 Human Review Process . . . 38
4.5 Results . . . 40
4.5.1 Individual Performance of Techniques . . . 41
4.5.2 Comparative Performance Between Techniques . . . 43
4.5.3 Mispredictions Across All Techniques . . . 44
4.6 Discussion . . . 46
4.6.1 Reflection on Affect Predictions . . . 46
4.6.2 Ethical Considerations . . . 47
4.6.3 Limitations . . . 48
4.7 Conclusion . . . 49

5 Preliminary Work: Chirp, A Platform to Investigate Self-Disclosure and Social Support in an Anonymous Space 50
5.1 Introduction . . . 51
5.2 Existing Disclosure Systems . . . 51
5.3 Developing The Chirp Application To Study Self-Disclosure in Online Spaces . . . 52
5.3.1 Design Considerations . . . 52
5.3.2 App Interface Design . . . 55
5.3.3 Privacy and Consent . . . 56
5.3.4 Data Flow for Account Creation and Recording Data . . . 56

6 Proposed Work 57
6.1 Preliminary Work Contributions . . . 57
6.1.1 Proposed Additional Contributions . . . 57
6.2 The Impact of Demographic Context on Understanding in Emotional Self-Disclosure Through Emojis . . . 57
6.2.1 Proposed Approach . . . 58
6.2.2 Proposed Analysis . . . 60
6.3 Chirp: The Impact of Private Online Self-Disclosure on Perceived Social Support . . . 61

6.3.1 Proposed Approach 61
6.3.2 Design Considerations 63
6.3.3 Proposed Analysis 64
6.4 Timeline 64

Bibliography **65**

List of Figures

Figure 3.1 The average self-reported online interaction frequency was compared before and during the COVID-19 pandemic. The average hours spent interacting on each platform per week (left), as well as frequency of self-disclosure (right) on each platform was calculated. College students saw an increase in usage across all platforms, with the largest being social media. A greater increase was found in frequency of self-disclosure across these platforms, and students on average spend more time disclosing over one-on-one platforms instead of group platforms, with the most being one-on-one messaging. 17

Figure 3.2 Data is grouped by those disclosing over messaging platforms at least weekly (frequent) or those disclosing monthly or less (infrequent). The specific emotions shared are plotted, revealing significant differences between groups across all emotions, with greater increases for more negative emotions (depression, anxiety, fear), with the exception of anger. It appears that those who are more willing to disclose are specifically referring to negative emotions, not disclosure of positive emotions. It appears that students who share negative emotions more frequently may perceive added benefits to social support. 19

Figure 4.1 Sochiatrist extracts, consolidates, and pseudonymizes the data to develop models predicting affect based on messaging data. 32

Figure 4.2 Histogram for PANAS(–) (pink) compared to the histogram for PANAS(+) (blue), with the overlap in purple. Note that PANAS(–) has a lower mean than PANAS(+) and is a more skewed distribution. 37

Figure 4.3 Self-reported PANAS(+) and PANAS(–) scores plotted against predicted scores. The x-axis is PANAS(+) in the top plots and PANAS(–) in the bottom plots while the y-axis is the algorithm prediction (each technique uses a different scale). The dots in black are the points labeled confident by human reviewers. 40

Figure 5.1 Chirp application main pages. Note that the main user timeline (b) was not shown to participants in the Individual group, who had the user profile (c) as their main app page. The Timeline and Profile buttons on the bottom of the screen were also not shown to Individual group users. 55

Figure 6.2 An example of a post made in Chirp. The first emoji show is the "main" emoji, and all following emojis are the "story" emojis. This image was shown to participants as part of the final study survey, and participants were asked the following question: What do you think the author of the following post meant to convey about their mood? The first of these emojis was chosen in response to the prompt "How are you feeling", while the others were chosen in response to a follow-up prompt "Want to share why?" 62

List of Tables

Table 3.1 Correlations for social support and loneliness by frequency ofself-disclosure were examined across each one-on-one and group platforms. Note that *higher* levels of social support and *lower* levels of loneliness are the desired outcomes. Online disclosure has a larger effect on both social support and loneliness than hours interacting online. Furthermore,self-disclosure had a stronger relationship specifically with social support than loneliness across all platforms. *p<0.05, **p<0.01, ***p<0.001. 17

Table 3.2 Correlations for social support and loneliness by frequency ofself-disclosure were examined across each one-on-one and group platforms.Self-disclosure had a stronger relationship with social support than loneliness across all platforms. Furthermore, disclosure across one-on-one platforms had a stronger relationship than disclosure across group platforms, with one-on-one messaging having the strongest correlation. *p<0.05, **p<0.01, ***p<0.001. 19

Table 3.3 Differences in the specific emotions disclosed by those who reported disclosing over messaging platforms at least weekly (frequent) or those disclosing monthly or less (infrequent) were examined. One-way ANOVAs were conducted across each emotion, revealing significant differences between groups across all emotions. The mean difference above represents the (mean of frequent)-(mean of infrequent), indicating greater differences for more negative emotions (depression, anxiety, fear), with the exception of anger. This suggests that students reporting higher willingness to disclose online are likely referring to more negative emotions. As increased self-disclosure online was found to be associated with increased perceived social support, it is likely that negative emotions play a larger role than positive emotions in students' perceived social support online during the COVID-19 pandemic. 26

Table 4.1 Sochiatrist Data Extractor example output, demonstrating how messages are collected across platforms and how names are anonymized. These are example messages, not actual participant data. 34

Table 4.2 Examples of negative and positive words used in the PANAS survey. Participants are asked to fill out a Likert scale from 1–5 for each word to answer the question "to what extent do you feel this way right now" 35

Table 4.3 Summary statistics of the messages send and received analyzed in this study, and the number of participants using each platform. Snapchat messages were unavailable at the time. 37

Table 4.4 Spearman's correlations between various techniques for estimating *positive* affect scores from PANAS, i.e. PANAS(+). Human review performs best when the reviewers feel confident about their estimate. LIWC and VADER are highly correlated, but less correlated with human review. They are less correlated with PANAS(+) compared to human review with confidence ignored ("All" column). $*p<0.05$, $**p<0.01$ 42
Table 4.5 Spearman correlations between various techniques for estimating *negative* affect scores from PANAS, i.e. PANAS(–). Human review performs better when the reviewers feel confident about their estimate, although sometimes not statistically significant. LIWC and VADER are highly correlated, but less correlated with human review. Even so, VADER performs similarly to human review with confidence ignored ("All" column) $*p<0.05$, $**p<0.01$ 42
Table 4.6 Perturbed examples where the PANAS scores by participants differed from those decided during human review. In these sessions, the PANAS(–) scores were similar, but the human labeled scores were lower than PANAS for sessions 410 and 707, higher in sessions 258, 749, and 948. All examples have been perturbed for privacy. Note that the average PANAS(+) score is 29.7, the average PANAS(–) score is 14.8, and both have a range of 10–50. Values above and below these averages can be treated as "relatively high" and "relatively low" values for both PANAS and human review scores. Though VADER is on a different scale (from 0 to 1), values above or below its means (0.21 for PANAS(+) prediction and 0.07 for PANAS(–)) can also be treated as "relatively high" or "relatively low". 44
Table 4.7 Different sources of text data from various prior literature report low correlations between PANAS and LIWC; † indicates a correlation between PANAS and VADER rather than PANAS and LIWC (Beasley & Mason, Beasley et al.); Note that Beasley & Mason achieve slightly higher correlations when using a wider time range than the one presented here. We only include the results for the time range shown as it is the time range most similar to ours. Compared to other studies, messaging data analyzed in our paper achieves a similar accuracy range with LIWC, but notably better with VADER, especially for PANAS(–). $*p<0.05$, $**p<0.01$ 47

Abstract

With rising rates of mental health struggles among undergraduate students, social support—specifically *perceived* social support—has become vital in maintaining emotional health. Perceived social support has long been directly tied to metrics of mental, emotional, and physical well-being, including reduced rates of depression, anxiety, and loneliness. Similarly, emotional self-disclosure has shown strong correlations with perceived social support. Online platforms have become a primary form of communication among undergraduates, and thus a central tool towards building social support. An understanding of the role of these platforms in encouraging self-disclosure and social support is a crucial step towards supporting student mental health.

This thesis aims to 1) understand the causal factors that contribute to increased self-disclosure and perceived social support in online spaces and 2) investigate the features of private online spaces that facilitate more open self-disclosure. An exploratory study on online communication habits of undergraduate students finds a strong correlation between comfort with self-disclosure in private messaging and perceived social support. Social support unexpectedly showed a stronger correlation with willingness to self-disclose over private messaging than over phone or video calls. A separate study investigates the accuracy of emotional disclosure over private and public online communication platforms using a unique data extraction platform, Sochiatrist. Compared to previous work, this study shows private online messages to more accurately represent emotional state than posts in public online spaces. Following the previous two studies, Chirp was developed as a social app that encourages emotional self-disclosure in a semi-public space through mood tracking posts. Chirp is built as a tool to individually study factors impacting self-disclosure and social support among cohorts of users without the need to compromise participant privacy.

This thesis proposes that honest self-disclosure and higher levels of perceived social support can be facilitated in online spaces by introducing aspects of private online messaging, such as privacy, perceived control over information spread, and the ability to view and reflect on previous communication.

Chapter 1

Introduction

1.0.0.1 Thesis Statement

Attributes of private messaging encourage emotional disclosure and social support among undergraduates in online spaces.

1.1 Motivation

With rising rates of known or diagnosed mental health disorders, such as depression [59] and anxiety [60], mental health has recently been centered in the national consciousness [61, 112]. This problem has been particularly prevalent among school- and college-age students [59, 60]. In response to this trend there has been a rising focus on social connection building [112], particularly in online spaces where people around college age are increasingly spending more time [139] and looking for connection [62].

Here we propose that attributes related to private textual communication – or messaging – encourage greater emotional disclosure and therefore social support and connection building among undergraduate students.

Past studies have shown that using social media can reduce rates of depression [45], encourage emotional self-disclosure [4], and provide a space for users to find camaraderie and support [6]. Self-disclosure has further been recognized as a fundamental human need [15], and is associated with intimacy and relationship strength [117, 142]. Conversely, social media use has been associated with negative mental health effects, such as negative body image [54], loneliness [123], and internet addiction.

Existing theories suggest that user intent and outside context are primary mediating factors in expected outcomes from social media use (e.g., Self Determination Theory [65, 81, 123], Social Comparison Theory [92], Interpersonal-Connection-Behaviors Framework [32]). However, we note that previous work has shown self-disclosure to be most frequent amongst dyads [142], and that many studies investigating self-disclosure and social support in online spaces have focused on semi-private spaces with shared context (e.g., [3, 58, 168]).

Thus we investigate whether and how private messaging spaces encourage emotional self-disclosure in undergraduate student users, and the impacts this may have on perceived social support, a primary protective factor in mental health contexts. This investigation is contextualized through three themes that together I use to define private messaging. These themes include:

- Written communication – users in private messaging spaces communicate through written information
- Audience selection – private messages are sent and visible to only a chosen set of users
- Social presence – private messages are sent to another user who has the ability to respond in some capacity

1.2 Overview of Contributions

1.2.0.1 Bridging the Social Distance: Offline to Online Social Support during the COVID-19 Pandemic

Gabriela Hoefer, Talie Massachi*, Neil G Xu, Nicole Nugent, Jeff Huang. In proceedings of CSCW 2022.*

In this study, we investigate the mental health impacts associated with the shift away from in-person socialization for online digital contact during the COVID-19 pandemic. We examine properties of online interactions that may affect loneliness and perceived social support, with a particular focus on the impact of **written communication** in private online spaces. Students were surveyed ($N = 827$) across 97 universities across the US during their first full semester impacted by the COVID-19 pandemic (Fall 2020). Private online interactions (messaging, phone call, video call) were found to have a comparable correlation to social support as face-to-face interactions, but public online interactions (social media) were associated with more negative outcomes. Among private platforms, messaging had the strongest correlation with social support – even more so than richer communication platforms such as phone or video calls. We speculate that the persistence of written communication in messaging contexts may allow for later reflection on supportive responses and on personal emotional state that may lead to greater perceived social support when compared to more ephemeral spoken conversations.

1.2.0.2 Sochiatrist: Signals of Affect in Messaging Data

Talie Massachi, Grant Fong*, Varun Mathur*, Sachin Pendse*, Gabriela Hoefer, Jessica Fu, Chong Wang, Nikita Ramoji, Nicole Nugent, Megan Ranney, Daniel Dickstein, Michael Armey, Ellie Pavlick, Jeff Huang. In proceedings of CSCW 2020.*

Many existing techniques for affect detection across social media platforms have relied on sentiment analysis software, including LIWC and VADER. However, the accuracy of these techniques have been called into question, with studies showing little to no correlation between ground truth reported affect and LIWC or VADER scores when run on public social media data. In this study we investigate these same measures as well as human review on private messaging data and affect scores from 25 participants, exploring differences in how and when each technique is successful. Results show that while human review predicts affect better than VADER, the best automated technique, when humans are judging positive affect ($r_s = 0.45$ correlation when confident, $r_s = 0.30$ overall), human reviewers only do slightly better than VADER when judging negative affect ($r_s = 0.38$ correlation when confident, $r_s = 0.29$ overall), indicating that VADER is nearly as accurate as human perception at affect detection in private spaces. Compared to prior literature, VADER correlates more closely with PANAS scores for private messaging than public social media, and similarly when compared to

more traditional personal spaces such as diary entries or spoken conversation. Our results indicate that users may more honestly self-disclose in private spaces than public spaces. We propose that this may be a reflection of both the more unedited nature of communication in these spaces, as well as increased feelings of **control over the audience** of these disclosures.

1.2.0.3 Chirp: A Social System to Encourage Semi-Private Emotional Self-Disclosure

Talie Massachi, Lauren Choi, John Roy, Gabriela Hoefer, Shaun Wallace, Jeff Huang

Given previous evidence of a relationship between online private messaging spaces and positive outcomes, we now seek to better understand relationships between attributes of private messaging and emotionally beneficial outcomes such as increased perceived social support. In order to do so, we need a controlled sandbox space in which to test individual attributes. Thus we developed CHIRP, an anonymous social media platform designed to explore the underlying effects of private online spaces on emotional health. CHIRP prompts users to self-disclose moods and emotions to a semi-private timeline using emojis, creating a small experimental space that emulates online private messaging spaces. In this way, CHIRP provides a way to safely investigate the *causal* effects of specific design decisions and factors influencing social media and online communication through direct experimental interventions.

1.3 Proposed Work and Research Goals

Here I propose two studies. First, I propose a further investigation following the theme of *written communication* in private messaging. In CHIRP we use emojis as a form of limited written communication. While previous work has shown that emojis can be understandable given specific temporal and spatial context [82], here I propose an investigation into how emojis used as responses to questions— as they are used in CHIRP— are understood, as well as how this is impacted by demographic differences. Second, I propose an experiment to better understand the *causal links* between attributes of private messaging and perceived social support. This includes both attributes with which we previously found a relationship with positive outcomes (i.e. written communication and audience selection) and with the concept of *social presence*, or the presence of another user to which disclosures are made and who has the potential to respond. This second study uses CHIRP as a representative online space, along with a between-subjects experimental setup to investigate the stated research goals.

1.3.1 Emoji Understanding and Interpretation with Limited Context

While my preliminary work has pointed towards the positive impact of written communication on well-being, particularly with regard to perceived social support, this has primarily been in the realm of textual communication. However, written communication beyond text can also impart meaning [82]. Based on the understanding that emoji can meaningfully transmit information while obscuring overly personal information, CHIRP (described in Chapter 5 was built using emojis as a primary mode of communication. Here I propose a study to investigate the limits of written, non-textual communication in the context of emotional self-disclosure. This study would investigate not only general understanding of emoji-based communication, but also how it is impacted by the demographics of the interpreter and the interpreter's impression of the demographics of the original poster.

1.3.2 The Impact of Written Disclosure in Semi-Private Online Spaces on Perceived Social Support

Prior work has found both correlational evidence between perceived social support and emotional self-disclosure in private online spaces (e.g., [71]), and qualitative evidence of self-disclosure and socially supportive messages in semi-private spaces (e.g., [3, 58]. However, there has been little work showing the causal relationship between self-disclosure and perceived social support in semi-private spaces [45]. I propose a study leveraging Chirp in a controlled experiment comparing participants instructed to post an emoji-based mood tracking entry in the app daily within a semi-private setting and participants that do not use the app at all.

1.3.3 The Impact of Social Presence When Disclosing in Online Spaces on Perceived Social Support

Previous work, such as that by Deters and Mehl have found that increased posting in a semi-public space (i.e. Facebook) decreased feelings of loneliness regardless of response rate on posts. Simultaneously, reciprocated self-disclosure has also been found to increase relationship strength [117]. These results raise the question of whether reciprocation or even the potential for reciprocation (i.e. social presence) necessary to impact perceived social support. Through the use of Chirp, I propose we test the impact of self-disclosure in a semi-private space (i.e. visible to other cohort members) compared to a fully private space (i.e. visible only to the poster). In doing so, we can experimentally isolate the impact of social presence on percieved social support.

Chapter 2

Related Work

2.1 The Benefits of Online Social Support

Literature in psychology and social science has long shown the benefits of social support. Defined as "an exchange of resources between two individuals perceived by the provider or the recipient to be intended to enhance the wellbeing of the recipient" by Shumaker and Brownell [137], social support, and specifically *perceived* social support, has been tied to many metrics of mental, emotional, and physical well-being [148]. This includes acting as a protective factor that improves emotional resilience in response to trauma [138]. Research has shown that feelings of social support reduce rates of depression and anxiety [148], and are tied to decreased feelings of loneliness [80].

The rise of social media and other online platforms a popular venues for interpersonal communication has led researchers to question the impact of these online communication platforms on feelings of support and other well-being outcomes. Numerous studies have shown the positive effects of online support groups on users (e.g., [38, 58, 79, 157]), including the creation of social support networks over social media platforms (e.g., [6, 58]). Studies have also shown strong correlations between direct online communication with friends or family and feelings of social support [24, 71].

Studies on social support in social media settings have been primarily descriptive, either describing or finding correlations between actions on existing platforms and users feelings [81, 144]. An exception to this, Deters and Mehl [45] ran a longitudinal study investigating the impact of increased Facebook posting on feelings of loneliness. The authors ran a between-subjects study with university students. Experimental group participants were asked to post more status updates on Facebook over the course of a week, while the control group was given no specific instructions. Both groups recorded feelings of loneliness before and after the study. Deters and Mehl found that participants that posted on Facebook more frequently showed lower levels of loneliness, *regardless of whether those posts were interacted with or responded to by their network.* The authors suggest a few potential reasons for this pattern, including that "the act of writing itself...might create a feeling of connectedness" as users have their social network in mind while writing their posts. The authors also propose that public posts may lead to increased rates of private messaging, something that they did not measure, or that the act of self-disclosure in participant posts might encourage conversation about those topics. Deters and Mehl note that in future work these different outcomes could be tested by having separate cases where users post to either a public or private space [45].

2.1.1 Measures of Social Support

Traditionally, perceived social support is measured by the Multidimensional Scale of Perceived Social Support (MSPSS) [172, 174] (shown in Appendix A). However, the questions included in the MSPSS imply in-person and strong-tie interaction and support (i.e., family, close friends, and significant others). To account for the kinds of interaction and connections we see online, Nick developed the Online Social Support Scale (OSSS), a 40-point scale that specifically asks about perceptions of social support in relation to online communities [111] (shown in Appendix B). Both scales are scored as a sum of all responses, where a higher score indicates increased feelings of social support.

In this study, we use the MSPSS as a measurement of offline levels of social support, and the OSSS as a main measure of changes in online social support.

2.2 Online Communication Over Private Platforms

The majority of previous research examines online communication by investigating only public social media platforms, without differentiating between other, private platforms [101, 171]. However, these platforms are different in nature: public platforms, which encompasses social media such as Facebook, Twitter, Reddit, and other public online groups, may reach a wide audience - often with bounds the user is unaware of, whereas private platforms (e.g., messaging, video calls, and phone calls) are intended only for those deliberately included in the interaction [101, 171]. Even fewer studies on mental health have differentiated between platforms [51, 56, 72, 94], despite previous work showing different interaction types between public and private communications [23]. Of the few studies that differentiate, private platforms have emerged as more beneficial for loneliness than public platforms [107]. Their benefits are often credited to the greater perception of anonymity and privacy [171]. Furthermore, there are differences within private platforms that may impact user behavior and should be accounted for. Private platforms allow for a variety of group dynamics, as they enable one-on-one communication (i.e., direct communication between two people) and group communication (i.e., communication between more than two people), which has been shown to impact offline interactions [142]. Thus in this thesis I focus primarily on the impacts of private online communication.

Much of the literature surrounding social support in online communities tie support back to the concept of disclosure (e.g., [5, 32, 68]). Studies of both on- and offline communication show that people feel closer when they emotionally self-disclose (i.e. discuss or reveal their emotions or feelings) and reciprocate [117]. Studies have also found strong direct correlations between the frequency of emotional self-disclosure online and feelings of social support [45, 71, 90, 91]. However, the causal direction of this relationship is still unknown, and is a focus of this study.

One prominent theory of online interaction that tries to explain this relationship is the Interpersonal-Connection-Behaviors (ICB) framework [32]. This framework proposes that the impact of social media on wellbeing, including feelings of social support, relies on the type of interactions that a user engages in. Specifically, that "active", or "connection promoting" behaviors (e.g., posting, commenting, reacting) lead to positive outcomes, while "passive", or non-connection promoting, behaviors (e.g., lurking, reading without engaging) lead to negative outcomes. In this framework, self-disclosure is specifically pointed out as "particularly relevant because technology-mediated self-disclosure is at least as frequent and as meaningful as face-to-face self-disclosure, [109]"(citation theirs) [32].

Chapter 3

Preliminary Work: Offline to Online Social Support during the COVID-19 Pandemic

3.1 Introduction

In recent years, online communication platforms have only become more popular as a space for communication and connection among students. The COVID-19 pandemic further led to a fundamental change in people's social lives on a worldwide scale [156]. At times, the United States had the highest number of reported cases of any country [151], shutting down college campuses nationwide [30]. The Fall 2020 season was a complex period where colleges had unusual arrangements for the student experience. Students suffered from disruption of their formative years, loss of social belonging and community, and worries about the future [21]. However, uniquely, this pandemic was the first major event that discouraged *in-person socialization* while an alternative of *online digital interaction* was as widely available, further driving a shift towards online platforms as a primary mode of communication. University and government lockdowns imposed new norms for student interactions, allowing for a timely investigation into the effects of online communication on student well-being.

Over the three month period from September 24, 2020 to December 24, 2020, we distributed an online survey investigating the impact of pandemic-induced online interactions on social support and loneliness in college students across the United States. The campaign was done directly through contacting academic deans to request they help distribute the survey, as well as through social media forums for college students. The survey was continuously sent out until a stratified sample of students from diverse types of colleges had responded, resulting in 827 complete responses during the crisis.

Each student participant specified the nature of their physical and digital socializing during the pandemic, as well as completed measurements of **loneliness**, defined as the "state experienced when a discrepancy exists between the interpersonal relationships one wishes to have, and those that one perceives they currently have," [121] and **social support**, defined as the "exchange of resources between two individuals perceived by the provider or the recipient to be intended to enhance the well-being of the recipient" [88].

Here we report on the psychological impact of the pandemic and associated lockdown orders on student well-being, finding that these lifestyle changes have led to increases in loneliness and, surprisingly, no significant change in feelings of social support. We identify how different forms of digital communication (e.g., social media, messaging, phone calls, and video calls) correlate with loneliness and perceived social support, with particular focus on emotional self-disclosure across different mechanisms of private communication (i.e. messaging, phone calls, and video calls). In comparing these different platforms, we investigate not only the impacts of online communication on well-being, but also how the difference in *modality* (public vs. private, written vs. spoken) impacts student outcomes. Finally we investigate the degree to which online socialization may be an effective substitute for in-person interaction in the context of loneliness and social support.

In this study ask the following research questions:

RQ1 How do public (social media) and private (messaging, phone calls, video calls) forms of online communication impact social support and loneliness?

RQ2 What is the impact of self-disclosure across different private online communication platforms on social support and loneliness?

RQ3 How does online interaction act as a vehicle for social support and a mitigating factor for loneliness during a period of intense physical isolation?

3.2 Related Work

3.2.1 The State of Mental Health During the COVID-19 Pandemic

Numerous studies reveal a decrease in overall mental well-being throughout the world during the COVID-19 pandemic (e.g., [125, 158, 160]), often attributed to psychosocial stressors associated with unforeseen increased isolation, such as life disruption (e.g., [93, 155]) and fear of illness or uncertainty (e.g., [51, 143]), therefore calling for increased prioritization of mental health (e.g., [152, 165]). Wang et al. found 54% of individuals showed moderate to severe psychological impact from the COVID-19 pandemic, with 29% reporting moderate to severe anxiety and 17% reporting moderate to severe depression in the population of China immediately following the first outbreak of the pandemic [160]. Furthermore, they conducted a follow-up study revealing that these symptoms persisted even a month later [161]. Tull et al. found similar trends in the United States population, revealing stay-at-home orders to be associated with increases in loneliness and decreases in social support [155]. These reports of declining mental health are replicated in studies from Australia [52], the United Kingdom [22], and among college students worldwide (e.g., [26, 51, 93, 95]). Furthermore, individuals in social isolation around the world reported lack of access to adequate mental healthcare, relying primarily on self-help techniques [40]. This matches trends in previous outbreaks, as research during SARS revealed higher levels of stress, anxiety, and depression in the general population [31], and patterns across previous pandemics reveal a lack of adequate mental health resources worldwide during or immediately following the outbreak [149].

3.2.1.1 Among College Students

The COVID-19 pandemic college students in particular; most experienced severe disruptions to academic routine, and many were forced to evacuate dorms, lost on-campus jobs due to the remote work environment, experienced increasing financial hardships, food and housing insecurity, and reduced access to academic material [13, 167]. Moreover, many students reported a decline in feelings of social belonging and community connectedness [167]. Current studies show college students report higher levels of stress, anxiety, and depression than the general population during the COVID-19 pandemic (e.g., [26, 37, 51, 143, 162]). Immediately following the initial outbreak, one large-scale longitudinal study (164,101 college student participants) revealed increased stress, anxiety, depression, and acute stress in college students [93] and another study found similar, persistent mental health trends in college students over multiple months of the pandemic [37]. Due to the clear psychological effect of COVID-19 on this group, our study seeks to specifically examine US college students in order to better understand the impacts and potential benefits of online interaction on social support and loneliness.

3.2.2 Online Communication as a Coping Mechanism During the COVID-19 Pandemic

3.2.2.1 Movement of Communication to Online Platforms

Numerous studies in Human-Computer Interaction examine crisis informatics, the study of the "intersecting trajectories of social, technical and information perspectives during the full life cycle of a crisis" [67]. After almost every disaster we see a regional unification of communities, materials, and information, known as geographical convergence [55]. Studies have found this behavior mirrored

in online interactions, as individuals gravitate towards online communities as a coping mechanism (e.g., [47, 114, 140]) creating broader social convergence without geographical barriers. This trend was found during the COVID-19 pandemic, as people around the world turn to virtual platforms for socialization while under quarantine guidelines (e.g., [56, 113, 156]). In Belgium, social media apps saw a 72% increase in usage, messaging apps saw a 64% increase, and phone calls saw a 44% increase in average time spent on each platform [113]. Similar patterns have also been found in Italy, revealing significant increases across video calls and online gaming platforms [56], and in the United States, revealing an increase in posting across public social media platforms [156].

Although some increases in online interaction were expected due to the nature of an international crisis, it is likely that online interaction was further escalated due to the global quarantine-induced isolation that impeded social interactions on a larger scale than seen in recent crises. This effect is supported by the Social Compensation Hypothesis, which suggests individuals will compensate for in-person social deficits by increasing their online interaction [103]. From this perspective, the abrupt halt in face-to-face interactions that many experience as a result of the pandemic may create inherent social deficits and draw individuals online as a mode of compensation.

In addition to online presence, online expression was amplified during the COVID-19 pandemic [14]; and the sentiment of public-facing social media posts was found to be significantly more negative [156]. Self-disclosure, "an interaction between at least two individuals where at least one intends to deliberately divulge something personal to another," [63] is associated with increased stressful life events [170] and recognized as a fundamental human need [15]. Furthermore, self-disclosure is associated with intimacy and relationship closeness, and is most frequent between dyads [142], suggesting that disclosure online may be more common in one-on-one chats, rather than in group chats or public social media.

Online disclosure is often met with reciprocated self-disclosure from others, and has been found to increase relationship strength [117] and facilitate social support [3]. Current work suggests there may be an increased likelihood of willingness to disclose during the COVID-19 pandemic [108], however, fails to examine changes in disclosure patterns across multiple online platforms. We seek to fill a gap in the literature by examining the specific emotions students disclose through online private platforms and the relationship of each of these emotions with perceived social support and loneliness during this time.

3.2.2.2 Social Support Compensation Through Online Interaction

Perceived social support has been found to improve resilience to trauma due to neurobiological factors as well as its inherent promotion of stress-regulating behaviors [138]. During the COVID-19 pandemic, increased social support was also shown to decrease crisis-related stress symptoms such as anxiety and depression [110].

Studies reveal online communities to be an effective method for receiving social support [3], however it is unclear whether online interactions can generate similar levels of social support derived from in-person interactions. During periods of physical isolation brought on by COVID-19, research shows college students report using online communication intentionally as a means of gaining support [162], and digital platforms were effective in truly increasing perceived social support during this time [56, 95]. Despite the positive effects of online interactions, other studies have found participants tend to believe virtual platforms are inadequate substitutes for in-person conversation [122]. The physical isolation during the COVID-19 pandemic thus provides us with an opportunity to investigate

whether online interactions are truly an effective substitute for receiving in-person social support.

3.2.2.3 The Potential for Mitigating Loneliness Online

Prior to and during the COVID-19 pandemic, studies have revealed inconsistent patterns between online interactions and loneliness. Some studies suggest technology is helpful for fostering feelings of social connection in isolated individuals [107], yet others reveal associations between increased public social media usage and increased feelings of loneliness [95, 124]. More recent research has also found increased levels of support seeking behavior on public social media platforms [95] and have found individuals are likely to be using public platforms as a coping mechanism to compensate for existing feelings of loneliness [107, 118], rather than social media causing feelings of loneliness [156]. This raises question to whether online interactions are truly mitigating feelings of loneliness. Our study seeks to better understand the factors influencing loneliness, specifically the differing effects of online and in-person communication as well as the impact from properties of separate online platforms (i.e., social media, messaging, phone call, video call).

3.2.2.4 Effect of Interaction Type and Intent

Individuals across the world appear to be using online platforms as a coping mechanism, however, it is unclear whether these platforms truly combat feelings of loneliness or provide social support during periods of isolation. Previous research has revealed public social media to be correlated with increased feelings of loneliness [156]; however, other research has found social media to promote coping techniques and the use of technologies to increase well-being (e.g., self-care, meditation videos) [132]. These contrasting patterns may be due to the way in which an individual is using the social platform.

The Interpersonal-Connection-Behaviors (ICB) Framework suggests that the use of online platforms in active relationship building leads to positive outcomes, whereas more passive usage yields fewer benefits [32]. This framework proposes categorizing usage within the platforms: (A) active, meaning engaging in interactions that promote user's active building of relationships (e.g., commenting, having a conversation), or (B) passive, meaning without any inherently relational engagement (e.g., scrolling without engagement) [32]. Active interactions are associated with greater positive outcomes, whereas passive interactions present negative associations such as increased feelings of loneliness [33]. Similarly, research conducted during the COVID-19 pandemic has also found that the total time spent interacting has less bearing on well-being than the amount of satisfaction one derives from interacting online [122], however, has not considered the active or passive nature of the interaction.

Recent work has challenged the active and passive categorization, suggesting that behavior previously classified as passive usage can still have relational qualities [20]. As a result, our study seeks to further investigate the passive and active comparison made in the ICB framework by more closely examining various factors and behaviors that may contribute to social support and loneliness outcomes associated with online interaction during the COVID-19 pandemic.

3.3 Methods

To better understand the psychosocial effects of the COVID-19 pandemic as well as investigate the implications of various online platforms on US college students, we distributed an online survey as-

sessing students' daily life, online interaction patterns, and critical aspects of their mental health, namely loneliness and social support. Upon completion of the survey, participants were given the opportunity to enter in a raffle to receive one of ten $25 digital gift cards.

All participants were US college students, comprising a group of people who were immediately disrupted by the pandemic, with many being forced to relocate out of their dorms and away from their primary social support groups within just a few days [167]. Moreover, research shows this age group uses online platforms more than any other age group in the US [28], allowing us to examine its implications during this time. Additionally, participants were limited to students actively enrolled in accredited US universities as countries around the world elected varying methods of crisis management and were at different stages of the COVID-19 crisis while the survey was distributed.

3.3.1 Survey Design

A qualitative questionnaire was drafted, refined and then underwent two rounds of pilot testing and a cognitive field test. In the first pilot test, the broader survey was evaluated for potential insights on 11 graduate and undergraduate student researchers. Experimenters then discussed the impact of online interactions and self-disclosure through various platforms with these researchers. Based on their feedback, we identified self-disclosure patterns as an area of potential impact and therefore included the Emotional Self-Disclosure Scale (ESDS) [141] to assess disclosure of specific emotions.

After incorporating feedback from student researchers, the survey was tested a second time through a pilot test with 26 college students across five U.S. universities, with additional qualitative follow-up questions in order to improve the pilot survey. Based on the findings, we reordered our quantitative survey questions to fit conversational conventions. The range for potential responses for usage of online platforms was also expanded in response to feedback. We also tested questions assessing trust across each platform (e.g., Please rate how much you trust the following communication platforms – Twitter). Students reported significant difficulty in answering this due to many confounding factors (e.g., data privacy, company scandals, misconceptions, etc), thereby increasing survey fatigue. As a result, all questions assessing trust of these platforms were removed.

Finally, we conducted a cognitive pretest and follow-up interview with a college student who had previously never seen the questionnaire. We observed the student participant as they filled out the questionnaire and gathered insights on their experience. The student expressed feelings of survey fatigue given the length of the questionnaire. In order to ensure accurate results, we incorporated validation questions (e.g., "What is the name of the current global pandemic?") and reverse wording. Any responses that failed to correctly answer validation questions were removed.

3.3.2 Survey Questions

The distributed questionnaire was quantitative in nature, and demographic questions included self-identification of gender, university, and level of study (i.e., graduate or undergraduate student). Additional demographic data such as age and degree major was not collected to preserve participant privacy. The remaining part of the questionnaire can be separated into three sections: physical isolation, online interaction behavior, and scales to measure social support and loneliness.

3.3.2.1 Assessing Amount of Physical Isolation from Others during the COVID-19 Pandemic

In the first part of the questionnaire, we address the level at which participants are physically isolated from other people during the pandemic by examining the number of in-person interactions they had. Questions first examined their behavior during the pandemic, through multiple-choice responses asking participants to indicate their adherence to standard government social distancing guidelines and select the level of quarantine that best matches their behavior, such as: (a) "*I am in mandatory quarantine (mandated by government to stay home)*", (b) "*I am in voluntary self-isolation (not completely mandatory by government, but I am staying home)*", (c) "*I am going about my daily business like usual (going to work, school, etc) but avoid restaurants, cinemas, etc*", (d) "*I am going about my daily business like usual (going to work, school, etc) AND visit restaurants, cinemas, etc*", (e) "*My work requires me to go out (e.g. medical, delivery, public transport) while others are recommended to stay home.*" Moreover, participants were asked to specify their in-person interactions from the previous week: including the number of days one spent self-isolating (not leaving the house), and the number of days one had face-to-face contact with another person for 15 minutes or more (including someone living in the same space). Furthermore, self-report questions asked participants to indicate the number of people they live with, the number of people they talk to, and their keyworker status. We also assess COVID-19 impacts through self-report questions such as relationship changes and whether the participant is located in a COVID-19 hotspot.

3.3.2.2 Measuring Online Interaction During the Pandemic

The second part of the questionnaire examined college student online interactions during the COVID-19 pandemic. This seeks to gain insights into whether students are building community or socializing through online platforms. Multiple choice questions assessed online behavior prior to and during the COVID-19 pandemic, including the amount of time spent per week interacting on each platform (specific options included: *"Social media (Twitter, Facebook, Reddit, etc.)", "Messaging (SMS, Facebook Messenger, DMs, etc.)", "Videochat" and "Phone call"*). Participants reported how frequently they disclose across private platforms, based on the number of people in the interaction (i.e., *"How often do you discuss your emotional state with friends/family over... Phone call with 1 person, Phone call with 2+ people,"...* across phone calls, messaging and video calls).

Furthermore, we measured how willing participants were to self-disclose eight specific emotions online, then analyzed the differences between these emotions in order to better understand self-disclosure and its effect on perceived social support and loneliness at this time. We include a modified version of the Emotional Self-Disclosure Scale (ESDS) [141] to measure participant comfort when sharing emotions to another person during the COVID-19 pandemic. The ESDS is a popular assessment of how often individuals share certain emotions to a given person (e.g., friend, family, etc) and was modified for this study to specifically reflect only one-on-one messaging during the pandemic (as in-person communication was not an option for many in quarantine). Participants respond on a 1–5 scale of the willingness they are to disclose each of the given 40 emotion topics (i.e., How comfortable are you disclosing – "Times when you felt depressed"). We include all of the original 40 emotion topics of the original ESDS but request "the extent to which you discuss these feelings and emotions through private messaging". The McDonald's ω for this scale in this study was 0.98, indicating this scale had good reliability.

3.3.2.3 Understanding Perceived Social Support and Loneliness

The final part of the questionnaire seeks to assess specific mental health trends. This included the short form of the University of California, Los Angeles Loneliness Scale (ULS-8), which includes 8-items and is scored continuously, with higher scores indicating higher levels of loneliness [70]. This measurement, as well as the 20-item long form scale, has been widely used in the research community as a measurement of loneliness [129]. The McDonald's ω of this scale for this study was 0.70, indicating this scale had good reliability.

We also included the Multidimensional Scale of Perceived Social Support (MSPSS), a 12-item measurement of social support from three collectives: family, friends, and significant others [173]. Participants indicate agreement with questions such as, "I can count on my friends when things go wrong," with higher mean scores indicating higher perceived social support. Previous research indicates that this measure has adequate psychometric properties for adults [25] and it is a popular tool for measurement during the COVID-19 pandemic specifically [93, 155]. The McDonald's ω of this scale in this study was 0.90, indicating this scale had good reliability. Scores were cross-analyzed to consider the context of physical and online interaction.

3.3.3 Participant Recruitment

Completed online survey results were collected from 827 actively enrolled college students across 97 accredited US institutions during the Fall 2020 semester (September–December, 2020). The majority of participants were undergraduate students (87.34%), and participants attended institutions across a wide variety of student body population sizes with 20.2% small (< 5,000), 10.2% medium (5,000–15,000), 32.7% large (15,000–30,000), 36.9% huge (30,000+). Across the 97 institutions, there were an average of 24 participants per school ($SD = 28.28$). Participants attended institutions across 34 states, with the majority of participants attending schools in the south (44%) and midwest (29%), followed by the west (17%) and northeast (12%) United States. 62% of students identified as female, 33% as male, and 3% identified as nonbinary. When asked about their household, 10.5% of students reported living completely alone, 81% lived with 1–4 people, and 9% of students lived with 5 people or more. The majority of participants (82%) did not identify as a government keyworker at the time of survey completion and 58% of students identified as living in a hotspot at some point during the pandemic.

Recruitment targeted students above the age of 18 actively enrolled in accredited colleges in the United States during the Fall 2020 semester (September–December, 2020). Schools targeted held varying profiles, including small liberal arts colleges, large urban research universities, and rural universities with a high proportion of commuting students. Participants were recruited using three methods: emails from university administrators, Reddit posts, and Facebook advertisements. Instagram, which displays Facebook advertisements, and Facebook were chosen as avenues for recruiting as they were the most widespread methods of communication among college-age people [28]. Reddit was also chosen as a recruitment avenue as it allowed us to more directly target students from specific schools.

For the first of our three recruitment methods, we contacted about 200 administrators (e.g., academic deans, directors of student affairs, etc) at academic institutions directly by email to advertise this study to their students. Academic administrators then passed along this advertisement through the appropriate channels, depending on their institution. At least 58% of responses were gathered

through this method.

Secondly, we posted advertisements using Reddit, a popular social media platform often used by other researchers to gather data during the COVID-19 pandemic [14, 169]. University-specific subreddits gave us the rare opportunity to target students across a variety of universities. Targeted universities were randomly selected from accredited schools from the US Department of Education's Database of Postsecondary Institutions and Programs [49]. We then determined if there was a subreddit for the institution, contacted the subreddit moderators for approval, and posted accordingly. In order to ensure well-rounded responses, schools were later narrowed down based on size and university COVID-19 response status [30], then randomly selected. Advertisements were posted on 80 university-specific subreddits, accounting for at least 30% of the responses gathered.

Finally, approximately 3% of participants were recruited through Facebook advertising. Using stratified sampling, participants were grouped based on university COVID-19 regulations in order to ensure adequate representation in our sample. As there were comparatively few responses from universities with fully or primarily in-person status, the goal of this final round of recruitment was to target students from universities in this sub-group. The advertisement audience was limited to students (age 18-30) who identified on Facebook as currently attending universities who had been reported as having fully in-person or primarily in-person courses over the Fall 2020 semester [30].

This study intended to collect data from a wide variety of participants for reliability purposes (i.e., $N > 500$). A total of 1,328 individuals accessed the survey, however, 288 did not complete the survey, 122 declared they were not above the age of 18, 26 did not attend an accredited university in the US and 1 did not pass the validation question. We also monitored the amount of time responses took; as this survey took an average of 20 minutes, the 56 individuals who completed the survey under 5 minutes and the 8 individuals who took over 3 hours were removed for reliability purposes. After filtering, the final sample included a total of 827 participants.

Previous studies have examined the correlation between online interactions and social support outside of the COVID-19 pandemic [102], finding a medium correlation ($r = 0.36$) between social support and number of hours interacting online. A power analysis was conducted for correlation sample size [73] with power β set at 0.10 and $\alpha = 0.05$, two-tailed. Results revealed that in order to receive a similar effect size ($r = 0.36$), any sub-populations that we use should consist of at least 77 participants to reach statistical significance at the .05 level.

3.4 Results

Here we examine trends among our survey responses. We first investigate the general impacts of the COVID-19 pandemic and associated lockdown on college student mental and emotional well-being. We then evaluate how these changes can be contextualized by changes in online platform usage. Finally, we look specifically at self-disclosure over private online spaces and its impact on loneliness and perceived social support.

Analysis using JASP and python scripts examined the linear relationship among various interaction variables with perceived social support and loneliness, as outlined in the sections below. Normality was checked using the Shapiro-Wilk test, showing all instruments followed a normal distribution. Descriptive analysis was determined for frequencies and percentages, Pearson's and Spearman's correlations were examined for linear relationships, and one-way analysis of variance (ANOVA) analyzed

the differences between select groups. Tukey post-hoc tests determined the honest significant differences between groups.

3.4.1 Impact of Physical Isolation on College Students

3.4.1.1 Social Impacts of Physical Isolation

Participants reported the impact of COVID-19 on their relationships: friendships seemed to suffer heavily, as 64% of students felt COVID-19 had a negative impact on their friendships, whereas only 36% of students felt COVID-19 had a negative impact on their familial relationships. In general, close personal relationships (e.g., best friend, close relatives) did not suffer as much as distant relationships (e.g., acquaintances), with 44% of students reporting a negative impact on close relationships, and 63% reporting a negative impact on distant relationships.

3.4.1.2 Emotional Impacts of Physical Isolation

We also investigated behavioral patterns to gauge how physical quarantine during the COVID-19 pandemic impacted perceived social support and loneliness. Participants self-reported their behavior in one of five groups: mandatory quarantine (1.67%), voluntary quarantine (25.24%), normal life but avoiding select public areas (47.98%), completely normal life (21.43%), and keyworkers going to public areas as part of a condition for employment (3.57%). In our study, when we compare across quarantine levels, we do not include individuals who identify as keyworkers, as keyworkers have been found to have higher levels of loneliness and lower levels of social support during the pandemic compared to the rest of the population [75].

A one-way ANOVA examining the differences between loneliness across quarantine levels showed a small significant variation among groups $F(21.8,793) = 9.67, p < 0.001, n^2 = 0.02$. Post-hoc Tukey tests revealed an honest significant difference ($p < 0.001$) between loneliness, as those living under strict quarantine (voluntary or mandatory) reported higher levels of loneliness ($Mean = 3.14$, $SD = 1.11$) than those living in normal life without quarantine ($Mean = 2.68$, $SD = 1.06$). An honest significant difference was almost found ($p = 0.06$) as those living under strict quarantine reported higher levels of loneliness ($Mean = 3.14$, $SD = 1.11$) than those living under some quarantine restrictions ($Mean = 2.95$, $SD = 1.03$). An honest significant difference ($p = 0.02$) was also found as those living under some quarantine restrictions reported higher feelings of loneliness ($Mean = 2.95$, $SD = 1.03$) than those living in normal life without quarantine ($Mean = 2.68$, $SD = 1.06$).

Interestingly, a one-way ANOVA revealed no differences in perceived social support between quarantine groups, as the effect size was zero $F(6.85,793) = 3.23, p = 0.04, n^2 = 0.00$, however those under stricter quarantine groups reported slightly lower social support ($Mean = 3.57$, $SD = 1.12$), followed by those under some restrictions ($Mean = 3.75$, $SD = 0.94$), and finally those living their life normally without quarantine reported the highest social support ($Mean = 3.82$, $SD = 1.12$). It appears that individuals under stricter quarantine are experiencing increased levels of loneliness, but do not exhibit differences in social support. As individuals in stricter quarantine are spending more time physically distant from others, it is possible they are receiving social support digitally.

In addition to quarantine status, we also ran Pearson's correlations to examine the linear relationship between the actual number of people participants interacted with during the pandemic (including both online and in-person interactions) and both perceived social support and loneliness. We found a

Correlations with Social Support and Loneliness

	In-Person Interaction	Hours Interacting Online	Disclosure Online
Loneliness	-0.23***	0.05	-0.10**
Social Support	0.20***	0.073*	0.29***

Table 3.1: Correlations for social support and loneliness by frequency of self-disclosure were examined across each one-on-one and group platforms. Note that *higher* levels of social support and *lower* levels of loneliness are the desired outcomes. Online disclosure has a larger effect on both social support and loneliness than hours interacting online. Furthermore, self-disclosure had a stronger relationship specifically with social support than loneliness across all platforms. *p<0.05, **p<0.01, ***p<0.001.

Change in Online Interactions During the COVID-19 Pandemic

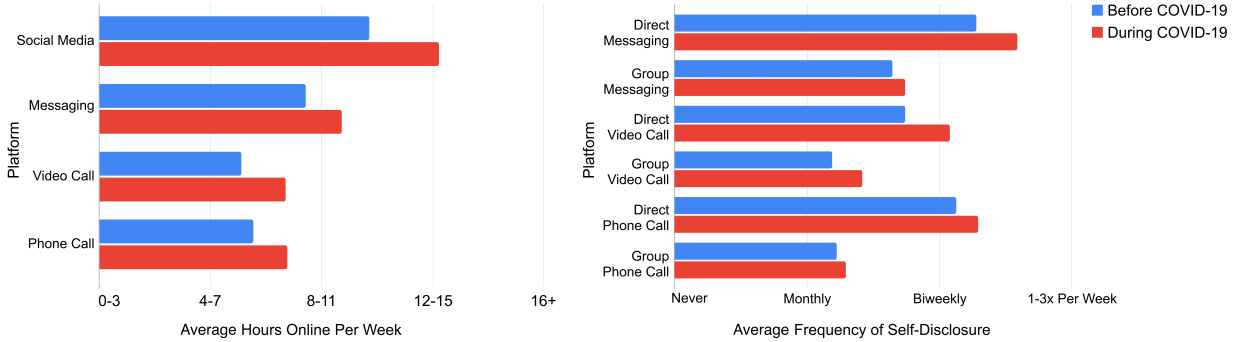


Figure 3.1: The average self-reported online interaction frequency was compared before and during the COVID-19 pandemic. The average hours spent interacting on each platform per week (left), as well as frequency of self-disclosure (right) on each platform was calculated. College students saw an increase in usage across all platforms, with the largest being social media. A greater increase was found in frequency of self-disclosure across these platforms, and students on average spend more time disclosing over one-on-one platforms instead of group platforms, with the most being one-on-one messaging.

very weak positive correlation ($r=0.11, p<0.001$) between an increase in the number of people interacted with and perceived social support. Moreover, a Pearson's correlation for the number of people interacted with and reported loneliness showed a weak negative correlation ($r=-0.17, p<0.001$).

As 84% of students reported feeling a large difference between online and in-person conversations, the implications of each on perceived social support and loneliness were examined individually. We compare each of amount of in-person interaction, amount of online interaction, and frequency of disclosure in private communication online with feelings of social support and loneliness, as shown in Table 3.1. In-person interaction was classified as the self-report number of days over last week in which a participant had a physical, face-to-face interaction with another person for at least 15 minutes (including someone living with them). It appears that both social support and loneliness have moderate correlation with in-person interactions, however, social support has much stronger correlations with online interaction, specifically self-disclosure (i.e., the frequency of discussing one's emotional state online). Although we find social support and loneliness to be moderately correlated with each other ($r=-0.41$, $p<0.001$), this suggests they may be impacted by different factors. This may point to loneliness being an emotion that can be more effectively mitigated through in-person contact than online interactions.

3.4.2 Changing Trends in Online Platform Usage

We next examined changes in platform usage during the pandemic to understand the effects of communicating with a greater number of people online. For the purpose of our study, we consider public online interactions as posts made on social media platforms (e.g., Facebook, Twitter, Reddit, etc.), that allow for a diverse audience of recipients; we consider private online interactions as messages sent directly to an individual or group of specific recipients (e.g. DMs, texts, phone calls, etc.).

We compare the average self-reported frequency of platform usage and self-disclosure across online platforms before and during the COVID-19 pandemic in Figure 3.1. We saw an increase in time spent across each online platform surveyed, in line with previous research showing general increases in student online activity during the pandemic [95]. Students spent at least 4 hours on average on each platform and spent the most amount of time on public social media.

Although students spent the most time on public social media, this was the only platform to show negative consequences with perceived social support and loneliness. More specifically, increased loneliness (Pearson's $r = 0.15$, $p < 0.001$) and, although insignificant, decreased social support (Pearson's $r = -0.06$, $p = 0.10$) was found to be associated with increased time spent interacting on social media. Interestingly, all other private platforms (messaging, video calls and phone calls) saw the opposite effect, and were associated with decreases in loneliness and increases in social support. As a result, we sought to better understand the positive effects of private platforms as well as investigate properties of these interactions that affect feelings of loneliness and perceived social support in the following section.

3.4.2.1 Comparing Online and In-Person Interaction for Social Support

As we see no significant difference in feelings of social support between quarantine groups and simultaneously we see an increase in online communication and self-disclosure in online spaces, we more closely investigate the relationship between social support, frequency of self-disclosure over private messaging, and frequency of in-person interactions. Figure **??** displays a comparison of social support across in-person interactions and online self-disclosure for each platform. Online self-disclosure across all platforms was shown to be associated with increased social support, even when individuals were more physically distant. Specifically, students who only interacted in-person and never disclosed their emotions over messaging reported roughly the same level of perceived social support ($Mean = 3.52$) as individuals who also disclosed their emotions frequently, but rarely interact in-person ($Mean = 3.84$). This indicates that regardless of the number of in-person interactions, self-disclosure across messaging has similar ties to social support. Furthermore, of individuals who do not interact in person, those who also did not self-disclose over messaging reported far lower levels of social support ($Mean = 2.92$) than those who frequently self-disclose over messaging ($Mean = 3.84$). This suggests that the practice of self-disclosure online could serve as an effective means of fostering social support even without in-person interactions.

3.4.3 Patterns in Self-Disclosure in Private Online Interactions

As previously mentioned, we saw an increase in usage across all private platforms during the pandemic (Figure 3.1). Of all the private platforms examined, users spent the most time messaging, and the least time communicating over video call. Overall, while we saw an increase in self-reported frequency

Emotions Shared Based on Frequency of General Messaging Disclosure

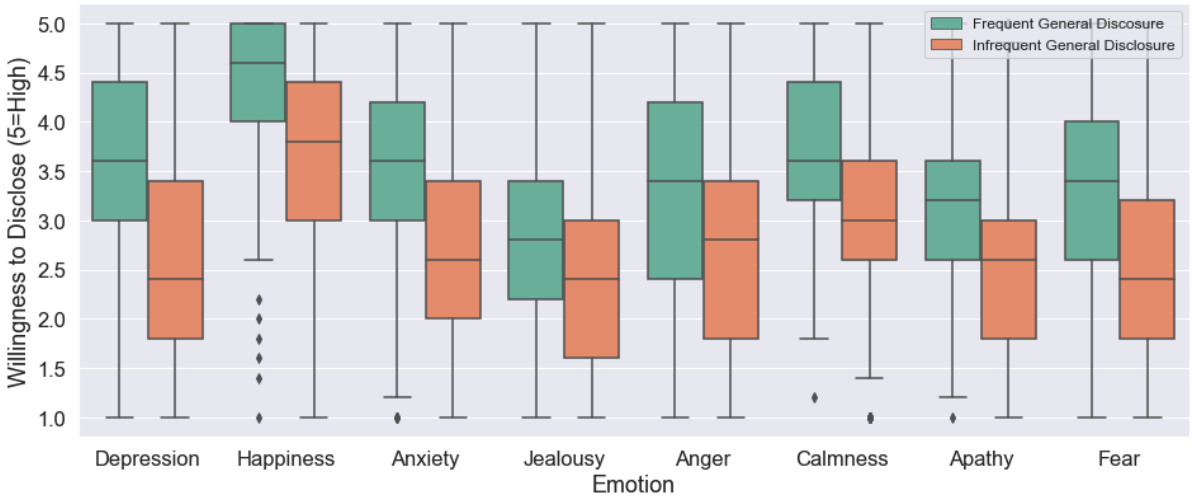


Figure 3.2: Data is grouped by those disclosing over messaging platforms at least weekly (frequent) or those disclosing monthly or less (infrequent). The specific emotions shared are plotted, revealing significant differences between groups across all emotions, with greater increases for more negative emotions (depression, anxiety, fear), with the exception of anger. It appears that those who are more willing to disclose are specifically referring to negative emotions, not disclosure of positive emotions. It appears that students who share negative emotions more frequently may perceive added benefits to social support.

of self-disclosure across all private platforms, trends between platforms seemed to stay consistent between the time before and during the pandemic. Similarly, it appears students most frequently self-disclose over messaging, and least frequently disclose over video call. Furthermore, students more frequently disclose their emotional state across one-on-one platforms ($Mean = 2.32$, $SD = 1.43$, indicating biweekly disclosure on average) than group platforms ($Mean = 1.49$, $SD = 1.04$, indicating monthly disclosure on average), with one-on-one messaging being the most frequent ($Mean = 2.59$, $SD = 1.45$), as shown in Figure 3.1.

Upon closer examination, we found the frequency of self-disclosure to have much stronger relationships with both social support and loneliness than the hours spent generally interacting online, as shown in Table 3.1. As a result, we compared self-disclosure across different private online platforms to see effects on social support. As shown in Table 3.2, Pearson's correlations revealed all one-on-one platforms ($0.19 \leq r \leq 0.25$, $p < 0.001$) to have a stronger relationship with perceived social support and

Self-Disclosure Across Online Private Platforms

	Social Support			Loneliness		
	Messaging	Phone Call	Video Call	Messaging	Phone Call	Video Call
One-on-one	0.25***	0.23***	0.20***	-0.05	-0.12***	-0.12***
Group	0.13***	0.10**	0.09**	-0.09**	-0.09**	-0.06

Table 3.2: Correlations for social support and loneliness by frequency of self-disclosure were examined across each one-on-one and group platforms. Self-disclosure had a stronger relationship with social support than loneliness across all platforms. Furthermore, disclosure across one-on-one platforms had a stronger relationship than disclosure across group platforms, with one-on-one messaging having the strongest correlation. *p<0.05, **p<0.01, ***p<0.001.

loneliness than group platforms ($0.09 \leq r \leq 0.13, p < 0.001$). Moreover, frequency of disclosure had a stronger relationship with social support than loneliness across all platforms, with the strongest being one-on-one messaging ($r = 0.25, p < 0.001$). In line with previous findings, social support has stronger relationships with self-disclosure than loneliness across each platform. One-on-one communication appears to have stronger ties than group communication, and specifically one-on-one messaging seems to have the strongest connection to social support.

3.4.3.1 Impacts of Disclosure by Emotion in Private Messaging

On average, we found the general population to be unsure of whether messaging was helping them cope ($Mean = 3.17, SD = 1.22$). This, in conjunction with our finding that frequency of disclosure over messaging is moderately correlated with feelings of social support calls for a deeper dive into disclosure over messages.

We sought to understand the specific emotions shared by those with increased willingness to disclose. Pearson's correlations analyzed student willingness to disclose a fixed set of emotions surveyed on the ESDS Scale (depression, happiness, jealousy, anxiety, anger, calmness, apathy, fear) [141] and perceived social support and loneliness. Increased willingness to disclose each emotion surveyed had moderate significant positive correlations with perceived social support ($0.27 \leq r \leq 0.33$, $p < .001$), as well as weak significant negative correlations with feelings of loneliness ($-0.010 \leq r \leq -0.007$, $p < .001$).

We examined the specific emotions disclosed by those who reported disclosing over messaging platforms at least weekly (frequent) or those disclosing monthly or less (infrequent), as shown in Figure 3.2. One-way ANOVAs were conducted across each emotion, revealing significant differences between groups across all emotions. However, we find that participants who disclose more frequently had greater increases in willingness to disclose for more negative emotions (depression, anxiety, fear) than more positive emotions (e.g. happiness, calmness), with the exception of anger, as shown in Appendix A, Table 3.3. As increased disclosure was found to be associated with increased perceived social support, this may indicate that negative emotions play a larger role than positive emotions in how individuals perceive social support online.

3.4.4 Main Takeaways

Overall, our results reveal the role of online online interaction during periods of physical isolation during the COVID-19 pandemic. We find that during this period, college students in stricter quarantine experienced higher levels of loneliness, but appeared to maintain their levels of social support. In conjunction, students who engaged more over public social media reported higher levels of loneliness and, though not statistically significant, lower levels of social support. However, students who engaged more over private communication channels reported lower levels of loneliness and higher levels of social support (RQ1). We find that the act of self-disclosure on private interaction platforms (messaging, phone, video calls) is associated with higher levels of perceived social support (RQ2), and that among these, users self-disclose most frequently over messaging and least frequently over video call. Furthermore, we find frequency of self-disclosure over messaging platforms is more strongly associated with perceived social support than either phone or video calls (RQ2), regardless of whether those conversations are one-on-one or in a group. Finally, we find that while increased feelings of social support

and decreased feelings of loneliness both correlated moderately with amount of in-person interaction, social support correlated with frequency of emotional disclosure online ($r = 0.29, p < 0.001$) much more strongly than loneliness did ($r = -0.1, p < 0.01$) (RQ3). In the specific case of social support, we find that students who self-disclose daily over online platforms show similar levels of social support as students who interact daily with others in person (RQ3).

3.5 Discussion

This study sought to investigate the impacts of public and private online communication platforms on college student loneliness and perceived social support during the COVID-19 pandemic. Below we emphasize three implications of this work.

3.5.1 Negative Relationship Between Public Posting and Well-Being

In contrast to private online interactions, which encompass more intimate interactions on private messaging, phone, and video calls, public online interactions include engagement within social media platforms (e.g., Facebook, Twitter) which allow for a diverse audience of recipients that, in many cases, may be unknown to the poster. We questioned the effect of public and private online interactions on feelings of social support and loneliness, and found that, while private messaging is associated with benefits to social support and reducing loneliness, the same does not hold for public online interactions.

By nature, public online interactions typically allow users to reach a wider number of people than private platforms. Similar to other studies conducted during the pandemic (e.g., [12, 56, 113]), our results show an increase in online interaction, particularly on social media. Participants reported spending more time on public social media than any other platform (e.g., messaging, phone call, video call), again matching previous findings in the US during the pandemic [156].

Given the large body of contrasting literature into the implications of social media use on various constructs related to mental health (e.g., [56, 107, 124, 132, 156]), including social support and loneliness; our findings that time spent on public platforms is correlated with higher levels of loneliness would seem to support the idea that public social media use has a negative effect on feelings of social support and loneliness. However, given the positive associations we find for private communication platforms, this explanation appears insufficient. While our analysis is not exhaustive, we propose a few considerations for why this pattern may exist.

We already expect to see a shift online with the beginning of the COVID-19 crisis, based on known behavioral patterns such as geographical convergence and its online equivalent [114, 140]. However, if this were the only factor, then we would expect to see similar social support and loneliness patterns across all online platforms. That we see differing patterns in public and private communication platforms indicates that other factors may be impacting college students.

Previous work has tied increases in online interaction to the Social Compensation Hypothesis [107], which suggests people gravitate towards online interactions to compensate for in-person social deficits [103]. Some people may attempt to increase their popularity online to compensate for their perceived lack of offline popularity [175] or increased introversion in offline social settings [103]. In the context of the COVID-19 pandemic, we may see students moving online to compensate for a lack of in-person interaction due to quarantine and social-distancing efforts.

Public social media platforms inherently enable users to interact with more people at once. They may therefore be the natural place for students to go to for companionship when they are feeling lonely or in need of support from a variety of perspectives. This would mean that it is the *existence of those negative feelings* that drives the correlation with social media use, rather than social media being the cause of negative feelings. However, what this does not explain is why we still see high levels of perceived social support associated with significant amounts of use and emotional disclosure across private platforms.

We can also examine this disparity between public and private messaging from the perspective of the Interpersonal Connection Behaviors (ICB) Framework. This Framework theorizes that outcomes of online platforms depend on whether users are actively using them for relationship building [32, 33].

Public social media usage in general does not inherently imply either active or passive use, and indeed some recent work has shown that traditionally "passive" activities could also be relational in nature [20]. However, traditional passive activities (e.g., reading/scrolling without commenting) are primarily associated with public platforms.

If we make the assumption that most time spent on public platforms is passive interaction, while a relatively large percentage of private online interaction (particularly when self-disclosing) is active, then our findings are aligned with the approach of the ICB Framework. We would expect to see negative outcomes associated with public platform usage in this case, along with more positive associations with private platform usage.

Finally, as social support and loneliness are, by their nature, emotions affected by interaction, maybe the lack of positive outcomes over platforms is a result of lack of interaction. That is to say, it may be that students gain less reciprocation and response over public platforms than in private communication. While it is possible to direct a public social media post towards a particular person (for example, Tweeting "at" another user), this is not the intended type of interaction for those platforms. Thus when a student posts on a public platform, there may not be significant or sufficient responses to elicit a positive outcome. In contrast, when a student communicates over a private platform, there is a specific person or group of people who are being contacted to engage, increasing the likelihood of a response.

This interaction pattern is reminiscent of known psychological phenomena, first being "diffusion of responsibility", where as the number of bystanders increases, an individual bystander shares the moral responsibility of intervening to assist the victim [89], bringing about nonintervention known as the "bystander effect" [57]. Recent work suggests this phenomena is linked to social media behavior and online victimization [99]. In the context of online interactions, this becomes more apparent when we look at an example: Say a user is feeling particularly negative and wants to reach out for support. Consider an identical message, say "I am feeling very sad today, please help cheer me up", posted on a public and private platform. In a public space, other users see the message but, knowing that there are many other people also seeing the same message, may believe that they do not have a satisfactory response, and that someone else would take up that emotional burden. In a private space, a user seeing this message is singled out, and is more likely to respond and engage.

3.5.2 Benefits of Private Communication

In contrast to public platform use, we see a more positive relationship between private platform usage and loneliness and perceived social support, despite public social media platforms showing the most

use and the greatest increase in use with the start of the pandemic. We also found self-disclosure over private platforms to be the strongest predictor of perceived social support, **even when compared to frequency of in-person interactions**. As shown in Figure **??**, the more time one spends disclosing online, regardless of the frequency of face-to-face conversations, is associated with greater social support when done so across private platforms.

This may indicate that some specific attributes common to private communication platforms yet absent from public platforms are able to facilitate perceived social support and mitigate feelings of loneliness among. In this study we see a trend where increased self-disclosure across private platforms correlated significantly with feelings of social support (as shown in Table 3.2). Thus we speculate that self-disclosure may be a factor impacting positive social support outcomes. With the known positivity bias in public social media content [46, 133], it is likely that users rarely disclose their true emotional state in public online spaces. It may be that users are more likely to disclose their true feelings in private spaces, which leads to increased social support. Further investigation on the differences in disclosure across public and private platforms are necessary to better understand this trend.

3.5.2.1 Benefits of Written Communication in Messaging as Opposed to Phone or Video Calls

Media Richness Theory suggests that the more information a platform gives (or the "richer" the platform is), the better outcomes for users will be. Following this theory, and also intuitively, we would assume that the closer an online platform feels to face-to-face interaction, the more it would be used for self-disclosure. We would expect this pattern regardless of whether the platform is intentionally selected to be used for the purpose of disclosure, or whether the platform is already in use and provides a comfortable space for self-disclosure.

With this in mind, there are several auditory features often found during face-to-face conversation that are not captured by messaging but are present in phone calls, most obviously tone and verbal pauses. Video call takes this a step further and also provides facial expression and body language. We would therefore expect students to disclose most frequently and effectively over video call as it is most similar to face-to-face interaction. However, as shown in Table 3.2, amount of self disclosure over *messaging* was actually found to correlate most strongly with feelings of social support, with students who reported disclosing daily over one-on-one messaging reporting 1.21x higher social support than those who disclosed rarely or never. This pattern defies our expectation and implies there is some level of benefit to the *written communication* inherent to messaging but not present in phone or video calls.

We speculate that the strong relationships between messaging platforms and perceived social support may be due to the **persistent nature of written communication** when compared to spoken communication. Messaging leaves a tangible record of past conversations, providing a space to revisit and reflect on both ones emotional state and on support received through the platform. By allowing users to look back at a reminder of past support, messaging may amplify the impact of supportive conversations on perceived social support.

We also note that messaging may provide an additional layer of perceived privacy to conversations, which previous work suggests facilitates social support [2]. The lack of spoken responses means that conversation participants cannot be overheard by a third party. Additionally, the lack of auditory and visual cues in messages also allows users to present their thoughts without emotional tells, something that may also be privacy-protecting. While this may seem counter-intuitive, there are situations

where emotional responses can hamper ones ability to communicate. For example, at times a speaker may cry when angry, making it difficult to effectively verbalize their thoughts and potentially leading listeners to ignore or dismiss the speaker.

Finally, previous work has shown people are more likely to reciprocate actions of self-disclosure over methods that are least expected for self-disclosure [76]. As common intuition tells us that disclosure is more expected over phone calls or video calls, it follows that disclosure is less expected over messaging. This would make any disclosure over messaging seem more meaningful and a symbol of a closer relationship, something that is suggested by the Hyperpersonal Interaction Model [76]. It is likely additional reciprocation may also be tied to higher feelings of social support.

3.5.3 Divergent Effects of Online Communication on Loneliness and Perceived Social Support

Through our analysis of several factors that seem to influence the behavior and outcomes of online interactions during COVID-19, we found that levels of perceived social support and loneliness are not necessarily affected by the same factors during periods of physical isolation. Therefore, this section analyzes the differences in how online communication seems to impact social support versus loneliness. In particular, we note that while perceived social support seemed to be positively associated with factors such as private online interactions, one-on-one communication, and self disclosure, the same factors did not result in subsequent decreases in feelings of loneliness.

This matches previous findings of persistent levels of loneliness across a seven week lockdown period [22] despite evidence of increased online communication during this time [56, 113, 156]. In comparison, our findings show that in-person factors seemed to have a much greater effect on loneliness than social support. While individuals living alone reported being significantly lonelier than those living with others, we find no significant difference in perceived social support between those living alone and those living with others. This is consistent with the work of Elmer et al., which found that although students reported increased loneliness, there were no significant shifts in social support during the COVID-19 pandemic [51]. However, the causal relationship behind these patterns is still unclear. It is possible that persistent levels of loneliness may stem from loneliness being a shorter-term feeling based on current experiences, whereas perceived social support is a measure of an individual's existing support and community network. Therefore, an individual's perceived social support prior to the COVID-19 pandemic may serve as a deterministic factor in the level of social support they feel even during physical isolation. Individuals with large existing support networks may simply have a larger network of people with whom they feel comfortable interacting and self-disclosing, whether that is in person or online. On the other hand, as loneliness is an emotion that can change from moment to moment, it may simply be difficult to mitigate online without in-person interaction. As our study was limited to correlation data surrounding these factors, future research should take a causal approach to investigate the impact of online interactions on social support and loneliness.

3.6 Limitations

There were a few areas where this study was limited. First, there is a potential limitation in the recruitment method used. As we advertised exclusively through online channels (e.g., university email services and social media), it is possible that students without regular internet access would

be under-represented in this study. Furthermore, our social media reach was limited by our use of Facebook advertisements (including on Instagram) and Reddit posts, so students that do not use those platforms were under-represented in this study. Online recruitment allows researchers to get a large reach of participants, including those under strict quarantine restrictions, however, individuals who do not use social media at all or are unresponsive to email prompts may be excluded from this sample.

In addition, we filtered participants based on the question, "Are you currently enrolled in a college/university in the United States?". Filtering based on this criteria includes all active students in our sample. This entails that our sample may have included remote international students who are enrolled in US colleges and universities, but live abroad. These students may live under varying COVID-19 regulations depending on their home countries, so their experiences may differ from students living in the US. Furthermore, the sample may include students who are enrolled part-time, another group whose experiences may differ from full-time university students. Regardless, all participants in our study are active members of the student body, and are affected by the regulations imposed by US universities.

Another limitation to this study was the inability to determine causation between the variables analyzed. Due to the observational nature of this study, we were only able to identify correlations between variables. For example, findings that correlate higher levels of social support with higher levels of messaging cannot be generalized to say that messaging causes students to feel more socially supported. To determine causality, future work should consider randomized controlled trials to better understand the causal relationships between online interactions and mental state.

Lastly, as with all voluntary survey-based studies, responses to this study were all self-reported, which could lead to self-selection biases or inaccuracies in survey responses. Participants willingly elected to complete the survey online and their decision to participate could reflect inherent, biased characteristics. Furthermore, self-report affords the possibility that participants might exaggerate their scores or struggle to introspectively assess their behavior accurately. In order to combat this, the questionnaire underwent several rounds of pilot testing, and incorporated validation questions (e.g., "What is the name of the current global pandemic?") and reverse wording to ensure accurate results. Survey-based studies of this form are very common methods of measurement in modern research, and the mental health scales distributed all showed a high level of internal consistency.

3.7 Conclusion

Previous research reveals young adults interact on social media more than any other age group in the US [28]. With the major upheaval in college student life caused by the COVID-19 pandemic, we see a further increase in usage across online platforms during the COVID-19 pandemic [56, 113, 156], one echoed in our results. However, the increased physical isolation during the COVID-19 pandemic provided the unique opportunity to investigate whether online interactions are truly an effective substitute for in-person interactions in mitigating loneliness and fostering social support. Online interactions were found to have stronger, positive ties with perceived social support than loneliness. Although we cannot prove causation, private digital platforms showed promise for fostering social support in physically isolated individuals.

This study reveals the positive relationship between private online communications on student social support during the COVID-19 pandemic. Private platforms were found to have greater posi-

tive ties to social support than public social media. We find self-disclosure across private platforms held the strongest ties to perceived social support, specifically on messaging platforms as opposed to phone and video calls, perhaps due to opportunity for reflection and increased privacy that messaging provides. Overall, we see clear ties between online communication and feelings of social support, and suggest further study into self-disclosure specifically in private platforms.

Appendix

Emotions Disclosed Based on Frequency of Messaging

Differences In Emotions Shared Between Frequency of General Messaging Disclosure

Emotion	df	Mean Difference	SE	Mean Square	F	p	n^2	p_{tukey}
Depression	1	-1.01	0.10	152.58	106.61	<.001	0.15	<.001
Fear	1	-0.86	0.10	110.728	77.46	<.001	0.11	<.001
Anxiety	1	-0.89	0.10	118.07	84.40	<.001	0.12	<.001
Anger	1	-0.66	0.10	65.58	41.46	<.001	0.06	<.001
Happiness	1	-0.65	0.10	61.94	38.80	<.001	0.06	<.001
Jealousy	1	-0.57	0.09	48.90	43.37	<.001	0.06	<.001
Calmness	1	-0.62	0.09	57.46	44.72	<.001	0.07	<.001
Apathy	1	-0.64	0.09	60.51	51.70	<.001	0.08	<.001

Table 3.3: Differences in the specific emotions disclosed by those who reported disclosing over messaging platforms at least weekly (frequent) or those disclosing monthly or less (infrequent) were examined. One-way ANOVAs were conducted across each emotion, revealing significant differences between groups across all emotions. The mean difference above represents the (mean of frequent)-(mean of infrequent), indicating greater differences for more negative emotions (depression, anxiety, fear), with the exception of anger. This suggests that students reporting higher willingness to disclose online are likely referring to more negative emotions. As increased self-disclosure online was found to be associated with increased perceived social support, it is likely that negative emotions play a larger role than positive emotions in students' perceived social support online during the COVID-19 pandemic.

Chapter 4

Preliminary Work: Signals of Affect in Messaging Data

4.1 Introduction

As mentioned previously, a major part of daily communication in today's world is digital, particularly among young adults. Digital communication spans across many disparate platforms, including both public (e.g., Facebook, Twitter, Instagram) and private (e.g., texting, Facebook Messenger, Kik, WhatsApp) spaces. As discussed previously, the persistent written records of conversations made through these applications may provide a space for reflection and re-visitation to both one's own emotional state over time, and on the support we receive from others.

In this chapter, we explore the potential for these public and private online spaces to act as a space for connection and bonding through emotional sharing. With our prediction that written communication fosters feelings of social support, we would expect communication over public social media posts to similarly correlate with perceived social support, a trend that we do not see reflected in our results. We propose that this disparity may be related to the positivity bias commonly seen in public social media [46,133], as well as evidence that users often consider self-presentation more when posting publicly, or to a wider audience [17,83]. We question whether emotional self-disclosure over public social media is therefore distinct from users' actual emotional state, and less accurate than emotional self-disclosure over private platforms.

It has not yet been studied whether messages sent in private messaging spaces actually correlate with user emotional state. Thus in this study we investigate the sentiment implied through messages sent in private messaging spaces (as measured through existing sentiment analysis techniques, LIWC and VADER, as well as human review), and how well this sentiment correlates to users' self-reported affect rated on a scale of 10–50 (using the well-known PANAS scale [164]).

However, one major obstacle complicates useful programmatic analysis of personal messaging data. Extracting private messaging data from multiple platforms in a privacy-preserving manner is difficult without specialized technical expertise and preparation. Thus in this paper we describe the implementation and maintenance of a cross-platform messaging extractor. This extractor has been active for three years and has extracted data from over 350 individuals from a patient population with clinical collaborators; the data from those extractions are part of clinical research which is currently unpublished. A separate study using the same extractor was conducted with a population of college students, which is presented in this paper.

While we find that human review predicts affect more accurately than VADER over all cases of positive affect and anytime when the panel is confident, VADER predicts negative affect nearly as well as human review ($r_s = 0.28$ and $r_s = 0.29$, respectively). Interestingly, we find that LIWC, the most common technique used in literature, does not significantly predict affect in any cases. When compared to previous work, we see that VADER analysis on private messages predicts affect at a comparable level to similar analysis on more traditional spaces for disclosure (i.e. diary entries [153] and speech [35]). We also see that all analyses on private spaces perform better than analyses on data from public social media platforms, suggesting that users may be more **honest** in their self-disclosures over private communication spaces.

In this study we contribute a better understanding of self-disclosure trends in private online communication spaces, as measured by popular sentiment analysis techniques LIWC and VADER. While we do find and report on some discrepancies between these techniques and human review, our findings still imply that messages sent over private communication *more accurately represent true user affect* than conversations in public social media spaces.

4.2 Related Work

4.2.1 Self-Report Measurements of Affect

Previous works on sentiment analysis, as well as other work from psychology [150, 164] commonly express emotional state in terms of two separate quantities: positive affect and negative affect. Positive affect refers to the extent to which an individual subjectively experiences positive moods and emotions, such as joy, interest, and alertness [105], while negative affect refers to the experiences of negative moods and emotions, such as anxiety, sadness, fear, anger, guilt, and shame [145]. A commonly held belief is that these two metrics are independent: a high negative affect does not necessarily imply a low positive affect and vice versa [164]. Common techniques used to measure positive and negative affect include the Positive And Negative Affect Schedule (PANAS) [164].

The widely used 20-item version of the PANAS has been shown to be an internally consistent self-report assessment of affect [10, 154, 164]. It contains a series of 20 mood-descriptive terms (e.g., interested, determined, upset, ashamed) where participants are asked to rate their level of agreement with each on a 5 point Likert scale ranging from 1 = *very slightly or not at all* to 5 = *very much*. Of these, 10 terms are representative of negative affect, while the remaining 10 represent positive affect. Upon completion, the totals of all negative affect and positive affect questions are summed, producing two scores (one for positive affect and one for negative) with a minimum value of 10 and a maximum of 50 [164].

Given its wide usage and overall reliability, we use PANAS as a ground truth measure of affect in this study. We correlate the results of different sentiment analysis techniques with positive and negative affect scores found through PANAS.

4.2.2 Sentiment Analysis Techniques for Social Media Text

In this study, we compare ground truth PANAS scores to two automated sentiment analysis techniques known as LIWC and VADER. Linguistic Inquiry and Word Count (LIWC) is one of the most widely used automated sentiment analysis techniques [119]. Additionally, Valence Aware Dictionary for sEntiment Reasoning (VADER) is a sentiment analysis model specifically built for use in contexts such as social media posts [74].

Unlike PANAS, LIWC and VADER are indirect measures used to automatically estimate affect. LIWC is text analysis system developed in 1993 [120] popularly used for sentiment analysis. It is a commonly held standard in both fields of computer science and psychology. After its initial creation, LIWC has been updated a number of times to follow linguistic trends and improve its accuracy, most recently in 2015. LIWC contains a weighted dictionary of words, word stems, and a selection of emoticons. Each of these is labeled with a list of categories and sub-categories that identify, among other things, the affect implied by each word. When given a written text, LIWC compiles scores for it by summing the weighted values assigned to each word in the passage. The original set of words in LIWC's dictionary was based on common emotional and affect rating scales, including PANAS [120]. Updates to the scale since then have been based on cycles of expert analysis and further methods of relevant term discovery [147].

VADER is a freely available rule-based sentiment analysis tool built to improve upon existing techniques, including LIWC and human review [74]. VADER utilizes a larger dictionary of terms, but otherwise works very similarly to LIWC, with a particular focus on analyzing text from social

media sources. While not as popular as LIWC, VADER has been used in previous research as a sentiment analysis tool [18, 19].

4.2.3 Predicting PANAS using Automated Sentiment Analysis

Previous work has found significant but weak correlations between PANAS and automated methods of affect prediction discussed above (ie. LIWC and VADER). This is consistent across texts collected from a variety of textual sources such as social media posts and diary entries [18,19,34,153]. Although all of these previous studies used PANAS scores as a more general measure of affect rather than an in-the-moment measure, they still provide useful comparison points for our purposes.

4.2.3.1 Affect Prediction in Personal Writing and Speech

Tov et al. [153] investigated LIWC predictions of PANAS scores over 21 days. However, instead of using a single PANAS score to describe that period, participants completed the PANAS each night, reflected on their affect that day, and wrote two short diary entries about a good and a bad event from that day. At the end of the study, all of a participant's diary entries were combined and used to generate a LIWC score. This score was then compared in a correlative analysis with the average of all PANAS scores obtained by the participant over the same period, producing a moderate-to-weak, but significant, correlation (positive: $r=0.21$, $p<0.01$ and negative: $r=0.22$, $p<0.01$).

Cohen et al. [34] wanted to investigate methods of personality detection from autobiographical text. Participants were asked to speak for three minutes on the topic of their choice, though they were subtly encouraged to talk about themselves. After participants finished speaking, they were given a series of psychological tests, including PANAS-X (a version of the PANAS with 60 items instead of 20). Each speech was transcribed and used to generate a LIWC score. Further analyses were run from there, including a correlative analysis between LIWC and PANAS scores. LIWC was found to be moderately and significantly correlated with PANAS positive and negative affect scores ($r=0.29$, $p<0.05$ and $r=0.24$, $p<0.05$, respectively).

4.2.3.2 Affect Prediction in Social Media

Beasley & Mason [18] investigated the ability of LIWC to predict general affect from public social media posts on Facebook and Twitter. Participants were asked to fill out the PANAS with regard to their general affect. Researchers then collected as many of the participant's posts from Facebook and Twitter as the platforms allowed. PANAS scores were then compared to a VADER analysis over the text of all collected Twitter and Facebook posts. PANAS was also compared to LIWC analyses over all collected posts (going as far back as the platform would allow), and posts only in the month, 6 months, and year preceding the study [18]. For the most part, correlations between PANAS and LIWC and PANAS and VADER were weak, with a highest correlation of $r=0.13$,$p<0.01$.

Following that study, Beasley et al. [19] investigated whether pre-filtering Facebook and Twitter data would improve VADER's accuracy in predicting PANAS scores. Methods for PANAS score generation and Facebook and Twitter data extraction were the same as those in Beasley & Mason [18]. After posts were collected, they were filtered for posts containing pre-selected "patterns of expression," for example,

"am" (e.g., "I am") + optional up to two words (e.g., "not very") + [affect-related words] (e.g., "happy")

For each participant, only posts that met these filtering criteria were included in the text given to VADER for analysis. A correlative analysis was then performed for PANAS as compared to both VADER scores for Facebook posts, and VADER scores for Twitter posts of participants containing greater than 36 Twitter posts after filtering. Despite this, all correlations were still weak ($r < 0.15$). See Table 4.7 for exact values.

In this study, we similarly correlate LIWC and VADER scores with PANAS. However, unlike previous studies, we use private messaging data as input for LIWC and VADER as opposed to public social media, diary entries, or transcribed speech. We cross-compare our results with those of previous studies to find patterns in input data that detect affect more successfully.

4.2.4 Personal Disclosure Across Multiple Social Media Platforms

In order to effectively predict affect from social media data, it is first important to understand how people interact with and communicate over these platforms.

Most previous work exploring affect detection in social media focuses on a single platform during analysis. The majority of these works specifically analyze Twitter data [41, 44, 48, 77, 116, 128], but some studies examined platforms such as Instagram [127] and Facebook [39]. Still others focus on two or more social media platforms [18, 19], a decision that seems prudent given that as of 2018, the Pew Research Center reported that 73% of social media users use multiple social media platforms (Median = 3) [139]. Furthermore, the social media landscape makes rapid advances, and platforms that are popular now may fade out over time. For example, in 2015, 71% of teens ages 13–17 reported using Facebook, but by 2018, the percent of teens 13–17 dropped to 51% [9].

Previous research on sentiment analysis also tends to focus on public social media data. This assumes that people primarily post about their emotions publicly. However, studies have shown contradictory results on whether people are more likely to share emotions publicly or privately [16,97,116]. User preference for sharing emotions more openly in public posts or private messages seems to be influenced by platform [16,116]. When investigating trends in self-disclosure on Twitter, Park et al. found some individuals prefer to share their emotions in public Tweets, rather than in Twitter private messaging [116]. They attribute this to individuals using public spaces on Twitter as an area for emotional reassurance and self-expression. In contrast, Bazarova et al. found that Facebook users disclosed more sensitive information in private channels than in public posts [16]. They suggest that when people cannot control the target of their disclosure, such as in public posts, then users are less likely to disclose. Lottridge & Bentley similarly found that users were more likely to share news links with a goal to start conversations in private chats rather than publicly [97]. Furthermore, studies have shown that people often present themselves differently on their public social media in order to self-promote (e.g., [87]) indicating that public social media posts may not accurately reflect a user's emotional state. In this study, we sought to bridge a gap in the literature by investigating the performance of automated methods of affect prediction over private messaging data. We believe that private messages extracted from a variety of platforms may more closely predict affect than previous social media studies [18, 19]. In many ways, private messages are more similar to direct speech [34] than to public social media posts. So based on the success by Cohen et al., private messages may be a viable textual

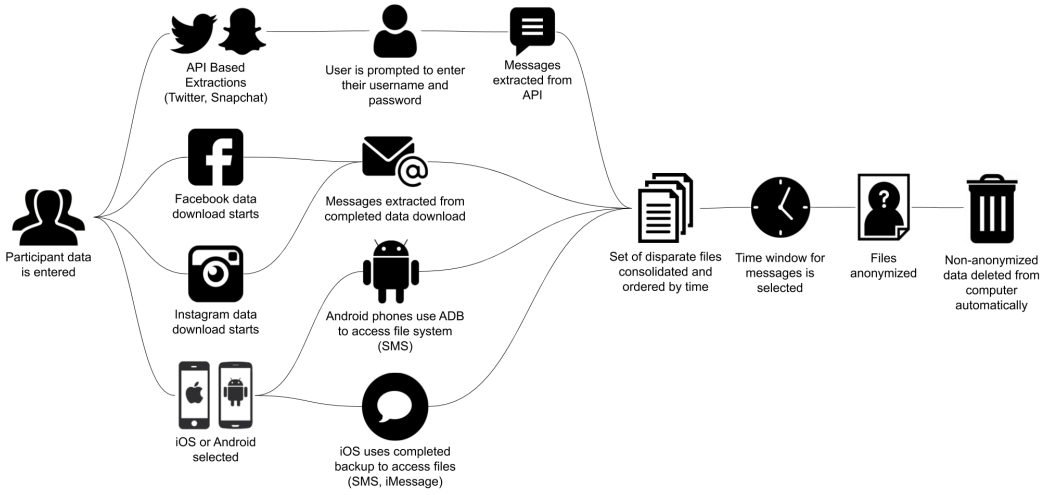


Figure 4.1: Sochiatrist extracts, consolidates, and pseudonymizes the data to develop models predicting affect based on messaging data.

source for affect detection. Furthermore, by extracting data from a variety of different platforms we can protect against the rapid changes in social media platform popularity and have a more holistic view of a user's communication, especially for the many users who do not use one platform exclusively.

4.3 Sochiatrist System

We developed Sochiatrist (a portmanteau of the words "social" and "psychiatrist"), an application that uses data-scraping methods to automate the retroactive extraction of a participant's messages and public posts from popular social media platforms (Figure 4.1). The system is retroactive in that it can collect social media data from within any specified past date range in a single run. It is non-invasive and does not require the participant to install any long-term software or tracking system. It is also built to be privacy-and-consent-first: participants must be physically present to input their login information for each messaging platform during an extraction and all extracted data is pseudo-anonymized. The system is unique because it allows efficient extraction of *both public posts and private messaging data* from many social media platforms. Code and instructions on using the Sochiatrist system will be released publicly at the completion of the clinical studies (currently funded by the National Institute of Health) that make use of the system.

4.3.1 Data Extraction Methods

The Sochiatrist system enables researchers to collect public posts and private messages from multiple web and mobile based platforms. Supported platforms have changed over time due to changes in data availability and social network popularity. At the time of writing this paper, Sochiatrist supports Facebook, Twitter, Instagram, Snapchat, Discord, and SMS/iOS Message extraction on both Android and iOS phones.

For platforms whose data is stored locally (e.g. SMS, WhatsApp), the Sochiatrist system extracts data directly from the phone, which must be connected to the computer being used for extraction.

The system creates a backup file from the phone and then reads this backup file to extract messages. For iOS devices, this backup file contains a `SQLite` database where text messages are stored. The location of this database is known from previous work in computer forensics to contain messaging data [36, 115]. For Android devices, the backup file is a CSV file that is read from the phone over ADB (Android Debug Bridge). In all cases, the participant must provide the password to their phone to allow this process to take place. Under no circumstance does extracting messages require rooting or jail-breaking, hard-to-reverse procedures that may damage a participant's phone, or unauthorized access to encrypted data.

For web-based platforms (e.g., Facebook, Twitter, Instagram, Snapchat), the Sochiatrist system downloads messages directly from the web, as a data download either provided by the website (Facebook, similar to Saha et al. [131], and Instagram), from an application programming interface (Snapchat), or directly from the website through the use of a web scraping script (Twitter). For all of these methods, the participant enters their username and password into the system and these credentials are used to provide the legitimate authentication required to gain access to the wanted data. There is no use of unauthorized access to data from any platform.

After extraction, messages from all platforms are compiled in a consistent format and sorted by timestamp. The user is prompted to specify a desired time range and only data within that time range is saved. To ensure privacy, the data is pseudo-anonymized (see Section 4.3.2.1) and all intermediary files created during the process are automatically deleted, leaving the participant's phone and the user's laptop untouched. The final pseudo-anonymized data is written to disk in CSV format on the computer running the application.

The Sochiatrist system has a graphical user interface. It displays clear, step-by-step instructions and allows for non-technical research assistants to run extractions with minimal training. It is also resilient to many types of user errors such as incorrect input formats or incorrect passwords, and attempts to provide helpful correcting instructions upon failure. Over the course of the past 3 years, the system has been used by clinical research assistants without computational backgrounds to successfully extract data from over 350 study participants (e.g., [64, 126]).

Currently, the Sochiatrist system is used in five different clinical studies. The system is fully supported by our team: we release regular updates with patches and offer technical support to resolve issues that we discover through dedicated automated bug reporting mechanisms.

4.3.2 Privacy and Consent

The Sochiatrist system is specifically designed to respect participants' consent and privacy. During data extraction, participants must be physically present to input their login information for each messaging platform used or the password to their phone, which builds participant consent into the core of the system. Special care is taken to ensure that there is no data persisted to the computer running the system that would ever allow a future unauthorized extraction. All data extractions are also legal and make use of legitimate authorization methods. There is a system that pseudo-anonymizes the final data, and all non-anonymized intermediary files are irrecoverably deleted. Table 4.1 shows an example output of the Sochiatrist system. For each message, the dataset includes the timestamp, the text of the message, whether the message was sent or received, an ID representing the sender's name, and which social media platform the message was exchanged on.

Timestamp	Status	Message	To/From	Platform
7/1/17 12:32	sent	I haven't been feeling great this past week	6af224	Facebook Messenger
7/1/17 12:32	received	Do you want to talk about it 7ac988?	6af224	Facebook Messenger
7/2/17 18:01	sent	Are you free at #:##?	72c7e0	Instagram DM

Table 4.1: Sochiatrist Data Extractor example output, demonstrating how messages are collected across platforms and how names are anonymized. These are example messages, not actual participant data.

4.3.2.1 Pseudo-Anonymization

Private messages are by definition personal and it is natural that participants may not want their identity or the identity of any of their conversation partners linked to extracted data. To respect this, for every message extracted, the "to/from" field is replaced by a randomly generated alphanumeric string. To maintain consistency, the same conversation partner is always replaced by the same string, although the same person messaging across different platforms will receive different identifiers for each platform.

However, there is also textual content in messages that may be identifying. As a simple example, if either conversational participant sends "Hi [name]", then identity is likely compromised. The free text in messages is therefore processed to remove names and numbers. The system has a constant database of common names that it augments with the participant's Facebook friends (obtained through the same data dump that contains Facebook messages) during anonymization. It uses this list to detect and remove names within the free text. As above, the names are replaced with a random alphanumeric string which is consistent across the use of the same name. Simple regular expression searches were used to detect different forms of the name such as possessives and different capitalization. All numbers in the free text are also replaced with the '#' character, which anonymizes shared phone numbers, account numbers, addresses, meeting times, and other such identifiers.

This pseudo-anonymization removes some sensitive information that messages can contain. However, this is of course an imperfect system and cannot be guaranteed to anonymize all identifying information in the extracted data, which is why we refer to it as pseudo-anonymization. This technique misses cases where people use nicknames for people that they message (e.g. Auntie, Honey), when people exchange messages with someone who has a common English word in their name (e.g. April, Hope), or cases when a name happens to not be in the custom dictionary of names generated. Locations are also not deidentified. However, the topic of how to completely anonymize free text is an extremely complex one, and it is our hope that as modern anonymization techniques improve so will the anonymization capabilities of our system.

4.4 Methods

We ran a study to investigate the relationship between affect and private messaging data, collecting private messaging data from a sample of undergraduate students at Brown University. Undergraduate students were studied due to the heavy usage of messaging platforms in this population [53]. Sentiment analysis and human review methods were used to estimate self-reported PANAS scores from their messages. The performance of these estimations were compared later during analysis. All procedures were reviewed by our institution's human subjects review office, (the IRB), and passed through a full board review (the highest level of review) on July 6, 2017.

Negative Words	Distressed	Guilty	Upset	Scared	Irritable
Positive Words	Alert	Excited	Strong	Inspired	Proud

Table 4.2: Examples of negative and positive words used in the PANAS survey. Participants are asked to fill out a Likert scale from 1–5 for each word to answer the question "to what extent do you feel this way right now"

4.4.1 Study Procedure

Participants were recruited using posters and Facebook posts in groups created for undergraduate students of various class levels. Participants were required to have an Android or iOS phone to join the study, and use at least one messaging application supported by the Sochiatrist system. Of 28 students who agreed to participate, 3 decided against participation during the study, 2 for privacy reasons and 1 for undisclosed reasons. Their data was not collected and thus excluded from the study. The final set of 25 participants included 20 females and 5 males, with an age range of 17–22 years of age (Mean = 19 years). 68% of our participants ($N = 19$) used an iOS device for the study while the other 6 used an Android device.

To maintain anonymity, we did not collect any personally identifying information beyond gender and age. Recruitment materials were distributed in four Facebook groups: one group for each undergraduate class. All students in each year received an invitation to their respective group upon university admittance, and these groups were commonly used for events, announcements, and general communication between students at the university. We therefore assume each group to be representative of the diversity of the student body as a whole, and we assume our sample to be similarly representative of the students.

We measured participants' affect through self-reported PANAS surveys collected via a method known as Ecological Momentary Assessment (EMA), commonly used in field studies pertaining to mood or affect [1, 166]. During EMA, participants are notified over text to complete a survey or task. This notification appears a set number of times per day, as chosen by the experimenter [135]. Notification time (and whether that stays consistent day-to-day) is also at the discretion of the experimenter. EMA aims to minimize recall bias, maximize ecological validity, and allow study of microprocesses that influence behavior in real-world contexts [136].

In line with typical EMA protocols, participants were prompted to complete the PANAS survey three times each day. The survey was sent to participants at a random point in each third of their day, based on their reported sleep and wake times. Links to surveys were sent via email or text, depending on the participant's preference, and were administered by an online Google form pre-filled with the participant's unique identifier. The survey presented each of the 20 PANAS attributes with a 5-point Likert scale, and asked participants how they felt at that specific moment in time. After receiving a prompt, participants had one hour to complete the survey and would receive a reminder prompt after 30 minutes if the survey was still not submitted. Extra PANAS surveys that were completed without a prompt were discarded from analysis.

The PANAS survey used for this study had 20 words, 10 measuring positive affect "PANAS(+)" and 10 measuring negative affect "PANAS(−)". Participants are asked to fill out a Likert scale from 1–5 for each word included in the PANAS in answer to the question "to what extent do you feel this way right now". The responses are simply summed up for a potential maximum score of 50 on each affect scale, and a minimum score of 10. Some examples of the words included in the survey are in Table 4.2.

We then collected participant's private messaging data from these two weeks using the Sochiatrist system, which consolidated the private messaging data they produced over the two weeks of the study. The ability to retroactively extract messages both simplifies the study procedure and reduces the risk of the study influencing naturalistic conversation behavior that would normally take place. This data includes messages extracted from the participant's online Facebook, Instagram, and Twitter services and messages extracted from WhatsApp, Kik and SMS (including iMessage) applications on the participant's Android or iOS phone. *Third-party messages (messages received by the participant) were not used during the analysis.* In other words, the conversations were only analyzed from the participant's sent messages.

Participants did not report challenges or objections with the Sochiatrist system, but rather, reported the study in general to be straightforward and understandable. That being said, there were two potential participants that decided not to continue with the study due to privacy reasons, and we must consider that participants in the study had already self-selected to be comfortable with sharing their data through the Sochiatrist system.

Upon completion of the study, participants were debriefed in the lab. They answered general questions about their opinions of mood tracking, issues they faced, and the process overall. Participants appreciated tracking their mood through EMAs and reported interest in continuing to track their emotional state, even after study termination. Finally, participants were compensated a maximum of $60 for their participation. Compensation was based on completing at least 95% of the PANAS surveys within the prompted 1 hour window ($35), the provision of some amount of social data ($5), and wearing the Microsoft Band (which was not used for this paper's analysis due to the focus on messaging data).

The study took place from November 14, 2017 to December 17, 2017. Over the course of the study, 1,009 PANAS surveys were collected. However, out of these, 55 were incomplete and were discarded. Survey compliance ranged from 81–100% with a median compliance rate of 98%. PANAS(+) and PANAS(−) scores were calculated from each of the 954 complete surveys. The mean PANAS(+) score was 23.9 and the mean PANAS(−) score was 16.9. See Figure 4.2 for a visualization of the distribution. These values are similar to the positive and negative affect population means estimated by Watson et al. [164] of 29.7 and 14.8 respectively. During the Sochiatrist download, 76,661 messages (Mean = 3,068 messages per participant, Median = 1,897, Range = 358–16,697) were collected across all participants. The messaging platform used by the most participants was SMS text messaging/iMessage ($N = 23$), while only one user received Twitter direct messages. Further messaging statistics can be seen in Table 4.3.

4.4.2 Data Processing and Session Generation

Third-party messages (messages received by the participant) were removed from all data collected. We matched each PANAS score with all the messages sent in a 24 hour window surrounding it ($\pm$12 hours of when the PANAS was completed); messages in one 24 hour window corresponded to a single PANAS score as its "session." We discarded all sessions that did not have at least one message sent on any tracked platform in the two hours surrounding its actual PANAS survey time, or where reported PANAS scores were both 10, as this likely indicated that the participant did not fill out the survey in accordance with their true affect. It is important to note that although we consider 24 hours of data in a session, the actual PANAS survey asked people how they feel *at that moment* in accordance with EMA.

Messaging Platform	Messages Sent	Messages Received	Total Messages	Participants
Facebook Message	16,506	24,348	40,854	22
Text Message	14,878	19,183	34,061	23
WhatsApp Message	329	1,206	1,535	4
Twitter DM	0	196	196	1
Instagram DM	15	0	15	5
Total	31,728	44,933	76,661	25

Table 4.3: Summary statistics of the messages send and received analyzed in this study, and the number of participants using each platform. Snapchat messages were unavailable at the time.

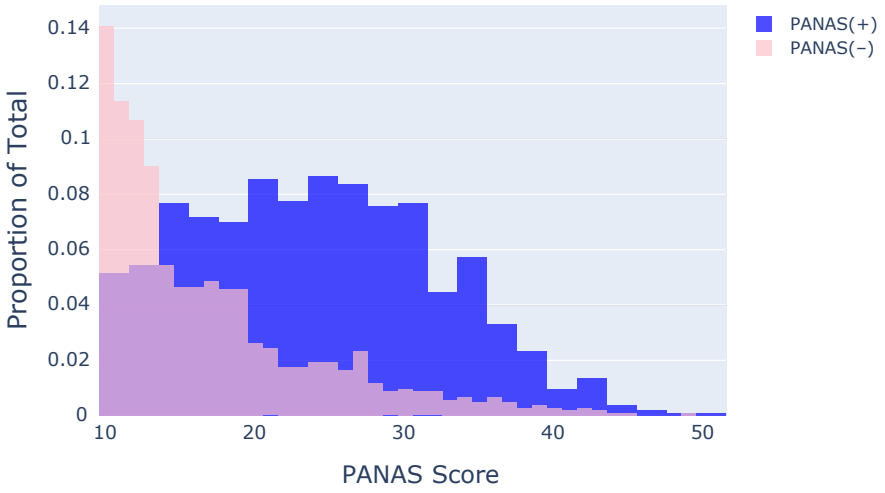


Figure 4.2: Histogram for PANAS(−) (pink) compared to the histogram for PANAS(+) (blue), with the overlap in purple. Note that PANAS(−) has a lower mean than PANAS(+) and is a more skewed distribution.

We chose to use a 24 hour window (±12 hours), as opposed to three intervals based on PANAS score report times for two reasons. First, we wanted the most accurate results from our human reviewers for a solid best case comparison. Human reviewers requested to see all messages occurring within one day of a PANAS score for more context, and we wanted to keep the time window for LIWC/VADER analysis consistent with human review for more accurate comparisons. Second, intervals between two PANAS score measurements varied due to the randomized prompting of the EMA (some as short as two hours) so some sessions would include a disproportionate number of messages due to the uneven windows, especially if a session included a period of sleep.

We also note that during investigation using shorter time periods for LIWC and VADER analysis (±2,4,8 hours) we found that the longer interval (±12 hours) correlated most strongly with the ground truth PANAS scores. Of the shorter intervals, PANAS(−) correlations were comparable (within 0.05 for LIWC, within 0.1 for VADER) to when sessions used a ±12 hour time period. PANAS(+) correlations for the ±8 hour set were also comparable to ±12 hours (within 0.01), but ±2 and ±4 hours correlated much less closely to PANAS(+) (a difference of more than 0.12).

4.4.3 Automated Sentiment Analysis

Two methods of sentiment analysis, LIWC and VADER, were used and their accuracy was compared. These are the most common methods used for identifying sentiment in social media messages (see Related Work).

4.4.3.1 LIWC

The authors of LIWC describe it as "...the gold standard in computerized text analysis. Learn how the words we use in everyday language reveal our thoughts, feelings, personality, and motivations." Of the 19 studies we reviewed pertaining to online communication and emotional state, 13 of them used LIWC as a measure of affect [18, 39, 42, 43, 44, 48, 84, 85, 86, 106, 116, 128, 153]. It is a commonly held standard in fields of both computer science and psychology as a sentiment analysis tool. Due to its widespread usage in research as a sentiment analysis tool, we included LIWC as a key estimator for affect from private messaging data.

The LIWC2015 tool was used to analyze all the messages in each entire session. For each session, the tool outputs many metrics. We took the `posemo` score as the PANAS(+) estimate and the `negemo` score as the PANAS(–) estimate. These scores are not on the same 10–50 scale that PANAS uses and do not necessarily scale linearly with PANAS scores. For this reason, all correlations computed later are non-parametric (i.e. Spearman's correlation).

4.4.3.2 VADER

Developed after LIWC, VADER was an attempt to produce a rule-based sentiment analysis method that was optimized for social media text, and is also popular in the field [74]. The authors of VADER release it as, "a gold-standard sentiment lexicon that is especially attuned to microblog-like contexts."

The NLTK distribution of VADER from the `nltk.sentiment.vader` package was used to analyze all the messages in each entire session. For each session, the tool produces a `pos`, `neg`, `neu` and `compound` score. We took the `pos` score as the PANAS(+) estimate and the `neg` score as the PANAS(–) estimate. As in LIWC, these scores were not scaled in any way.

4.4.4 Human Review Process

Human Review is important for two reasons. Firstly, analysis of the differences between human review and sentiment analysis methods may provide insight into what may be lacking from these methods in PANAS predictions so that we may build better estimators. Automated sentiment analysis can reveal how existing methods perform on textual message content, but does not project what is possible with additional context from the conversation. Human reviewers reading the conversation can get a broader understanding of the pace of conversation from timestamps, tone, and changing topics, and can even infer the relationship of the participant and conversational partner. Secondly, it allows us to calibrate the task. With access to just a snippet of a conversation—no context about the person, the scenario, etc.—it may be extremely difficult to predict a PANAS score. Human labels may allow us to understand better what is perceived as good performance on this task and what we may hope to achieve in the future.

A group of three reviewers estimated PANAS(+) and PANAS(–) from a sample of sessions. These three reviewers are authors on this paper, but they did not participate in running the study nor did they interact with any of the participants. To select the sessions, a stratified sample of sessions over participants was taken by randomly selecting between 4–6 eligible sessions per study participant (for a total of 110 sessions). There were a few restrictions to prevent bias in the reviewers labels. Any session previously rated by any reviewer (in tests or pilots for reviewing) was removed from the set of eligible sessions. Any session with fewer than two messages within two hours of the PANAS rating time were also removed from the set of eligible sessions. Lastly, in order to prevent large of amounts of overlap in the text between rated sessions, we only included one session per participant per day in the set of sessions to be labeled. We selected this set of sessions to get as broad a range of different participants as possible without over representing any single participant, and subsequently over representing one particular texting style.

When labeling a session, human reviewers were shown the timestamp of the PANAS score along with the corresponding 24 hours ($\pm$12 hours) of messages. For each message, reviewers could see the text, its timestamp, the anonymized numeric participant ID, and the anonymized alphanumeric ID for the conversation the message was a part of. In addition, reviewers knew the population means for positive and negative affect of 29.7 and 14.8, respectively, as estimated by Watson et al. [164]. Note that this is not the same as the sample mean of PANAS(+) or PANAS(–) from the sample of sessions labeled, or even the sample mean from all the sessions present in our data.

To start, all reviewers familiarized themselves with the PANAS survey. As mentioned above, they were given the mean PANAS(+) and PANAS(–) scores estimated by Watson et al. [164]. Before any discussion, reviewers also each individually rated a set of messages with a shorter time window ($\pm$2 hours) as a test. Reviewers were shown the scores of the other reviewers, as well as the true PANAS scores, for the sessions in this test set. Following the test, reviewers requested that a full day of messages be included in a session for context, and that they would prefer to be able to discuss scores with another before deciding on a rating. The set of sessions in this test set was fully disjoint from the sessions used in final analysis.

For the final human review process whose data was included in this study, all three reviewers were placed together in a video call. With the data described above available to them, each reviewer individually proposed overall PANAS(+) and PANAS(–) scores (each between 10 and 50) for a given session. When reviewers disagreed about the values they proposed (which happened in the majority of cases), they were given the opportunity to simply accept another reviewer's proposed score. If a unanimous decision was still not reached at this point, then each reviewer explained their reasoning for the scores they proposed. One reviewer would then propose a new set of PANAS predictions based on the given explanations. The other reviewers could then accept the proposed values or propose an alternate set of scores, with explanation if appropriate. Reviewers repeated this last step until unanimous consensus on positive and negative affect scores were agreed upon, resulting in two final scores, one for PANAS(+) and one for PANAS(–) . Time to reach consensus averaged about four minutes. Scores were predicted on a scale of 10 to 50 to be consistent with the PANAS scale. Reviewers looked for cues in content and tone, as well as time of day the survey was taken, time of year, and their own previous experience with the situations described in the text. When reviewers struggled to identify any signal in a session, they rated it as "neutral", falling back on population means mentioned above. Reviewers also each took individual notes on each EMA, which included the group's reasoning, as well

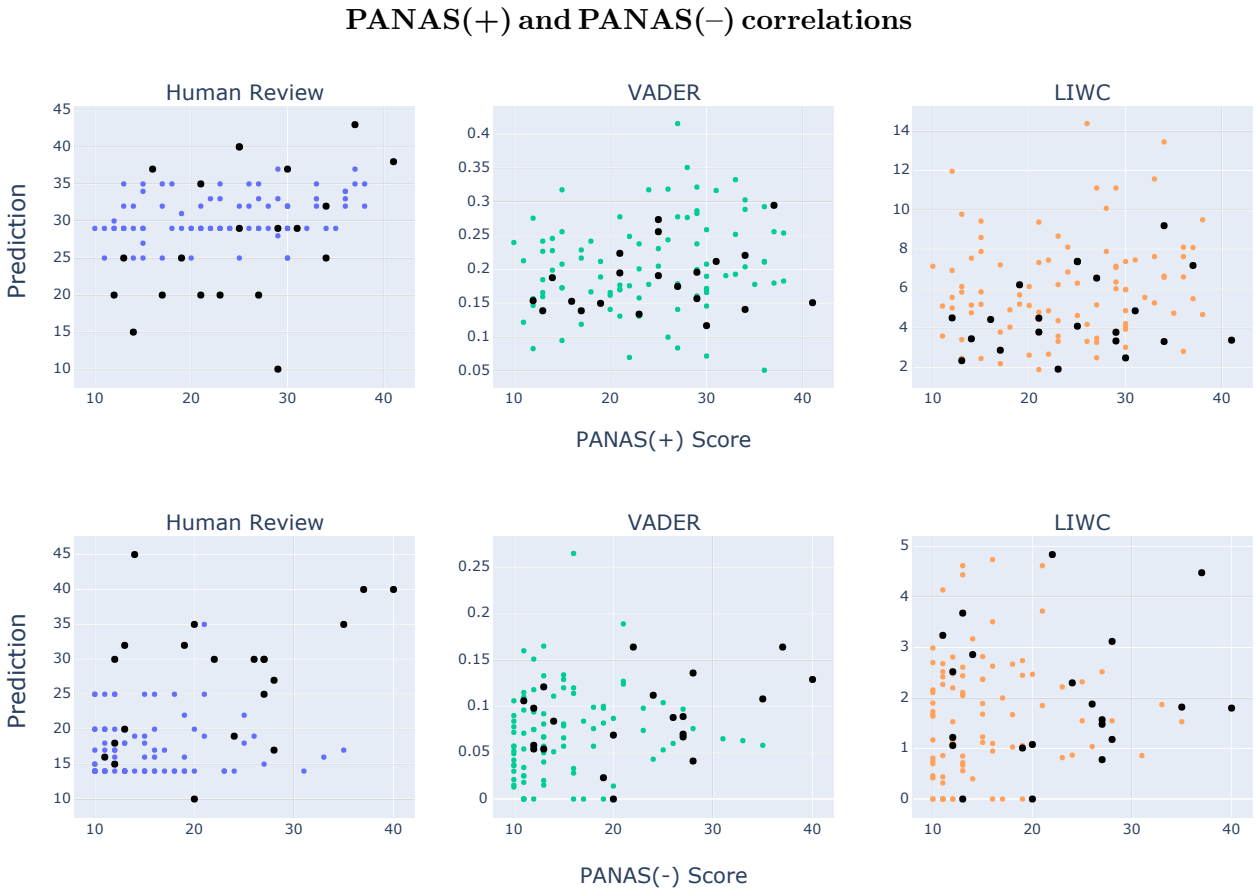


Figure 4.3: Self-reported PANAS(+) and PANAS(−) scores plotted against predicted scores. The x-axis is PANAS(+) in the top plots and PANAS(−) in the bottom plots while the y-axis is the algorithm prediction (each technique uses a different scale). The dots in black are the points labeled confident by human reviewers.

as any particularly important aspects of the text. Finally, reviewers came to a unanimous decision as to whether they felt confident in the rating or not. Confidence was a binary outcome (yes/no). It was recognized that confidence would be a subjective decision and there were no strict heuristics used by reviewers. Some things they expected to use to determine confidence were the strength of a particular emotion and how clearly the participant expressed their feelings in their messages.

4.5 Results

We use LIWC and VADER as our primary systems for sentiment analysis in this study, as they have been used in previous studies and therefore offer a fair comparison [18, 19, 34, 153]. We also include human review as a baseline "gold standard" sentiment analysis technique with which to confirm the accuracy of LIWC and VADER. Correlations were calculated using the non-parametric Spearman's correlation due to the lack of a linear relationship between the sentiment analysis scores and PANAS (thus, a Pearson's correlation is not used). Predicted versus actual PANAS score comparisons for each sentiment analysis technique can be seen in Figure 4.3.

4.5.1 Individual Performance of Techniques

Tables 4.4 and 4.5 show both individual performance measures (how well a sentiment analysis technique correlates with PANAS affect scores), as well as the correlation between techniques.

Self-reported positive and negative PANAS scores, PANAS(+) and PANAS(–), were treated as the ground truth for the participant's affect. In absolute terms, PANAS(+) scores tended to be higher ($Mean = 23.95$, $SD = 8.20$) than PANAS(–) scores ($Mean = 16.87$, $SD = 7.26$). Between the two scores, PANAS(+) and PANAS(–) scores had a weak, insignificant correlation (Pearson's coefficient $r = 0.16$, $p = 0.09$). This reflects the expected independence of these measures, where high positive affect does not necessarily mean having low negative affect and vice versa.

For the stratified sample of 110 sessions chosen for analysis, PANAS(+) ($Mean = 24.04$, $SD = 7.87$) was higher than PANAS(–) ($Mean = 16.51$, $SD = 6.88$). For the subset of sessions that human reviewers reported confident, PANAS(+) was slightly higher than the 110 analyzed ($Mean = 27.70$, $SD = 9.04$) and PANAS(–) was substantially higher ($Mean = 27.75$, $SD = 9.25$). Other techniques were used to predict the PANAS affect scores with the messaging data, and are reported in the following subsections.

4.5.1.1 Human Review

The three human reviewers examined the timestamped private messages for the sampled 110 sessions and attempted to predict their associated PANAS(+) and PANAS(–) scores, marking 21 of these sessions as confident. As expected, these ratings moderately predicted the PANAS(+) and PANAS(–) from the user-reported EMAs. For PANAS(+), human review was more accurate compared to automated techniques, with a correlation ($r_s = 0.31,\ p < 0.01$). For PANAS(–), human review still correlated moderately ($r_s = 0.28,\ p < 0.01$).

When only the sessions where reviewers were confident of their scores ($N = 21$) are considered, human review outperformed all other analyzed techniques when estimating PANAS(+). Among these sessions, the human review estimates correlated more strongly with PANAS(+) than when not confident ($r_s = 0.45,\ p = 0.04$). PANAS(–) estimates showed similar behavior ($r_s = 0.38,\ p = 0.09$), though this correlation was not significant. The human reviewers performed much better when confident than when not confident, meaning they are able to judge whether there was enough signal in the conversations to accurately predict affect. When reviewers were less confident, their accuracy was relatively inconsistent. Reviewers' ability to predict PANAS(+) remained moderately accurate ($r_s = 0.28,\ p < 0.01$), while ability to predict PANAS(–) dropped sharply ($r_s = 0.11,\ p = 0.29$). These findings indicate that although the task of generally predicting affect is difficult, there may be identifiable scenarios where the task is more tractable.

Reviewers tend to be more confident when PANAS(–) was higher. The mean PANAS(–) for confident predictions is 27.0 (SD=9.45) , as compared to the mean for not confident predictions of 17.1 (SD=3.97) . However, this trend does not hold true across PANAS(+) scores. This may be due to the fact that all human language has a known positivity bias [46], meaning that negative speech is relatively rare. Therefore, when negative emotions are expressed, it may become a more clear indicator of a high PANAS(–) that raters are able to detect and feel confident in their detection.

	VADER	LIWC	Human Review		
			All	Confident	Not Confident
PANAS(+)	*0.20	0.14	**0.30	*0.45	**0.28
VADER	-	**0.75	**0.27	0.31	*0.22
LIWC	-	-	**0.29	0.36	*0.24

Table 4.4: Spearman's correlations between various techniques for estimating *positive* affect scores from PANAS, i.e. PANAS(+). Human review performs best when the reviewers feel confident about their estimate. LIWC and VADER are highly correlated, but less correlated with human review. They are less correlated with PANAS(+) compared to human review with confidence ignored ("All" column). *$p<0.05$, **$p<0.01$

	VADER	LIWC	Human Review		
			All	Confident	Not Confident
PANAS(–)	**0.28	0.17	**0.29	0.38	0.11
VADER	-	**0.77	**0.31	0.32	*0.26
LIWC	-	-	*0.22	0.31	0.19

Table 4.5: Spearman correlations between various techniques for estimating *negative* affect scores from PANAS, i.e. PANAS(–). Human review performs better when the reviewers feel confident about their estimate, although sometimes not statistically significant. LIWC and VADER are highly correlated, but less correlated with human review. Even so, VADER performs similarly to human review with confidence ignored ("All" column) *$p<0.05$, **$p<0.01$

4.5.1.2 LIWC

When trying to predict affect, LIWC weakly correlated with both PANAS(+) and PANAS(–) scores (not statistically significant in both cases). Over the 110 sessions in our randomly sampled testing set, LIWC received a Spearman's $r_s = 0.14$, $p = 0.14$ with regard to PANAS(+) scores, indicating a non-statistically significant but weak correlation. For PANAS(–) scores, LIWC received a Spearman's $r_s = 0.17$, $p = 0.08$, which is a similarly weak, not statistically significant correlation.

This may indicate that words used at a specific time have some correlation to the affect of a person at that same time. There seems to be some relationship between the messages an individual sends and their affect, but it is hard to draw conclusions in this case given the lack of significance. Given that LIWC performs significantly worse than human review at detecting affect, we do not use LIWC as a representative measure of existing sentiment in private messages.

4.5.1.3 VADER

Across both PANAS(+) and PANAS(–) scores, VADER performed moderately well. In particular when PANAS(+) scores were treated as ground truth, VADER outperformed LIWC, though both were outstripped by human review, earning a Spearman's correlation of $r_s = 0.20$, $p = 0.03$. VADER exceeded our expectations across ground truth PANAS(–) scores, obtaining $r_s = 0.28$, $p < 0.01$ and performing better than LIWC and nearly as well as human review.

This indicates that VADER is a clearly better choice than LIWC for identifying affect from private messages. It suggests that VADER approaches a potential upper bound for how well an automated

technique can perform given that VADER analysis approaches human accuracy, especially in the case of negative affect detection. Even when selecting only the most confident of predictions for negative affect, human reviewers only reach $r_s = 0.38$ correlation, which is only slightly better than VADER's prediction. The similarity in accuracy to human review indicates that VADER's prediction of affect is a valid measure of actual affect presented in private messaging.

4.5.2 Comparative Performance Between Techniques

Aside from their correlation with PANAS scores, each technique can be compared with each other to identify discrepancies. We can use these discrepancies to detect how we might be able to improve the techniques in the future.

LIWC and VADER were found to be significantly strongly correlated across both PANAS(+) ($r_s = 0.75$, $p < 0.01$) and PANAS(–) ($r_s = 0.77$, $p < 0.01$) scores. This is as expected, since VADER was originally partially based on LIWC's lexical dictionary. Furthermore, since both techniques were developed with human predictions of textual indications of emotion [74, 120], we would assume both to have high correlations to human review scores. In accordance, both LIWC and VADER correlated more strongly with human review than with PANAS scores when taking into account all samples, especially for PANAS(+) scores (Tables 4.4 and 4.5).

LIWC's correlation with human review was found to be statistically significant in PANAS(+) ($r_s = 0.29$, $p < 0.01$) and PANAS(–) ($r_s = 0.22$, $p = 0.02$). VADER followed this trend, obtaining a statistically significant correlation in both PANAS(+) and PANAS(–) ($r_s = 0.27$, $p < 0.01$, and $r_s = 0.31$, $p < 0.01$, respectively), as well as outperforming LIWC in the PANAS(–) case.

Similar trends of LIWC and VADER prediction of PANAS scores were seen when comparing sessions where human raters were confident in their scores. Although some were moderate, none of these correlations were found to be statistically significant. This lack of significance is likely due to the low statistical power due to the lack sparsity of confident ratings ($N = 21$). In these sessions, human reviewers predicted PANAS scores more accurately than both LIWC and VADER. VADER correlated more strongly with PANAS(–) ($r_s = 0.35$, $p = 0.11$) than PANAS(+) ($r_s = 0.25$, $p = 0.27$). VADER also outperformed LIWC in these cases.

Interestingly, unlike VADER, LIWC correlated more strongly with confident scores in PANAS(+) than in PANAS(–) ($r_s = 0.16$, $p = 0.48$ and $r_s = 0.08$, $p = 0.71$, respectively). Despite its similarity to VADER, it is notable that LIWC severely under-performed comparatively.

In comparison, across sessions when reviewers were not confident, the accuracy of reviewer scores and LIWC and VADER's predictions dropped for the most part (LIWC has a high correlation to PANAS(–) on the not-confident sessions). As before, for PANAS(+) VADER ($r_s = 0.21$, $p = 0.05$) outperformed LIWC ($r_s = 0.14$, $p = 0.19$), but both performed worse than human reviewers ($r_s = 0.28$, $p < 0.01$). Note that both human review and VADER correlated significantly, but LIWC did not. However, for these sessions where human reviewers were not confident, VADER actually predicted PANAS(–) ($r_s = 0.26$, $p = 0.01$) better than both LIWC ($r_s = 0.18$, $p = 0.08$) and human review ($r_s = 0.11$, $p = 0.29$). This means that human reviewers' relatively successful performance for PANAS(–) prediction was heavily dependent on sessions with confident ratings (21 of 110 sessions). Also notable here is that, despite the relatively high number of sessions included in the set ($N = 89$), human review correlation with PANAS(–) was not significant (Tables 4.4 and 4.5).

VADER's overall higher performance in comparison to LIWC over all PANAS score predictions

Session	PANAS		Human		VADER		Select example illustrating different types of mispredictions
	(+)	(−)	(+)	(−)	(+)	(−)	
410	34	35	25	35	0.14	0.11	"...i really wish i could spend less time venting but i dont know how to not breakdown... i am just emotionally mentally and physically exhausted..."
707	29	37	10	40	0.20	0.16	"i cried myself to sleep for like 30 minutes once u left ... yesterday", "during the day i'm just feeling hopeless"
258	14	11	32	14	0.20	0.10	"HAPPY BIRTHDAY", "Amazing! I think it's great so I hope they will too!! Yayyy!"
668	18	25	35	22	0.16	0.10	"I'm sad and tired just laying in bed" and "I almost started sobbing"
749	13	28	25	27	0.14	0.04	"I tried chatting with my professor today about getting my paper deadline pushed", "I'm asking my chiropractor... she can probably get permission quicker than my physical therapist"
948	13	15	35	14	0.17	0.06	"Guess who might have just won ##$??", "bro thats rough", "don't worry I'm going home"

Table 4.6: Perturbed examples where the PANAS scores by participants differed from those decided during human review. In these sessions, the PANAS(−) scores were similar, but the human labeled scores were lower than PANAS for sessions 410 and 707, higher in sessions 258, 749, and 948. All examples have been perturbed for privacy. Note that the average PANAS(+) score is 29.7, the average PANAS(−) score is 14.8, and both have a range of 10–50. Values above and below these averages can be treated as "relatively high" and "relatively low" values for both PANAS and human review scores. Though VADER is on a different scale (from 0 to 1), values above or below its means (0.21 for PANAS(+) prediction and 0.07 for PANAS(−)) can also be treated as "relatively high" or "relatively low".

raises several questions. One might expect LIWC and VADER to perform comparatively given their similar structure and the high correlation between their scores. However, because VADER is specifically optimized for social media [74], VADER is likely better than LIWC at interpreting patterns of speech common to social media platforms, and therefore more closely mirror patterns of human review.

4.5.3 Mispredictions Across All Techniques

Differences in LIWC, VADER, and human review ratings can occur for a variety of reasons. While VADER and human review showed similar correlations with actual user affect, their correlation with each other was not as strong as expected. This indicates that while automated techniques and human review are likely picking up on different attributes of communication to identify affect. Here we have identified several cases where we observed common mispredictions for each of the three techniques when predicting self-reported affect. To pinpoint mispredictions, we manually reviewed the performance of each metric. Particularly, we selected sessions where sentiment analysis methods were highly unsuccessful in emotion prediction, or where different methods arrived at substantially different scores for the same session. We then conducted a qualitative analysis of message content for these selected sessions and found several common patterns at points of inaccurate prediction.

4.5.3.1 Over-reliance on Tone or Context

LIWC and VADER function fundamentally by making predictions based on the vocabulary and tone used in a session. Human reviewers, however, rely much more heavily on context and situational

content when trying to predict. Both of these focuses can lead to major errors in affect prediction, depending on the session.

We find exclusive reliance on tone, as LIWC and VADER do, to be an inconsistent indicator of true affect values, likely because it does not scale for situational context. We see many cases where human reviewers note using contextual clues to help predict affect, and then proceed to predict much more accurately than LIWC and VADER on the same session. One example of this is session 749, where a participant is trying to reschedule a paper deadline, and requires a dean's permission and a doctor's note, as shown in Table 4.6. Since the language used is not angry and contains few traditionally negative words, LIWC and VADER predict a low PANAS(–) score while the human review accurately predicts a high PANAS(–) score, based on the shared knowledge that situations such as these are often very frustrating.

In contrast, human reviewers on occasion focused too much on message content, especially when confident. For instance, in session 948 (shown in Table 4.6) one participant mentioned winning money from a study they had taken part in, and using it to buy gifts for their parents. The reviewers rated this as high PANAS(+), and noted that they had imagined the participant would be excited over winning money. However, the participant had an unusually low PANAS(+), a fact that was identified by the automated techniques. While the human reviewers made an assumption about the participant's mood based on the content of the conversation they read, LIWC and VADER's focus on sentence tone was able to more accurately predict ground truth scores.

Further investigation into additional latent clues for affect may be warranted. As a result, it is unclear which technique is better in this sense, as both have clear shortcomings.

4.5.3.2 Message Weight Based on Temporal Distance

Human reviewers reported placing very different weight on messages based on their temporal distance from PANAS completion, giving messages closer to PANAS completion time greater weight than those that are further. LIWC and VADER, on the other hand, place equal weight on all messages in a given time frame.

We would intuitively assume that messages sent very long before or after a PANAS measurement would have little bearing on the measurement itself. And indeed, there are instances where we see that discounting very distant messages seems to improve prediction accuracy. However, it is important to also consider situations where very distant messages can provide evidence of the actual affect of a participant. For instance, in one session reviewers noted that the message, "I just wanted to make myself feel a little better" was noted as indicative of negative affect, but was ultimately discounted when making their prediction because it was sent the previous day, and the tone had become less negative in recent messages. The true PANAS(–) was higher than reviewers predicted. It may be that this message was indicative of a broader negative mood state, and human reviewers missed the signs of this. LIWC and VADER, however, would perform better in these situations as they do not discount statements based on temporal location.

4.5.3.3 Inability to Dissociate Positive and Negative Affect

As mentioned earlier, PANAS(+) and PANAS(–) are considered independent dimensions of affect [164]. While intuitively we would assume a high PANAS(–) to indicate a low PANAS(+) and vice

versa (i.e. a negative correlation), this is not found to be the case with actual PANAS scores. PANAS scores in our data actually showed the opposite with a positive correlation ($r = 0.20$, $p < 0.01$).

Despite this, we found a negative correlation between predicted PANAS(+) and PANAS(–) scores for all sentiment analysis techniques. This correlation was particularly prevalent in human review ($r = -0.51$, $p < 0.01$), though LIWC and VADER both had weak, non-statistically significant, negative correlations between PANAS(+) and PANAS(–) ($r = -0.06$, $p > 0.05$, and $r = -0.16$, $p > 0.05$, respectively). The general failure to dissociate positive and negative affect is clearly exemplified in situations where participants report high scores for both PANAS(+) and PANAS(–). For example, in session 410 shown in Table 4.6, the participant messaged, "i am just emotionally mentally and physically exhausted." Despite this they reported both a high PANAS(+) and a high PANAS(–). Both human reviewers and VADER underestimated PANAS(+) scores for this session, likely due to its explicit account of negativity. Situations such as this one explain many of the mispredictions in human review ratings. The fact that human reviewers were the most susceptible to this type of error may have interesting implications in the clinical setting: would a licensed psychiatrist make the same "mistake"?

4.6 Discussion

4.6.1 Reflection on Affect Predictions

4.6.1.1 Features of Private Messaging Shared with Other Successful Textual Sources

When compared with previous research on affect detection, private messaging is found to be a promising textual source for accurate affect detection. As shown in Table 4.7, previous research comparing LIWC or VADER to PANAS scores has used public social media data [18, 19], personal diary entries [153], or spoken responses [34]. Across multiple studies examined, social media posts were consistently outperformed by all private sources. This may indicate a level of emotional distancing from public social media posts as opposed to something more private like a diary, or something relatively unfiltered like speech. Overall, direct speech and our VADER analysis over private messaging data were the most accurate predictors of affect, aside from human review.

Across all studies discussed, LIWC analysis over natural speech performed with the highest accuracy [34]. Private messaging data and direct speech share several key components, which likely contributes to their increased reflection of affect. For instance, both encourage rapid communication and have an assumed lack of statement revision. A study conducted by Lyddy et al. revealed texts to serve as a method of fast, unedited communication, with the average word count being 14.3 words and 17% containing misspellings (SD=12.0) [98]. Thousands of initialisms (e.g. LOL, brb, and ttyl) have been developed to aid in communicating quickly and further limiting a need for revision.

Because **private messages are less likely to be revised** to fit a wider audience, messaging data is likely to be more reflective of true affect than public data. When users revise statements, original emotional disclosures and tone may be edited out. We speculate that one of the main reasons LIWC and VADER were able to detect affect from direct speech was because speech affords little time for editing and influencing tone. In contrast, public posts on social media may be edited heavily before posting. This calls for further investigation into the influence of statement revisions on the accuracy of sentiment analysis techniques, as well as further explanations for the strong reflection of affect found in private messages.

	Beasley & Mason				Tov et al.	Cohen et al.	Beasley et al.†		This Paper	
	LIWC		VADER†							
	Facebook	Twitter	Facebook	Twitter	Diary Study	Speech	Facebook	Twitter	LIWC	VADER
PANAS(+)	-0.07	0.05	0.05	0.07	**0.21	*0.24	**0.14	*0.14	0.14	*0.20
PANAS(–)	-0.11	0.04	0.02	0.09	**0.22	*0.29	**0.13	**0.16	0.17	**0.28

Table 4.7: Different sources of text data from various prior literature report low correlations between PANAS and LIWC; † indicates a correlation between PANAS and VADER rather than PANAS and LIWC (Beasley & Mason, Beasley et al.); Note that Beasley & Mason achieve slightly higher correlations when using a wider time range than the one presented here. We only include the results for the time range shown as it is the time range most similar to ours. Compared to other studies, messaging data analyzed in our paper achieves a similar accuracy range with LIWC, but notably better with VADER, especially for PANAS(–). *$p<0.05$, **$p<0.01$

The second most successful technique found in previous studies was LIWC analysis over short daily diary entries [153]. The shared relative success of our method and this one may be due to the similarities between diary entries and private messages. Both are private methods of written communication and thus have a level of perceived control the author has **the ability to select the audience that will see the text**. While messages can be forwarded or shared with others outside of the conversation, users in a private messaging conversations still decide who sees the message. This is especially apparent when compared with public social media posts, which are visible to anyone, even people entirely outside of the poster's network.

In addition to the previous points about rapid communication and lack of revision, this idea of privacy and control can also be extended to the Cohen et al. speech study based on how it was conducted [34]. Participants were speaking directly to a single group (i.e. the researchers), and could reasonably assume that no one but the researchers would ever have access to that information. It may be due to the combination of all of these factors that LIWC analysis over natural speech by Cohen et al. was so successful.

4.6.1.2 Suggestions for Future Work

In this study, we see evidence that users are more honest in their self-disclosures in private messaging spaces than in public social media spaces. We attribute this trend to the fact that users in private spaces have control over the audience of their messages, leading to fast, less edited, and more honest disclosures. Our qualitative results indicate that while the automated techniques of LIWC and VADER relied more heavily on tone to correctly identify affect, human review put particular significance on decribed context. Given that VADER and human review both correlated moderately with ground truth affect scores, this may indicate that both implicit disclosures through tone and and explicit contextual disclosures are communicated through private messaging contexts. However, we note that our results are primarily correlational, and that further investigation on the impact of audience selection and control in private spaces on perceived social support would be prudent.

4.6.2 Ethical Considerations

A key requirement for the functionality of Sochiatrist is informed consent and a benefit to the user themselves. Participants are physically present for all data extraction, and they enter their own login information after consenting to extraction for each platform. The participants have an opportunity to review the messages that will be extracted, and to remove specific lines from the dataset. These

procedures are consistent with the human subjects full board review that occurred before the start of the study.

Studies in the past have addressed ethical concerns in a few ways. Most other studies collected publicly available posts only (e.g. [18,19,41,44,48]), others only collect data participants specifically record for the study (e.g. diary entries [153] or recorded speech [34]), while yet others ask participants to copy over messages by hand in the lab [17]. All of these methods aim to only collect data that participants are actively comfortable sharing with the researchers or the general public (in the case of public posts). However, we believe that by requiring participants to specifically opt-in to each individual platform, and also providing them the option to additionally remove any individual messages or messaging threads after extraction, achieves the same goal. This is supported by the fact that none of our participants mentioned objections or discomfort with sharing their data during post-study followup interviews.

It is important to note that only messages sent from a participant are included in this study. Therefore messages from other individuals to the participant (third parties) are not collected in this study, as those senders were unable to consent to message extraction. Additionally, pseudo-anonymization reduces the risk that participants expose third parties' names or identifying numbers in their own messages.

Once we move beyond this study and into day to day usage, this kind of data extraction can be used as a personal tracking tool. Users may use message extraction and analysis to track mood events and gain personalized insights into their own emotional triggers. Other mainstream self-tracking systems store data in remote servers, while the extraction and analysis for our system is local, and therefore can be used without an external party accessing the data. By storing this data locally, extracted data is no more accessible than users' private messages on applications they already have installed on their phones.

Clinicians can also benefit from a more overarching or long-term view of patient mood states. Sochiatrist or a system like it could be used by patients to generate data they are comfortable sharing. This would allow clinicians to gain deeper understanding of patients' affect without the need for invasive procedures like forcing them to explain and relive traumatic experiences or going through and monitoring patients' messaging history by hand. Patients may find it less stressful to present information to their therapists using a data extraction system such as Sochiatrist, rather than explaining a situation verbally.

4.6.3 Limitations

There were a few areas where this study was limited. First, we were unable to make a direct comparison between public and private messages to prove strongly our hypothesis that private messages may be a richer data source for estimating affect. To examine the specific utility of direct message data as a space for people to express their most intense and private emotions [16], we purposely chose to test the extreme case where no public data is used at all. In practice, Sochiatrist can extract both public and private data for analysis, but for this study, nearly 99.9% of the messages in our collected data from the participants are from private messaging sources and thus there is an insufficient sample of public messages to make any meaningful comparison between the two. However, the fact that public messages are such a rare occurrence in our study demographic is an indicator that further investigation into private messages may be required.

The study population was limited to college undergraduates in the United States, so due to differences in expression across cultures and age groups, the results may not be generalizable to other demographics. Additionally, the timeframe of this study could have introduced several confounds. Participant data was collected from November to December, around the end of the fall semester. During this period of time, many college students were likely to be focused on academics and exams, which could influence their affect and introduce bias into the data. However, we still believe that the methods and systems provided here make it possible to reproduce similar studies across other demographics.

We also may not have had access to all of participants' messaging history. There are common messaging apps that Sochiatrist does not support, such as Telegram and Signal. Our study design ensured that participants must use one of the Sochiatrist-supported applications, but does not guarantee that it is their "main" texting application or that there is not a significant amount of data missing.

The length of our study was shorter than previous studies on mood and affect, which had data collection periods that ranged from 30 days [100] to 3 years [134]. This means that we don't have a long enough timespan per-participant to train participant-specific models. Furthermore, the number of participants ($N = 25$) was also small. It is generally more difficult to collect a large dataset of private messages than a large dataset of public posts, since people are hesitant to share their private data. However, in the future it would be ideal to collect data over a longer period of time with a larger sample of participants.

Due to resource limitations, there was also not as broad a set of human reviewers as we would have liked. The same three reviewers, all from an American collegiate background, reviewed each session. This makes it difficult to conclude that the human reviewed scores can be produced consistently. A further study with reviewers from more diverse backgrounds would be needed to make broader claims about human review. Crowdsourcing solutions are unfortunately not always a viable method of human review, since the data is only pseudo-anonymized and sharing private data publicly on the internet quickly runs afoul of privacy standards. Future work computing the inter-rater agreement for this task is required to reveal how many human reviewers are required for a low variance estimate; if inter-rater agreement is low, then many reviewers are required, whereas if the inter-rater agreement is high, fewer are needed.

4.7 Conclusion

In this paper, we investigated the degree to which emotional signals were present in private messaging data. While not as accurate at predicting affect as human review, the automated sentiment analysis algorithm, VADER, was found to be similarly accurate with other automated sentiment analysis on traditionally expressive information (i.e. personal diary entries, direct speech). VADER analysis of private messaging data was additionally found to be more accurate than similar automated sentiment analysis of public social media data. These results suggest that discussion in private spaces, including private messaging, lends itself to more honest emotional sharing. We propose that this may be a result of user control over their recipient audience in private spaces. We suggest that future studies investigate experimental setups to test the causal relationships between honest emotional disclosure and social support.

Chapter 5

Preliminary Work: Chirp, A Platform to Investigate Self-Disclosure and Social Support in an Anonymous Space

5.1 Introduction

While previous work described in Chapters 3 and 4 suggests a positive relationship between social support and online disclosure and private communication use, these studies are only indicative of relationships between these factors, and are unable to determine the causal direction of these relationships. Similarly, previous work shows strong relationships between communication in private and semi-private spaces and positive aspects of well-being. Analyses of different online support groups show they can encourage self-disclosure and provide significant social support (e.g., [3,6,58,68,168]). However, to our knowledge, most studies on the topic are either correlational studies based on cross-sectional surveys or qualitative analyses of current users [65,81]. With some exceptions (e.g., [45]), there have been few longitudinal studies that have been able to test the causal direction behind these correlations.

Several theories exist that try to explain why these relationships may exist. For example, Self Determination Theory (SDT) suggests that people use social media to try to cover psychological needs that are unmet offline [65,81,123]. The Interpersonal-Connection-Behaviors (ICB) Framework also notably proposes that *active* versus *passive* engagement leads to positive and negative outcomes respectively [32]. Across these theories, we see a trend where intent and outside context are the main mediating factors for expected outcomes.

However, social media platforms incorporate various design elements that can influence user interactions and subsequently impact outcomes [65]. Previous work in Chapters3 and 4 support this, suggestions differences in type and quality of communication that are made on **public** (visible to any user on a platform) versus **private** (visible only to invited users) or **semi-private** (visible to any user, but likely only viewed by a specific cohort of users) platforms [71,101]. In these studies we predict that specific aspects of private and semi-private online communication lead to the positive outcomes in personal well-being.

Thus we developed Chirp to allow for experimental intervention in order to test the causal impacts of different design primitives in online communication systems. Chirp provides a space that prompts cohorts of anonymous users to emotionally self-disclose through emoji-based mood tracking posts. Chirp is a highly modular mobile app whose primary intention is to act as a safe space where we can study cohorts of users when exposed to different design decisions, without the need to compromise participant privacy.

5.2 Existing Disclosure Systems

As Chirp is intended as a social mood tracker, here we discuss similar existing systems. There are many commercial journaling and mood tracking apps that aim to improve users' daily routines and well-being (e.g., Daylio [29]). This also includes well-being apps with mood tracking features, such as Headspace[1] or Calm[2] [163]. These apps primarily intend to act as spaces for self-reflection and self-improvement, and generally do not include public or social spaces.

Previous work in CSCW has also included a number tracking systems similar to Chirp. The app Opico, developed by Khandekar et al., uses emojis as the basis of communication, much like we do

[1] https://www.headspace.com/
[2] https://www.calm.com/

in Chirp [82]. In the Opico app, users can select a location on a public map and leave a string of emojis as a "reaction" to it. These reactions can also encompass emotional or mood-related responses to certain location, and therefore show similarity to our "posts". Through the app, the authors find that, given context, user interpretations of emojis can be accurate to original intent.

Significant Otter [96], and IntimaSea [78] were both created as ways to communicate users' current emotional state to partners or friends that are already close with the user. Significant Otter allows users to send animated otter characters from their own smart watch to that of a partner [96]. The app uses user biosignals detected through the watch to predict potential animations that match participant "states" (including emotions, activities, greetings, and acts of affection). By sending these states users can communicate without the use of text. Similarly, IntimaSea uses automatic stress tracking to share stress levels within a group of friends [78]. However, unlike both Significant Otter and Chirp, IntimaSea allows users to share additional content (i.e., text, images, drawings) with others in their group.

While these systems each touch on aspects that we address in Chirp, none specifically focus on emotional and mood disclosure and social support in private social media settings.

5.3 Developing The Chirp Application To Study Self-Disclosure in Online Spaces

Chirp[3] is an open source mood tracker[4] and social media mobile application that uses emojis as the only form of user expression. It is built as a sandbox space to explore a simplified social media system that may allow users to build social support through emotional self-disclosure. The system does not require users to share any personally identifiable information. Moreover, every user must explicitly agree to an information privacy and consent form before using the app. The app is publicly available on the Apple and Google app stores, under the name "Chirp Social Mood Tracker".

5.3.1 Design Considerations

Based on our previous work around emotional self-disclosure and feelings of social support in Chapters 3 and 4, Chirp is built to **encourage emotional self-disclosure** in a **semi-private** space. Furthermore, in order to keep our results as generalizable as possible, we **incorporate common social media platform design choices**.

As researchers, we are obligated to protect our participants as much as possible. However, the social nature of the app means we cannot guarantee the interactions participants will encounter within the app. Thus, we **preserve participant privacy as much as possible**, including by preventing users from voluntarily sharing identifiable information. This makes it substantially more difficult for negative interactions to occur during the use of the app, particularly the forms of negative interactions seen in cyberbullying on other social media platforms [8].

During the design phase of app development, we ran four separate user studies to refine the design and test the viability and interest in the platform.

[3] https://chime.cs.brown.edu/

[4] We note that Chirp asks users the question "how are you feeling?", and includes main emojis that may more accurately be labeled as emotions rather than moods. However, by convention we use the term "mood tracker" to describe the form of the system.

5.3.1.1 Mood Tracking

CHIRP uses the underlying functionality of a mood tracker to encourage emotional self-disclosure. Although we considered alternatives such as messaging apps and online forums, we concluded that an anonymous social mood tracker would offer the best opportunity to safely study the impacts on social support. By including both a traditional mood tracker "profile page" and a semi-public social "timeline" where users can "post", we allow for a non-social version of the app (sans timeline) to still look and feel like a completed app for control cases. This simultaneously allows us to provide one of the most common shared characteristics between social media sites, a profile page [66].

Across two separate phases of early user testing, testers were placed into a Discord server[5] and asked to post about their emotional state daily using only emojis (testers were prevented from posting text via an automated Discord bot). While testers reported it was an interesting experience, they complained that the variety of emoji options made it hard to choose an initial emoji to use as a general descriptor. For this reason, along with ease of analysis, we implemented the concept of a "main" emoji for mood tracking (See Figure 5.1).

We restricted users to choosing their "main" emoji from 9 options:

- Happy
- Satisfied
- Excited
- Tired
- Bored
- Stressed
- Angry
- Sad
- Scared

This subset was based on the six universal emotions as defined by Ekman: happiness, sadness, anger, fear, disgust, and confusion [50]; as well as the six moods tested in the Profile of Mood States (POMS): tension, depression, anger, vigor, fatigue, and confusion [104]. To select which main emotions and moods to include, we released a pilot survey asking student responders to report which emotions or moods they had felt that day (from POMS and Ekman), as well as each of the terms rephrased into layman's terms. We additionally added some neutral options, such as bored and content, and asked for any final thoughts. Based on this survey, we included the most commonly referenced options from POMS (fatigue, tension), Ekman's list (happiness, sadness, anger, fear), and our added options or rephrasings (excited, bored, content). We rephrased these nine selected states to make sense as a response to the question "How are you feeling?" which is shown in the app (Figure 5.1). Finally, user testing reports showed that testers were confused by the term "content" (as a state of satisfaction) as opposed to "content" (as the things that are held or included in something); thus we used the term "satisfied" in our final study design.

5.3.1.2 Anonymity

Interactions over social media and other online communication are not solely positive [8]. Sharing personal information can lead to negative online interactions or bullying, for example doxxing [7]. Given the potentially sensitive nature of the posts shared in the app, we keep CHIRP anonymous. To maintain anonymity we: 1) do not collect or display any personal information about users, and 2) do not allow users the opportunity to voluntarily share personal information through the app (for example posting "my name is <name> I live in <location>"). The latter case is discussed further below.

[5] https://discord.com/

While we wanted to keep users anonymous, we still wanted to maintain popular elements of social media platforms. This meant including usernames and profile images (i.e., avatars) [66]. To avoid personal information such as names or images of users in the app, we assign automatically-generated usernames and avatar images for each user. Avatars were generated using Gravatar[6]. A random combination of an adjective and an animal provided a potentially memorable username and has been used in past studies to identify participants (e.g., [159]).

5.3.1.3 Emoji Only Communication

To preserve anonymity, we prevent users from sharing personal information in the app. This means that we cannot allow free text in social spaces. We chose emojis as an alternative form of communication since they have been shown to be very understandable given context [82]. In our case, shared context is provided in two ways: all posts in the app are in a specific format describing mood; and, in the study, all participants are first year undergraduate students at the same institution.

Including restrictions on user content are also a trend seen in many social media platforms. For example, Twitter's character limit, or Instagram which requires every post to include an image. We additionally note that not every social network relies on text, for example, the social networks Yo[7] or Emojili[8]. During user testing, testers expressed that they found the requirement to only use emojis to be a fun exercise. We hope this gamified experience felt reminiscent of other social media applications.

We note that to avoid confusion or ambiguity due to different art for Unicode emojis on Android versus iOS platforms, we use emojis from a third-party library called JoyPixels[9]. Thus all emojis appeared identical for all app users regardless of the type of phone they had.

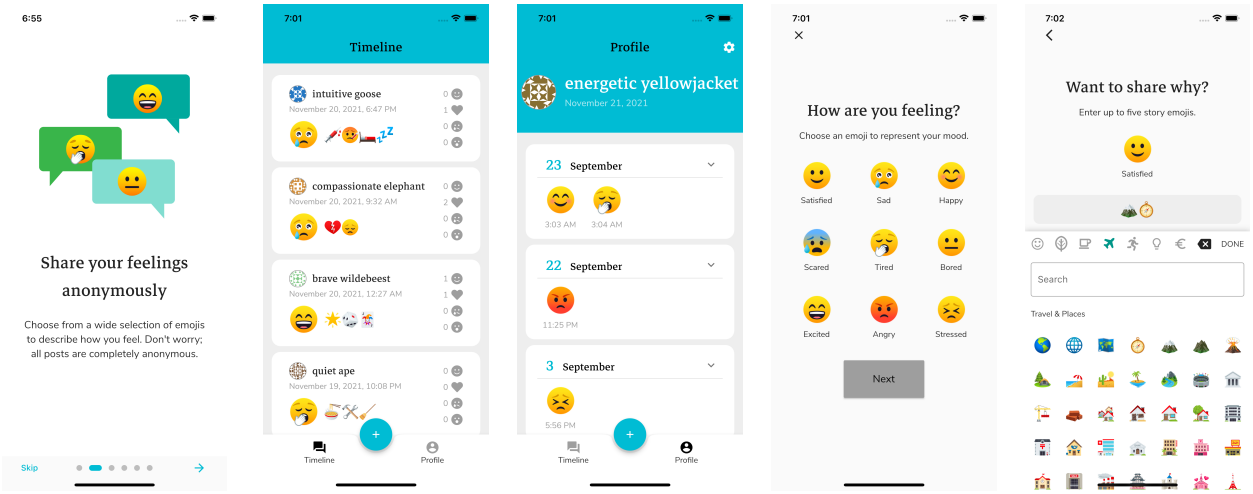
5.3.1.4 Limiting Responses to Reactions

The ICB framework describes how active social media use involves connection promoting behaviors [32]. As Chirp was intended as a social media platform where active use is possible, it needed some form of connection promoting— or interactive— features. As mentioned previously, we do not allow users to enter text in the app to preserve anonymity. Thus we cannot include text comments or responses. Furthermore, comments are an easy way to respond negatively, even without text (imagine a user posting about a sad day when a pet died and the comment response "😂😂🤪").

A different, very common feature in social media platforms is a system of reactions or likes [66]. Examples include reactions on Facebook or in Twitter DMs, likes/favorites on Instagram posts or Tweets, and upvotes/points systems such as Reddit. Studies have shown that reacts across different platforms can have a positive effect on online communities and interpersonal relationships, and can be socially supportive [27, 69, 79]. Thus we included reacts as a form of interaction and support-showing in the Chirp app.

We chose to include four reaction types: a smiling face, a crying face, a gasping face, and a heart. These reactions were based on common reactions between Facebook, Twitter, Instagram, TikTok, and LinkedIn. The exception to this is the "angry react", which is included in both the Facebook and Twitter DM reaction options but we do not include it in the Chirp app. The authors concluded

[6] https://en.gravatar.com/
[7] https://en.wikipedia.org/wiki/Yo_(app)
[8] https://emoj.li/
[9] https://joypixels.com/



(a) Onboarding Screen **(b)** Main User Timeline **(c)** User Profile **(d)** Main Emoji Selection **(e)** Story Emoji Selection

Figure 5.1: CHIRP application main pages. Note that the main user timeline (b) was not shown to participants in the Individual group, who had the user profile (c) as their main app page. The Timeline and Profile buttons on the bottom of the screen were also not shown to Individual group users.

that the angry react could be used as a negative response, and thus it was not included as a reaction option at this time. Borrowing conventions from these popular social media apps, users may select only one reaction to a post.

5.3.2 App Interface Design

When users first log into CHIRP, it presents them with our user privacy and consent form. This privacy and consent form informs users that "[their] anonymized data may be used or shared for future research." All users read this form before they are given the option to continue to use CHIRP. After a series of short onboarding screens, CHIRP shows users a timeline where they can see posts made by other users and add their reactions to their posts (as described in Section 5.3.1.4; see also Figure 5.1).

While users can see reactions to their posts on the timeline page (Figure 5.1b), they cannot view reactions to their posts on their profile page (Figure 5.1c). This is because the primary function of the profile page is as a mood-tracking log, somewhat separate from the social aspects of CHIRP.

In the app, users can create their own posts. They first start by reading the prompt: "How are you feeling?" and accordingly choose from 9 different "main emojis" (as described in Section 5.3.1.1).

After users choose a main emoji, they are presented with a page asking them to "... share why" by creating a story: a string of at least 1, but up to 5 emojis describing the main emoji they initially inputted. We provide a broad yet limited library of emojis to use in the story.

To ensure that the app appears as similarly across user devices as possible, we used Google's Flutter app framework, which does not use native iOS or Android components and instead renders the app pixel by pixel. Furthermore, as mentioned previously, we use the JoyPixels emoji font to ensure that all emojis appeared identically for iOS and Android users.

5.3.3 Privacy and Consent

We designed Chirp to ensure the privacy of our participants, as discussed in Section 5.3.1. Instead of personally identifiable information like email addresses or phone numbers, Chirp relies on the hashed identification code of our participants' mobile devices to identify individual accounts. A user's client device hashes their device ID before being securely sent over HTTPS to our web server. Users must agree to our terms of service about how our study will use their data before any data is stored.

5.3.4 Data Flow for Account Creation and Recording Data

As mentioned in Section 5.3.1.2 users are assigned a random avatar and username. We give users these anonymous identifiers to create a system that feels similar to major social media applications while providing memorable anonymous tokens to identify users. When creating an account, users are also placed in a "group" and assigned a group ID. They can only see and share posts within their group. Our university-managed server stores the user's salted and hashed private device ID, group ID, and anonymous public information in a database.

Chirp records a user's posts and reactions within a centralized database using a RESTful API. When a user creates a post, their device sends their post data to the server, which uses the cookies sent along with the request body to identify an authenticated session. When users react to a post, their device first updates their reaction status and then sends this data to the server to store in the database. When a user queries a post, Chirp will retrieve aggregated reaction counts and the latest reaction from the user making the query.

Chapter 6

Proposed Work

6.1 Preliminary Work Contributions

- Chapter 3: messaging in private spaces correlates with perceived social support more strongly than phone or video call, implying that the persistent and reflective nature of written communication may facilitate perceived social support
- Chapter 4: sentiment analysis techniques more accurately predict ground truth affect when run on private data than public data, implying that the ability to select and control ones audience may encourage more honest self-disclosure
- Chapter 5: A system to evaluate causal relationships in private and semi-private social media

6.1.1 Proposed Additional Contributions

1. Investigate the salience of alternative forms of written communication in self-disclosure context and the factors that impact understanding
2. Investigate the causal relationship between written self-disclosure in semi-private online spaces and perceived social support
3. Investigate the causal relationship between perceived social support and self-disclosure in a social versus non-social space

6.2 The Impact of Demographic Context on Understanding in Emotional Self-Disclosure Through Emojis

Here I propose a survey study investigating how demographics and knowledge of a poster's demographics can influence understanding of emojis. I specifically focus on emojis here as they are a relatively rich form of written communication while simultaneously obscuring particularly personal information that has the potential to be abused. For these reasons, we also use emoji as the main form of communication between users in Chirp.

6.2.1 Proposed Approach

6.2.1.1 Participants and Recruiting

Participants will be recruited from a variety of online sources (notably Reddit, Twitter, and Mastodon), and locally through placing flyers and posters in the community. With the platform from which participants found the study (Reddit, physical media, etc) noted and stored.

Recruiting primarily through online sources allows us to reach a demographically diverse population. In accordance, the survey can be advertised in general survey taking subreddits in order to avoid limiting the participant pool to people specifically interested in emojis, including r/SurveyCircle, r/PaidStudies, etc. However, as active online users (especially on Reddit) are more likely to be tech savvy than the average person, the survey will also be advertised using physical posters in order to get a diverse population in terms of online experience.

Participants will be required to be over the age of 18, reside in the United States, and not have participated in the previous CHIRP study held in 2022 as these same posts were visible to participants of that study. Research demonstrated that emoji interpretation between different national cultures is very varied [130, 146], however this study focused on other demographic markers (i.e. gender, age, and student status). Thus participants are limited to those who currently reside in the United States to control for cultural variance.

6.2.1.2 Study Procedure

Survey Contents Within the survey, participants will be presented four emoji posts from the CHIRP application. The survey displays the posts with the context that when creating a post in the CHIRP application, users are first asked "How are you feeling?" and allowed to select one main emoji from a set of nine "expression" emojis. Following the selection of a main emoji, users are asked "Want to share why?" and are allowed to select up to five additional emojis. Displayed posts in the survey will present the main emoji first and at double the size of the following story emojis.

Each participant will be shown the same four emoji posts in a randomized order. With posts chosen from the previous CHIRP study for which "ground truth labels" were collected.

Additionally each selected post has a different main emoji in order to understand the nuances of the interpretations of different emotions. Ultimately, the survey included one post with a main emoji that was clearly positive, one with a main emoji that was clearly negative, one in which the main emoji's positive or negative sentiment was ambiguous, and another in which the main emoji's meaning altogether was ambiguous. Meaning interpretations of these emojis were based on participant interpretations in a previous study. The four selected posts can be seen in Figures 6.1a, 6.1b, 6.1c, and 6.1d.

(a) The selected "happy" chime post

(b) The selected "angry" chime post

(c) The selected "satisfied" chime post

(d) The selected "tired" chime post

For each of these posts, a ground truth label was provided by the original poster and used for accuracy scoring. One example of a ground truth label is the following explanation provided for the "tired" post:

> *"They were tired so they got coffee; then they grabbed their backpack and took the bus to work. They did work then saw their boyfriend/girlfriend later."*

Additionally, each post will be accompanied by randomly selected demographic information across three categories: gender, age group, and student status. Within these three categories, two "opposite" were selected (male/female for gender, young adult/adult for age, student/non-student for student status). Please note that for non-binary participants, the "opposite" value was randomly selected between male and female. These options were chosen in order to simplify the way in which demographic information was provided with each post and easily track the options that could be displayed.

With these three categories, there are five options of the actual information that would be provided with the post:

1. **The Ground Truth Demographic Information.** The post would be accompanied by all three categories as they actually were for the original poster.

2. **The Opposite of the Ground Truth Demographic Information.** The post would be accompanied by the exact opposite of the truth for each category of demographic information. For example, if the post was actually made by a male young adult who is a student, the post would be shown as being made by a female adult who is not a student.

3. **A Reflection of the Participant's Demographic Information.** The post would be displayed with whatever the survey taker's demographic information is. This information is collected in the initial part of the survey, and then is coded to be stored and later displayed in this portion. The demographic information of the poster has no impact on this option.

4. **The Opposite of the Reflection of the Participant's Demographic Information.** The post would be displayed with the exact opposite demographic information of the survey taker. For example, if the survey taker was a female adult who is a student, the post would be displayed as being made by a male young adult who is not a student. The demographic information of the poster has no impact on this option.

5. **No Demographic Information.** The post would show no demographic information at all (for any of the three categories).

For each post, participants will be asked to share what they believe the poster meant to convey in the post. Participants will also be asked to rate the confidence of their interpretation on a scale of one to ten.

After each post is displayed, participants will taken to the next screen in which they were shown "another person's interpretation of the post" (in truth the ground truth label of the post). The ground truth labels displayed in this portion of the survey are perturbed to remove any personal information and sound as though they are coming from an external perspective. After being shown this other interpretation, participants will be asked if their confidence in their initial interpretation had changed (on a Likert scale). They were also asked to explain the change (or lack of change) in short answer text.

Compensation Participants who complete the survey will be able to enter their email addresses through a separate form to join a raffle. 10% (one in ten) of participants that enter their email addresses in this raffle will be awarded a $25 virtual gift card, sent to the provided email.

6.2.2 Proposed Analysis

For this study, I propose a combination of quantitative and qualitative analysis.

We must first perform content analysis of provided post interpretations, getting a numerical measure of accuracy per interpretation when compared to the ground truth descriptions.

We can then use this measure to quantitatively measure accuracy by participant demographic, by accuracy of the displayed demographic information for the poster, and by similarity between participant and poster demographics. These analyses will provide an understanding of how demographic context impacts interpretation of emojis.

We can also look at aspects such as current participant emotional state (asked at the beginning of the survey) or individual demographics of posters (such as posters being male) to see how these factors may have influenced interpretation accuracy. Much like how we viewed confidence in Chapter 4, we can also look at confidence of participants in their interpretations and how this impacted accuracy. Additionally, we can see if there are any trends in confidence before and after viewing the ground truth labels.

Using more qualitative measures, I propose that we can analyze interpretation responses for reasoning behind interpretations. Though participants were not asked to explain their interpretations, many opted to do so, and these explanations can provide insight into how and why participants might interpret emojis in specific ways. We can also do thematic analysis and see if specific themes or thought processes are more commonly seen among different demographic groups. We also have access to short descriptions of why participants felt more or less confident after seeing the ground truth label per post, which we can further analyse for trends.

6.3 Chirp: The Impact of Private Online Self-Disclosure on Perceived Social Support

Here I propose a between-subjects study using Chirp as a platform to better understand the impacts of online self-disclosure in private online spaces.

6.3.1 Proposed Approach

6.3.1.1 Participants

Participants will be recruited solely from first-year undergraduate students at Brown. For recruitment, we can post physical posters around campus, including in and around first-year dormitories, and in popular hangout spots near campus. As well as post advertisements online in social media groups created as conversational spaces for first-year undergraduates from the authors' institution (i.e., Facebook, Instagram, Reddit). All participants must be 18 or older.

First-year university students are the target population since many students are no longer physically near their primary social support network (e.g., parents, friends) once they start college, and therefore are likely to benefit from an online source of social support [11].

In order to be sure that all participants are under the same university-mandated COVID safety policy, it makes sense to recruit only from a single university. Furthermore, as mentioned previously, previous studies show that emojis are understandable *given a shared context* [82]. For example, within a specific Facebook group, shared friend group, or known location. By using a cohort of students from the same university, we can ensure an appropriate level of shared context that would ensure understandable communication between participants. Ideally we want to recruit at least 75 participants.

6.3.1.2 Study Procedure

I propose a two week study, with participants randomly separated into one of three groups, each group with differing levels of access and features within Chirp.

1. **Social (27 total participants)**: using the fully-featured version of Chirp as described above. Participants in an isolated study-only group. These participants can only share posts and interact with users in this group.

2. **Individual (25 total participants)**: using only Chirp's profile page as a mood tracker with no access to the social timeline.

3. **Control (25 total participants)**: did not use Chirp at all.

Opting for three groups in a between-subjects study (as opposed to a within-subjects study) gives two benefits: first, this allows us to avoid influences from familiarity with the platform over time, as we might see in a within-subjects study; second, we can collect data on all three groups over the *same period of time*, which prevents temporal influences on participant feelings of social support. The Social group acts as our primary experimental group, testing the impact of self-disclosure over the full social Chime app, while the Individual and Control groups both act as baseline groups. As we are studying students, there are common outside factors that are likely to impact feelings of social

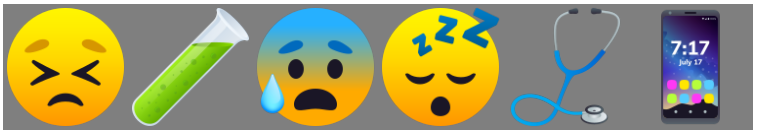


Figure 6.2: An example of a post made in CHIRP. The first emoji show is the "main" emoji, and all following emojis are the "story" emojis. This image was shown to participants as part of the final study survey, and participants were asked the following question: What do you think the author of the following post meant to convey about their mood? The first of these emojis was chosen in response to the prompt "How are you feeling", while the others were chosen in response to a follow-up prompt "Want to share why?"

support, for example, changes to university COVID policy, school breaks, or (as we encountered in this study) periods of mid-terms or finals. The Control group provides a baseline of comparison when considering those outside factors. By contrast, the Individual group allows us to separate the impact of mood tracking and self-reflection from that of the actual social interactions made in the CHIRP app.

Pre-Study Eligibility and Onboarding All participants should complete an eligibility survey as well as an onboarding survey containing a consent form and setup instructions and informing them of the study start date (as all participants across groups started and ended the study on the same days). Participants in the Social and Individual groups additionally must install and join the correct versions of the CHIRP app as part of the survey, but not begin posting in the app until the study start date. Once the study begins, Social and Individual participants are incentivised and requested to post in the app at least once per day. Participants in the Social group are also instructed that all users posting in the app were first-year undergraduates at their institution. This information was intended to create a natural "cohort" for the participants in the Social group, similar to a Facebook group or a subreddit community.

Study Surveys During the two-week study, all participants complete three study surveys—on the start date, one week after the start date, and two weeks after the start date. Weekly surveys allowed us to test *trends* in feelings of social support and outside factors rather than just a start and end point. The participants must fill out the surveys to the best of their ability within two days of receiving the email requesting a survey response.

All three study surveys include self-report questions on general social media and messaging time ("On average, how many hours per day do you spend [reading social media/posting on social media/messaging]"). The surveys also include psychological scales for participants to fill out, including the MSPSS and the OSSS. The MSPSS was included as a baseline measure of offline social support, as changes in outside factors related to social support networks could impact participant use or relationship to online social support. The OSSS is therefore used as a comparative measure to detect changes in perceived social support. Finally, all surveys asked participants what emotions they felt each of the 9 main emojis from the CHIRP app represented.

In both follow-up surveys, participants are asked if they had gone to someone for support in the past week, and if so who and when. These surveys also ask participants in the Social and Individual groups whether they had shared their CHIRP username with anyone else, and to explain the intended meaning of a message from a provided recent post they had made in the CHIRP app. Additionally, two of these provided posts are chosen and presented to all participants in the Social and Individual groups

during the second (mid-experiment) survey and to all participants in the final survey. Participants are asked to describe what they thought the original poster meant to express in the given post.

The final survey additionally asks whether participants in the Social group felt that they connected particularly well or poorly with any other users of the app, and asked participants in the Individual and Social groups about their feelings on the emoji-only communication within the app. Finally, all participants are given a space for any final thoughts. In this final survey, Control group participants are also shown the same two selected posts as the other participants and asked to describe what they thought the original poster meant to express.

Compensation Participants are paid $10 for each survey completed as part of this study. Participants in the Social and Individual groups were additionally paid $1 for each day during the two week study in which they posted, if these participants posted in the app every day of the study and completed all surveys then they were paid a $7 bonus (equaling a maximum of $50 in compensation). This incentive structure was chosen to make sure that participants consistently engage with the app, with the option to engage more often than we required.

6.3.2 Design Considerations

6.3.2.1 Use of Emoji

As mentioned previously, we built Chirp to prioritize participant privacy. Part of this meant preventing participants from posting in text. Emoji presented themselves as a popular form of communication that to most university students is fairly straightforward that maintained enough information given context [82] while obscuring privacy-breaking details.

6.3.2.2 Lack of Comment Functionality

In this study we aim to answer two questions, around self-disclosure and around social spaces. With the latter of these in mind, giving participants the ability to comment on posts would introduce significant noise to any results. When comparing participants in the Individual group to the Social group, if comments were available then the content of those contents would likely have a big impact on results. Thus in an effort to remove confounding variables, we should not include commenting functionality.

6.3.2.3 First-year Students

As mentioned above, first-year university students are often at a point of transition in their lives becoming physically separated from friends and family back home. This makes the first year of university a time when forming social support networks is more urgent, and therefore more likely to happen. First years are therefore a population most likely to benefit from the results we find, as well as most likely to show results.

6.3.2.4 Students at Brown

There are two reasons to limit participants to students at Brown. First, being in a similar cohort where most events are known across the group helps ensure that participants in the Social group have enough context to understand each others' posts. Second, as there are various lockdown and remote

work policies across universities that will strongly impact ability to interact in-person and connect in traditionally socially-supportive ways, we can ensure greater consistency for students in the study by keeping them at one university where policies will be consistent for all students if not over time.

6.3.2.5 Two-week Time Period

I propose a two week study here. This is based on previous similar studies that ran for one [45] or two [96] weeks and found significant results. Thus two weeks should be a sufficient length of time. Additionally, it is helpful to keep the study shorter to avoid dropout and fatigue.

6.3.3 Proposed Analysis

For this study, I propose to complete a combination of qualitative and quantitative analysis.

We can perform quantitative analysis across many dimensions. Most obviously, a comparison between perceived social support across all three of the study groups. Other basic statistics such as posting frequency and reaction frequency can be used to get an idea of user interest.

It will also be useful to look at the impact of reactions (both number and type) on emotional state and perceived social support of participants in the Social group. We can additionally use the actual mood posts themselves as labels, and see how moods transition over time.

We know that there will be some question skipping and dropout over the course of the study. While participants that fully dropped out can be removed from the study, we can account for missing responses to individual questions in larger scales (such as the 40 question OSSS) either through imputation, careful re-calculation of values, or through more complex methods such as more modern growth modelling approaches.

In terms of qualitative analysis, we can do some analysis of the accuracy of Chirp post interpretation using techniques such as content analysis. This can be compared to self-report data on how much participants felt they did or did not understand. In doing so we can compare accuracy across or within groups. We can also in our qualitative analysis look at how participants report to have used Chirp, and in what ways they found it helpful or unhelpful. We can also see how well Social group participants felt they connected with other users in the Social group.

6.4 Timeline

1. January 16, 2024: submit revision of Chirp paper (Section 6.3)
2. Spring 2024: analysis and writing for Emoji Understanding paper
3. July 18, 2024: submit Emoji Understanding paper (Section 6.2)
4. Summer-Fall 2024: complete dissertation and any required revisions for Chirp and Emoji Understanding papers

Bibliography

[1] Marije aan het Rot, Koen Hogenelst, and Robert A. Schoevers. Mood disorders in everyday life: A systematic review of experience sampling and ecological momentary assessment studies. *Clinical Psychology Review*, 32(6):510 – 523, 2012.

[2] Mariek Vanden Abeele, Alexander P Schouten, and Marjolijn L Antheunis. Personal, editable, and always accessible: An affordance approach to the relationship between adolescents' mobile messaging behavior and their friendship quality. *Journal of Social and Personal Relationships*, 34(6):875–893, 2017.

[3] Nazanin Andalibi. What happens after disclosing stigmatized experiences on identified social media: Individual, dyadic, and social/network outcomes. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–15, New York, NY, USA, 2019. Association for Computing Machinery.

[4] Nazanin Andalibi and Andrea Forte. Announcing pregnancy loss on facebook: A decision-making framework for stigmatized disclosures on identified social network sites. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pages 1–14, 2018.

[5] Nazanin Andalibi, Oliver L Haimson, Munmun De Choudhury, and Andrea Forte. Social support, reciprocity, and anonymity in responses to sexual abuse disclosures on social media. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 25(5):1–35, 2018.

[6] Nazanin Andalibi, Pinar Ozturk, and Andrea Forte. Sensitive self-disclosures, responses, and social support on instagram: the case of #depression. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*, pages 1485–1500, 2017.

[7] Briony Anderson and Mark A Wood. Doxxing: A scoping review and typology. *The Emerald international handbook of technology-facilitated violence and abuse*, pages 205–226, 2021.

[8] Monica Anderson. A majority of teens have experienced some form of cyberbullying, 2018.

[9] Monica Anderson and Jingjing Jiang. Teens, Social Media & Technology 2018, May 2018.

[10] Michael F Armey, Janis H Crowther, and Ivan W Miller. Changes in ecological momentary assessment reported affect associated with episodes of nonsuicidal self-injury. *Behavior Therapy*, 42(4):579–588, 2011.

[11] Jeffrey J Arnett, Rita Žukauskienė, and Kazumi Sugimura. The new life stage of emerging adulthood at ages 18–29 years: Implications for mental health. *The Lancet Psychiatry*, 1(7):569–576, 2014.

[12] Gökmen Arslan, Murat Yıldırım, and Masood Zangeneh. Coronavirus anxiety and psychological adjustment in college students: Exploring the role of college belongingness and social media addiction. *International Journal of Mental Health and Addiction*, 19(1):1–14, 2021.

[13] Marianne Aubin Le Quere, Maria Antoniak, Tegan Wilson, Alexa VanHattum, Griffin Berlstein, Elizabeth Ricci, Andrea Cuadra, Sachi Angle, and Sharifa Sultana. Survey of 106 computing grad students highlights covid-19 stresses, possible solutions, 2020. [Online; posted 28-June-2020].

[14] Valerio Basile, Francesco Cauteruccio, and Giorgio Terracina. How dramatic events can affect emotionality in social posting: The impact of covid-19 on reddit. *Future Internet*, 13(2):29, 2021.

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

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

[17] Natalya N Bazarova, Jessie G Taft, Yoon Hyung Choi, and Dan Cosley. Managing impressions and relationships on facebook: Self-presentational and relational concerns revealed through the analysis of language style. *Journal of Language and Social Psychology*, 32(2):121–141, 2013.

[18] Asaf Beasley and Winter Mason. Emotional States vs. Emotional Words in Social Media. In *Proceedings of the ACM Web Science Conference*, pages 1–10, Oxford, United Kingdom, 2015. ACM Press.

[19] Asaf Beasley, Winter Mason, and Eliot Smith. Inferring emotions and self-relevant domains in social media: Challenges and future directions. *Translational Issues in Psychological Science*, 2(3):238–247, 2016.

[20] Robin N. Brewer, Sarita Schoenebeck, Kerry Lee, and Haripriya Suryadevara. Challenging passive social media use: Older adults as caregivers online. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW1), apr 2021.

[21] Matthew HEM Browning, Lincoln R Larson, Iryna Sharaievska, Alessandro Rigolon, Olivia McAnirlin, Lauren Mullenbach, Scott Cloutier, Tue M Vu, Jennifer Thomsen, Nathan Reigner, et al. Psychological impacts from covid-19 among university students: Risk factors across seven states in the united states. *PloS one*, 16(1):e0245327, 2021.

[22] Feifei Bu, Andrew Steptoe, and Daisy Fancourt. Loneliness during a strict lockdown: Trajectories and predictors during the covid-19 pandemic in 38,217 united kingdom adults. *Social Science & Medicine*, 265:113521, 2020.

[23] Moira Burke, Robert Kraut, and Cameron Marlow. Social capital on facebook: Differentiating uses and users. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 571–580, New York, NY, USA, 2011. Association for Computing Machinery.

[24] Vanessa Caba Machado, David Mcilroy, Francisca M Padilla Adamuz, Rebecca Murphy, and Susan Palmer-Conn. The associations of use of social network sites with perceived social support and loneliness. *Current Psychology*, pages 1–14, 2022.

[25] Janie Canty-Mitchell and Gregory D Zimet. Psychometric properties of the multidimensional scale of perceived social support in urban adolescents. *American journal of community psychology*, 28(3):391–400, 2000.

[26] Wenjun Cao, Ziwei Fang, Guoqiang Hou, Mei Han, Xinrong Xu, Jiaxin Dong, and Jianzhong Zheng. The psychological impact of the covid-19 epidemic on college students in china. *Psychiatry research*, 287:112934, 2020.

[27] Caleb T Carr, D Yvette Wohn, and Rebecca A Hayes. As social support: Relational closeness, automaticity, and interpreting social support from paralinguistic digital affordances in social media. *Computers in Human Behavior*, 62:385–393, 2016.

[28] Pew Research Center. Social media fact sheet, 2021.

[29] Beenish M Chaudhry. Daylio: mood-quantification for a less stressful you. *Mhealth*, 2, 2016.

[30] Christie Chen and Yang yu Wang. Here's our list of colleges' reopening models, 2020.

[31] Cecilia Cheng and Mike WL Cheung. Psychological responses to outbreak of severe acute respiratory syndrome: a prospective, multiple time-point study. *Journal of personality*, 73(1):261–285, 2005.

[32] Jenna L. Clark, Sara B. Algoe, and Melanie C. Green. Social network sites and well-being: The role of social connection. *Current Directions in Psychological Science*, 27(1):32–37, 2018.

[33] Jenna L Clark and Melanie C Green. The social consequences of online interaction. In *The Oxford Handbook of Cyberpsychology*. Cambridge University Press (CUP), Oxford, United Kingdom, 2019.

[34] Alex Cohen, Kyle Minor, Lauren Baillie, and Amanda Dahir. Clarifying the linguistic signature: Measuring personality from natural speech. *Journal of personality assessment*, 90:559–63, 12 2008.

[35] Sheldon Cohen and Thomas A Wills. Stress, social support, and the buffering hypothesis. *Psychological bulletin*, 98(2):310, 1985.

[36] Andrew Coles. iPhone backup database hashes: which filenames do they use? *iPhone Backup Extractor: Recover Your Lost Data*, 2012.

[37] William E Copeland, Ellen McGinnis, Yang Bai, Zoe Adams, Hilary Nardone, Vinay Devadanam, Jeffrey Rettew, and Jim J Hudziak. Impact of covid-19 pandemic on college student mental health and wellness. *Journal of the American Academy of Child & Adolescent Psychiatry*, 60(1):134–141, 2021.

[38] Neil S Coulson. Receiving social support online: an analysis of a computer-mediated support group for individuals living with irritable bowel syndrome. *Cyberpsychology & behavior*, 8(6):580–584, 2005.

[39] Lorenzo Coviello, Yunkyu Sohn, Adam D.I. Kramer, Cameron Marlow, Massimo Franceschetti, Nicholas A. Christakis, and James H. Fowler. Detecting emotional contagion in massive social networks. *PLoS One*, 9(3), 3 2014.

[40] Marianne Lucena da Silva, Rodrigo Santiago Barbosa Rocha, Mohamed Buheji, Haitham Jahrami, and Katiane da Costa Cunha. A systematic review of the prevalence of anxiety symptoms during coronavirus epidemics. *Journal of Health Psychology*, 26(1):115–125, 2021.

[41] Dmitry Davidov, Oren Tsur, and Ari Rappoport. Enhanced sentiment learning using twitter hashtags and smileys. In *Coling 2010: Posters*, pages 241–249, Beijing, China, August 2010. Coling 2010 Organizing Committee.

[42] Munmun De Choudhury and Scott Counts. The nature of emotional expression in social media: measurement, inference and utility. *Human Computer Interaction Consortium (HCIC)*, 2012.

[43] Munmun De Choudhury, Michael Gamon, and Scott Counts. Happy, nervous or surprised? classification of human affective states in social media. In *Sixth International AAAI Conference on Weblogs and Social Media*, pages 435–438, 2012.

[44] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. Predicting Depression via Social Media. pages 128–137. AAAI, July 2013.

[45] Fenne Große Deters and Matthias R Mehl. Does posting facebook status updates increase or decrease loneliness? an online social networking experiment. *Social psychological and personality science*, 4(5):579–586, 2013.

[46] Peter Sheridan Dodds, Eric M. Clark, Suma Desu, Morgan R. Frank, Andrew J. Reagan, Jake Ryland Williams, Lewis Mitchell, Kameron Decker Harris, Isabel M. Kloumann, James P. Bagrow, Karine Megerdoomian, Matthew T. McMahon, Brian F. Tivnan, and Christopher M. Danforth. Human language reveals a universal positivity bias. *Proceedings of the National Academy of Sciences*, 112(8):2389–2394, February 2015.

[47] Michal Dolev-Cohen and Azy Barak. Adolescents' use of instant messaging as a means of emotional relief. *Computers in Human Behavior*, 29(1):58–63, 2013.

[48] Fabon Dzogang, Stafford Lightman, and Nello Cristianini. Circadian mood variations in Twitter content. *Brain and Neuroscience Advances*, 1:239821281774450, January 2017.

[49] U.S. Department Of Education. Database of accredited programs and institutions, 2021. data retrieved from the Database of Accredited Programs and Institutions, `https://ope.ed.gov/dapip/#/home`.

[50] Paul Ekman. Universals and cultural differences in facial expressions of emotion. In *Nebraska symposium on motivation*. University of Nebraska Press, 1971.

[51] Timon Elmer, Kieran Mepham, and Christoph Stadtfeld. Students under lockdown: Comparisons of students' social networks and mental health before and during the covid-19 crisis in switzerland. *Plos one*, 15(7):e0236337, 2020.

[52] Jane RW Fisher, Thach D Tran, Karin Hammarberg, Jayagowri Sastry, Hau Nguyen, Heather Rowe, Sally Popplestone, Ruby Stocker, Claire Stubber, and Maggie Kirkman. Mental health of people in australia in the first month of covid-19 restrictions: a national survey. *Medical journal of Australia*, 213(10):458–464, 2020.

[53] Andrew J. Flanagin. IM online: Instant messaging use among college students. *Communication Research Reports*, 22(3):175–187, 2005.

[54] Royal Society for Public Health. Status of mind: Social media and young people's mental health, 2017.

[55] Charles E Fritz and Harry B Williams. The human being in disasters: A research perspective. *The Annals of the American Academy of Political and Social Science*, 309(1):42–51, 1957.

[56] Alessandro Gabbiadini, Cristina Baldissarri, Federica Durante, Roberta Rosa Valtorta, Maria De Rosa, and Marcello Gallucci. Together apart: The mitigating role of digital communication technologies on negative affect during the covid-19 outbreak in italy. *Frontiers in psychology*, 11:2763, 2020.

[57] Stephen M Garcia, Kim Weaver, Gordon B Moskowitz, and John M Darley. Crowded minds: the implicit bystander effect. *Journal of personality and social psychology*, 83(4):843, 2002.

[58] Robert P Gauthier, Mary Jean Costello, and James R Wallace. "i will not drink with you today": A topic-guided thematic analysis of addiction recovery on reddit. In *CHI Conference on Human Factors in Computing Systems*, pages 1–17, 2022.

[59] Renee D Goodwin, Lisa C Dierker, Melody Wu, Sandro Galea, Christina W Hoven, and Andrea H Weinberger. Trends in us depression prevalence from 2015 to 2020: the widening treatment gap. *American Journal of Preventive Medicine*, 63(5):726–733, 2022.

[60] Renee D Goodwin, Andrea H Weinberger, June H Kim, Melody Wu, and Sandro Galea. Trends in anxiety among adults in the united states, 2008–2018: Rapid increases among young adults. *Journal of psychiatric research*, 130:441–446, 2020.

[61] US Government. Fact sheet: Biden-harris administration announces new actions to tackle nation's mental health crisis, 2023.

[62] Kris Gowen, Matthew Deschaine, Darcy Gruttadara, and Dana Markey. Young adults with mental health conditions and social networking websites: seeking tools to build community. *Psychiatric rehabilitation journal*, 35(3):245, 2012.

[63] Kathryn Greene, Valerian J Derlega, and Alicia Mathews. *Self-disclosure in personal relationships*. Cambridge University Press, New York, NY, US, 2006.

[64] Lauren R Grocott, Anneliese Mair, Janine N Galione, Michael F Armey, Jeff Huang, and Nicole R Nugent. Days with and without self-injurious thoughts and behaviors: Impact of childhood maltreatment on adolescent online social networking. *Journal of Adolescence*, 94(5):748–762, 2022.

[65] Maya Gudka, Kirsty LK Gardiner, and Tim Lomas. Towards a framework for flourishing through social media: a systematic review of 118 research studies. *The Journal of Positive Psychology*, 18(1):86–105, 2023.

[66] Jiajing Guo. Designing the" front": An overview of profile elements on social network sites. In *Companion Publication of the 2022 Conference on Computer Supported Cooperative Work and Social Computing*, pages 130–134, 2022.

[67] Christine Hagar. Crisis informatics. In *Encyclopedia of Information Science and Technology, Third Edition*, pages 1350–1358. Chandos Publishing, Oxford, 2015.

[68] Oliver L Haimson, Jed R Brubaker, Lynn Dombrowski, and Gillian R Hayes. Disclosure, stress, and support during gender transition on facebook. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*, pages 1176–1190, 2015.

[69] Ren-Whei Joanna Harn. *The visual language of emojis: A study on college Students' social support communication in online social networks*. PhD thesis, University of Kansas, 2017.

[70] Ron D Hays and M Robin DiMatteo. A short-form measure of loneliness. *Journal of personality assessment*, 51(1):69–81, 1987.

[71] Gabriela Hoefer, Talie Massachi, Neil G Xu, Nicole Nugent, and Jeff Huang. Bridging the social distance: Offline to online social support during the covid-19 pandemic. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2):1–27, 2022.

[72] Susan Holtzman, Drew DeClerck, Kara Turcotte, Diana Lisi, and Michael Woodworth. Emotional support during times of stress: Can text messaging compete with in-person interactions? *Computers in Human Behavior*, 71:130–139, 2017.

[73] Stephen B Hulley, Steven R Cummings, Warren S Browner, Deborah G Grady, and Norman M Goldfarb. Designing clinical research (vol. 4th). *Philadelphia: LWW*, 4:14–55, 2013.

[74] C.J. Hutto and Eric Gilbert. Vader: A parsimonious rule-based model for sentiment analysis of social media text. pages 216–225, 2015.

[75] Ru Jia, Kieran Ayling, Trudie Chalder, Adam Massey, Elizabeth Broadbent, Carol Coupland, and Kavita Vedhara. Mental health in the uk during the covid-19 pandemic: cross-sectional analyses from a community cohort study. *BMJ Open*, 10(9):1–14, 2020.

[76] Crystal L Jiang, Natalie N Bazarova, and Jeffrey T Hancock. The disclosure–intimacy link in computer-mediated communication: An attributional extension of the hyperpersonal model. *Human communication research*, 37(1):58–77, 2011.

[77] Long Jiang, Mo Yu, Ming Zhou, Xiaohua Liu, and Tiejun Zhao. Target-dependent twitter sentiment classification. pages 151–160, 01 2011.

[78] Yanqi Jiang, Xianghua Ding, Xiaojuan Ma, Zhida Sun, and Ning Gu. Intimasea: Exploring shared stress display in close relationships. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–19, 2023.

[79] Bethany L Johnson, Margaret M Quinlan, and Nathan Pope. " sticky baby dust" and emoji: Social support on instagram during in vitro fertilization. *Rhetoric of Health & Medicine*, 3(3):320–349, 2020.

[80] Warren H. Jones and Teri L. Moore. Loneliness and social support. *Journal of Social Behavior and Personality*, 2(2):145, 1987. Last updated 2013-02-22.

[81] Betul Keles, Niall McCrae, and Annmarie Grealish. A systematic review: the influence of social media on depression, anxiety and psychological distress in adolescents. *International journal of adolescence and youth*, 25(1):79–93, 2020.

[82] Sujay Khandekar, Joseph Higg, Yuanzhe Bian, Chae Won Ryu, Jerry O. Talton Iii, and Ranjitha Kumar. Opico: a study of emoji-first communication in a mobile social app. In *Companion proceedings of the 2019 world wide web conference*, pages 450–458, 2019.

[83] Funda Kivran-Swaine and Mor Naaman. Network properties and social sharing of emotions in social awareness streams. In *Proceedings of the ACM 2011 Conference on Computer supported cooperative work*, pages 379–382, 2011.

[84] Adam D.I. Kramer. An unobtrusive behavioral model of "gross national happiness". In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, page 287–290, New York, NY, USA, 2010. Association for Computing Machinery.

[85] Adam D.I. Kramer. The spread of emotion via facebook. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, page 767–770, New York, NY, USA, 2012. Association for Computing Machinery.

[86] Adam DI Kramer, Jamie E Guillory, and Jeffrey T Hancock. Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24):8788–8790, 2014.

[87] Nicole Krämer and Stephan Winter. Impression management 2.0: The relationship of self-esteem, extraversion, self-efficacy, and self-presentation within social networking sites. *Journal of Media Psychology: Theories, Methods, and Applications*, 20:106–, 01 2008.

[88] Catherine Penny Hinson Langford, Juanita Bowsher, Joseph P Maloney, and Patricia P Lillis. Social support: a conceptual analysis. *Journal of advanced nursing*, 25(1):95–100, 1997.

[89] Bibb Latane and John M Darley. Group inhibition of bystander intervention in emergencies. *Journal of personality and social psychology*, 10(3):215, 1968.

[90] Andrew M Ledbetter, Joseph P Mazer, Jocelyn M DeGroot, Kevin R Meyer, Yuping Mao, and Brian Swafford. Attitudes toward online social connection and self-disclosure as predictors of facebook communication and relational closeness. *Communication Research*, 38(1):27–53, 2011.

[91] Kyung-Tag Lee, Mi-Jin Noh, and Dong-Mo Koo. Lonely people are no longer lonely on social networking sites: The mediating role of self-disclosure and social support. *Cyberpsychology, Behavior, and Social Networking*, 16(6):413–418, 2013.

[92] Kaitlin M Lewin, Morgan E Ellithorpe, and Dar Meshi. Social comparison and problematic social media use: Relationships between five different social media platforms and three different social comparison constructs. *Personality and Individual Differences*, 199:111865, 2022.

[93] Yuanyuan Li, Jingbo Zhao, Zijuan Ma, Larkin S McReynolds, Dihuan Lin, Zihao Chen, Tong Wang, Dongfang Wang, Yifan Zhang, Jinfang Zhang, et al. Mental health among college students during the covid-19 pandemic in china: A 2-wave longitudinal survey. *Journal of affective disorders*, 281:597–604, 2021.

[94] Chieh-Peng Lin. Assessing the mediating role of online social capital between social support and instant messaging usage. *Electronic Commerce Research and Applications*, 10(1):105–114, 2011.

[95] Ellie Lisitsa, Katherine S Benjamin, Sarah K Chun, Jordan Skalisky, Lauren E Hammond, and Amy H Mezulis. Loneliness among young adults during covid-19 pandemic: The mediational roles of social media use and social support seeking. *Journal of Social and Clinical Psychology*, 39(8):708–726, 2020.

[96] Fannie Liu, Chunjong Park, Yu Jiang Tham, Tsung-Yu Tsai, Laura Dabbish, Geoff Kaufman, and Andrés Monroy-Hernández. Significant otter: Understanding the role of biosignals in communication. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–15, 2021.

[97] Danielle Lottridge and Frank R Bentley. Let's hate together: How people share news in messaging, social, and public networks. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2018.

[98] Fiona Lyddy, Francesca Farina, James Hanney, Lynn Farrell, and Niamh Kelly O'Neill. An analysis of language in university students' text messages. *Journal of Computer-Mediated Communication*, 19(3):546–561, 2014.

[99] Robert D Lytle, Tabrina M Bratton, and Heather K Hudson. Bystander apathy and intervention in the era of social media. In *The Emerald International Handbook of Technology Facilitated Violence and Abuse*. Emerald Publishing Limited, Bingley, UK, 2021.

[100] Yuanchao Ma, Bin Xu, Yin Bai, Guodong Sun, and Run Zhu. Daily mood assessment based on mobile phone sensing. In *Wearable and implantable body sensor networks (BSN), 2012 ninth international conference on*, pages 142–147. IEEE, 2012.

[101] Talie Massachi, Grant Fong, Varun Mathur, Sachin R Pendse, Gabriela Hoefer, Jessica J Fu, Chong Wang, Nikita Ramoji, Nicole R Nugent, Megan L Ranney, et al. Sochiatrist: Signals of affect in messaging data. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2):1–25, 2020.

[102] Matthew A McDougall, Michael Walsh, Kristina Wattier, Ryan Knigge, Lindsey Miller, Michalene Stevermer, and Bruce S Fogas. The effect of social networking sites on the relationship between perceived social support and depression. *Psychiatry research*, 246:223–229, 2016.

[103] Katelyn YA McKenna and John A Bargh. Plan 9 from cyberspace: The implications of the internet for personality and social psychology. *Personality and social psychology review*, 4(1):57–75, 2000.

[104] Douglas M McNair, Maurice Lorr, Leo F Droppleman, et al. Manual profile of mood states. 1971.

[105] David N. Miller. *Positive Affect*, pages 1121–1122. Springer US, Boston, MA, 2011.

[106] Margaret Mitchell, Kristy Hollingshead, and Glen Coppersmith. Quantifying the language of schizophrenia in social media. In *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pages 11–20, Denver, Colorado, June 5 2015. Association for Computational Linguistics.

[107] Kathleen Anne Moore and Evita March. Socially connected during covid-19: online social connections mediate the relationship between loneliness and positive coping strategies. *Res. Square*, 3:1–14, 2020.

[108] Teagen Nabity-Grover, Christy M.K. Cheung, and Jason Bennett Thatcher. Inside out and outside in: How the covid-19 pandemic affects self-disclosure on social media. *International Journal of Information Management*, 55:102188, 2020. Impact of COVID-19 Pandemic on Information Management Research and Practice: Editorial Perspectives.

[109] Melanie Nguyen, Yu Sun Bin, and Andrew Campbell. Comparing online and offline self-disclosure: A systematic review. *Cyberpsychology, Behavior, and Social Networking*, 15(2):103–111, 2012.

[110] Michael Y Ni, Lin Yang, Candi MC Leung, Na Li, Xiaoxin I Yao, Yishan Wang, Gabriel M Leung, Benjamin J Cowling, and Qiuyan Liao. Mental health, risk factors, and social media use during the covid-19 epidemic and cordon sanitaire among the community and health professionals in wuhan, china: cross-sectional survey. *JMIR mental health*, 7(5):e19009, 2020.

[111] Elizabeth A Nick, David A Cole, Sun-Joo Cho, Darcy K Smith, T Grace Carter, and Rachel L Zelkowitz. The online social support scale: Measure development and validation. *Psychological assessment*, 30(9):1127, 2018.

[112] Office of the Surgeon General. Advisory: The healing effects of social connection, 2023.

[113] Jakob Ohme, Mariek MP Vanden Abeele, Kyle Van Gaeveren, Wouter Durnez, and Lieven De Marez. Staying informed and bridging "social distance": Smartphone news use and mobile messaging behaviors of flemish adults during the first weeks of the covid-19 pandemic. *Socius*, 6, 2020.

[114] Stephen F Ostertag and David G Ortiz. Can social media use produce enduring social ties? affordances and the case of katrina bloggers. *Qualitative Sociology*, 40(1):59–82, 2017.

[115] Kenneth M Ovens and Gordon Morison. Forensic analysis of kik messenger on ios devices. *Digital Investigation*, 17:40–52, 2016.

[116] Myoungouk Park, D.W. McDonald, and M. Cha. Perception differences between the depressed and non-depressed users in twitter. *Proceedings of the 7th International Conference on Weblogs and Social Media, ICWSM 2013*, pages 476–485, 01 2013.

[117] Namkee Park, Borae Jin, and Seung-A Annie Jin. Effects of self-disclosure on relational intimacy in facebook. *Computers in Human Behavior*, 27(5):1974–1983, 2011.

[118] Sungkyu Park, Inyeop Kim, Sang Won Lee, Jaehyun Yoo, Bumseok Jeong, and Meeyoung Cha. Manifestation of depression and loneliness on social networks: A case study of young adults on facebook. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '15, page 557–570, New York, NY, USA, 2015. Association for Computing Machinery.

[119] James Pennebaker, Martha Francis, and Roger Booth. Linguistic inquiry and word count (liwc). 01 1999.

[120] James W. Pennebaker, Ryan L. Boyd, Kayla Jordan, and Kate Blackburn. The development and psychometric properties of liwc2015. 2015.

[121] Daniel Perlman and L Anne Peplau. Toward a social psychology of loneliness. *Personal relationships*, 3:31–56, 1981.

[122] Caroline Pitt, Ari Hock, Leila Zelnick, and Katie Davis. The kids are / not / sort of all right*. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 21(352), 2021.

[123] Brian A Primack, Sabrina A Karim, Ariel Shensa, Nicholas Bowman, Jennifer Knight, and Jaime E Sidani. Positive and negative experiences on social media and perceived social isolation. *American Journal of Health Promotion*, 33(6):859–868, 2019.

[124] Brian A Primack, Ariel Shensa, Jaime E Sidani, Erin O Whaite, Liu yi Lin, Daniel Rosen, Jason B Colditz, Ana Radovic, and Elizabeth Miller. Social media use and perceived social isolation among young adults in the us. *American journal of preventive medicine*, 53(1):1–8, 2017.

[125] Ravi Philip Rajkumar. Covid-19 and mental health: A review of the existing literature. *Asian journal of psychiatry*, 52:102066, 2020.

[126] Megan Ranney, Kyler Lehrbach, Nicholas Scott, Nicole Nugent, Alison Riese, Jeff Huang, Grant Fong, and Rochelle Rosen. Insights into adolescent online conflict through qualitative analysis of online messages. page 9. Proceedings of the 53rd Hawaii International Conference on System Sciences, 2020.

[127] Andrew G. Reece and Christopher M. Danforth. Instagram photos reveal predictive markers of depression. *EPJ Data Science*, 6(1):1–12, December 2017.

[128] Andrew G. Reece, Andrew J. Reagan, Katharina L. M. Lix, Peter Sheridan Dodds, Christopher M. Danforth, and Ellen J. Langer. Forecasting the onset and course of mental illness with Twitter data. *Scientific Reports*, 7(1):13006, December 2017.

[129] Dan Russell, Letitia Anne Peplau, and Mary Lund Ferguson. Developing a measure of loneliness. *Journal of personality assessment*, 42(3):290–294, 1978.

[130] M Sadiq et al. Learning pakistani culture through the namaz emoji. In *2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, pages 1–8. IEEE, 2019.

[131] Koustuv Saha, Larry Chan, Kaya De Barbaro, Gregory D Abowd, and Munmun De Choudhury. Inferring mood instability on social media by leveraging ecological momentary assessments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(3):95, 2017.

[132] Gaia Sampogna, Ioannis Bakolis, Sara Evans-Lacko, Emily Robinson, Graham Thornicroft, and Claire Henderson. The impact of social marketing campaigns on reducing mental health stigma: Results from the 2009–2014 time to change programme. *European Psychiatry*, 40:116–122, 2017.

[133] Lara Schreurs and Laura Vandenbosch. Should i post my very best self? the within-person reciprocal associations between social media literacy, positivity-biased behaviors and adolescents' self-esteem. *Telematics and Informatics*, page 101865, 2022.

[134] Sandra Servia-Rodríguez, Kiran K Rachuri, Cecilia Mascolo, Peter J Rentfrow, Neal Lathia, and Gillian M Sandstrom. Mobile sensing at the service of mental well-being: a large-scale longitudinal study. In *Proceedings of the 26th International Conference on World Wide Web*, pages 103–112, 2017.

[135] Saul Shiffman, Arthur A. Stone, and Michael R. Hufford. Ecological momentary assessment. *Annual Review of Clinical Psychology*, 4(1):1–32, 2008. PMID: 18509902.

[136] Saul Shiffman, Arthur A Stone, and Michael R Hufford. Ecological momentary assessment. *Annu. Rev. Clin. Psychol.*, 4:1–32, 2008.

[137] Sally A Shumaker and Arlene Brownell. Toward a theory of social support: Closing conceptual gaps. *Journal of social issues*, 40(4):11–36, 1984.

[138] Lauren Sippel, Robert Pietrzak, Dennis Charney, Linda Mayes, and Steven Southwick. How does social support enhance resilience in the trauma-exposed individual? *Ecology and Society*, 20:10, 10 2015.

[139] Aaron Smith and Monica Anderson. Social Media Use 2018: Demographics and Statistics, March 2018.

[140] Brian G Smith, Staci B Smith, and Devin Knighton. Social media dialogues in a crisis: A mixed-methods approach to identifying publics on social media. *Public relations review*, 44(4):562–573, 2018.

[141] William E Snell, Rowland S Miller, and Sharyn S Belk. Development of the emotional self-disclosure scale. *Sex Roles*, 18(1-2):59–73, 1988.

[142] Cecilia H Solano and Mina Dunnam. Two's company: Self-disclosure and reciprocity in triads versus dyads. *Social Psychology Quarterly*, 48:183–187, 1985.

[143] Changwon Son, Sudeep Hegde, Alec Smith, Xiaomei Wang, and Farzan Sasangohar. Effects of covid-19 on college students' mental health in the united states: Interview survey study. *Journal of medical internet research*, 22(9):e21279, 2020.

[144] Hayeon Song, Anne Zmyslinski-Seelig, Jinyoung Kim, Adam Drent, Angela Victor, Kikuko Omori, and Mike Allen. Does facebook make you lonely?: A meta analysis. *Computers in Human Behavior*, 36:446–452, 2014.

[145] Deborah M. Stringer. *Negative Affect*, pages 1303–1304. Springer New York, New York, NY, 2013.

[146] Satomi Sugiyama. Kawaii meiru and maroyaka neko: Mobile emoji for relationship maintenance and aesthetic expressions among japanese teens. *First Monday*, 2015.

[147] Yla R. Tausczik and James W. Pennebaker. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*, 29(1):24–54, March 2010.

[148] Shelley E Taylor. Social support: A review. 2011.

[149] Steven Taylor. *The psychology of pandemics: Preparing for the next global outbreak of infectious disease*. Cambridge Scholars Publishing, Newcastle upon Tyne, 2019.

[150] Auke Tellegen. Structures of mood and personality and their relevance to assessing anxiety, with an emphasis on self-report. page 681–706, 1985.

[151] The New York Times. Coronavirus world map: Tracking the global outbreak. 2022.

[152] Julio Torales, Marcelo O'Higgins, João Mauricio Castaldelli-Maia, and Antonio Ventriglio. The outbreak of covid-19 coronavirus and its impact on global mental health. *International Journal of Social Psychiatry*, 66(4):317–320, 2020.

[153] William Tov, Kok Leong Ng, Han Lin, and Lin Ge Qiu. Detecting well-being via computerized content analysis of brief diary entries. *Psychological assessment*, 25 4:1069–78, 2013.

[154] Timothy J Trull, Marika B Solhan, Sarah L Tragesser, Seungmin Jahng, Phillip K Wood, Thomas M Piasecki, and David Watson. Affective instability: measuring a core feature of borderline personality disorder with ecological momentary assessment. *Journal of Abnormal Psychology*, 117(3):647, 2008.

[155] Matthew T Tull, Keith A Edmonds, Kayla M Scamaldo, Julia R Richmond, Jason P Rose, and Kim L Gratz. Psychological outcomes associated with stay-at-home orders and the perceived impact of covid-19 on daily life. *Psychiatry research*, 289:113098, 2020.

[156] Danny Valdez, Marijn Ten Thij, Krishna Bathina, Lauren A Rutter, and Johan Bollen. Social media insights into us mental health during the covid-19 pandemic: Longitudinal analysis of twitter data. *Journal of medical Internet research*, 22(12):e21418, 2020.

[157] Cornelia F van Uden-Kraan, Constance HC Drossaert, Erik Taal, Bret R Shaw, Erwin R Seydel, and Mart AFJ van de Laar. Empowering processes and outcomes of participation in online support groups for patients with breast cancer, arthritis, or fibromyalgia. *Qualitative health research*, 18(3):405–417, 2008.

[158] Nina Vindegaard and Michael Eriksen Benros. Covid-19 pandemic and mental health consequences: Systematic review of the current evidence. *Brain, behavior, and immunity*, 89:531–542, 2020.

[159] Shaun Wallace, Brendan Le, Luis A Leiva, Aman Haq, Ari Kintisch, Gabrielle Bufrem, Linda Chang, and Jeff Huang. Sketchy: Drawing inspiration from the crowd. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2):1–27, 2020.

[160] Cuiyan Wang, Riyu Pan, Xiaoyang Wan, Yilin Tan, Linkang Xu, Cyrus S Ho, and Roger C Ho. Immediate psychological responses and associated factors during the initial stage of the 2019 coronavirus disease (covid-19) epidemic among the general population in china. *International journal of environmental research and public health*, 17(5):1729, 2020.

[161] Cuiyan Wang, Riyu Pan, Xiaoyang Wan, Yilin Tan, Linkang Xu, Roger S McIntyre, Faith N Choo, Bach Tran, Roger Ho, Vijay K Sharma, et al. A longitudinal study on the mental health of general population during the covid-19 epidemic in china. *Brain, behavior, and immunity*, 87:40–48, 2020.

[162] Xiaomei Wang, Sudeep Hegde, Changwon Son, Bruce Keller, Alec Smith, and Farzan Sasangohar. Investigating mental health of us college students during the covid-19 pandemic: cross-sectional survey study. *Journal of medical Internet research*, 22(9):e22817, 2020.

[163] Akash R Wasil, Emma H Palermo, Lorenzo Lorenzo-Luaces, and Robert J DeRubeis. Is there an app for that? a review of popular apps for depression, anxiety, and well-being. *Cognitive and Behavioral Practice*, 29(4):883–901, 2022.

[164] David Watson, Lee Anna Clark, and Auke Tellegen. Development and validation of brief measures of positive and negative affect: the panas scales. *Journal of Personality and Social Psychology*, 54(6):1063–1070, 1988.

[165] Jiaqi Xiong, Orly Lipsitz, Flora Nasri, Leanna Lui, Hartej Gill, Lee Phan, David Chen-Li, Michelle Iacobucci, Roger Ho, Amna Majeed, and Roger McIntyre. Impact of covid-19 pandemic on mental health in the general population: A systematic review. *Journal of Affective Disorders*, 277, 08 2020.

[166] Yong Sook Yang, Gi Wook Ryu, and Mona Choi. Methodological strategies for ecological momentary assessment to evaluate mood and stress in adult patients using mobile phones: Systematic review. *JMIR mHealth and uHealth*, 7(4), 1 2019.

[167] Yusen Zhai and Xue Du. Addressing collegiate mental health amid covid-19 pandemic. *Psychiatry research*, 288:113003, 2020.

[168] Ben Zefeng Zhang, Tianxiao Liu, Shanley Corvite, Nazanin Andalibi, and Oliver L Haimson. Separate online networks during life transitions: Support, identity, and challenges in social media and online communities. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2):1–30, 2022.

[169] Jason Shuo Zhang, Brian C. Keegan, Qin Lv, and Chenhao Tan. A tale of two communities: Characterizing reddit response to covid-19 through /r/china_flu and /r/coronavirus, 2020.

[170] Renwen Zhang. The stress-buffering effect of self-disclosure on facebook: An examination of stressful life events, social support, and mental health among college students. *Computers in Human Behavior*, 75:527–537, 2017.

[171] Renwen Zhang, Natalya N. Bazarova, and Madhu Reddy. Distress disclosure across social media platforms during the covid-19 pandemic: Untangling the effects of platforms, affordances, and audiences. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 21(644), 2021.

[172] Gregory D Zimet, Nancy W Dahlem, Sara G Zimet, and Gordon K Farley. The multidimensional scale of perceived social support. *Journal of personality assessment*, 52(1):30–41, 1988.

[173] Gregory D Zimet, Nancy W Dahlem, Sara G Zimet, and Gordon K Farley. The multidimensional scale of perceived social support. *Journal of personality assessment*, 52(1):30–41, 1988.

[174] Gregory D Zimet, Suzanne S Powell, Gordon K Farley, Sidney Werkman, and Karen A Berkoff. Psychometric characteristics of the multidimensional scale of perceived social support. *Journal of personality assessment*, 55(3-4):610–617, 1990.

[175] Jolene Zywica and James Danowski. The faces of facebookers: Investigating social enhancement and social compensation hypotheses; predicting facebook™ and offline popularity from sociability and self-esteem, and mapping the meanings of popularity with semantic networks. *Journal of Computer-Mediated Communication*, 14(1):1–34, 2008.