🎓🙏☺️: Building Online Social Support Systems Through Investigation of Undergraduate Private Messaging Spaces

Talie Massachi

Department of Computer Science
Brown University

talie_massachi@brown.edu
https://talie.town

Version 1.0.1

October 30, 2024

Thesis Committee:

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

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 dissertation 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.

This dissertation includes three studies investigating how users self-disclose in online spaces, as well as how that disclosure relates to perceived social support. An exploratory study on online communication habits of undergraduate students before and during the COVID-19 pandemic 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 using the data extraction platform Sochiatrist investigates the accuracy of emotional disclosure over private and public online communication platforms. Compared to previous work, this study shows private online messages to more accurately represent emotional state than posts in public online spaces. Following these two studies, Chirp was developed as a social app centering privacy, perceived control over information spread, and ability to reflect on emotional self-disclosures over time. User feedback of Chirp gives evidence that when we leverage these aspects of private communication, users can form meaningful relationships which can lead to social support, even in semi-public, anonymous spaces with extremely limited communication.

Through these studies, this dissertation investigates how aspects that define private online communication may lead to more social support and connection-building behavior. Specifically, it focuses on the aspects of written communication, audience selection, and social presence. Showing that integrating these aspects of private online communication into a public online communication space can lead to connection and support between users.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Thesis Statement

As a widely-known protective measure for mental health challenges, perceived social support is the direct and indirect focus of many programs aimed at increasing undergraduate mental health. While traditionally these programs have focused on in-person connection-building, for example through student support groups, we instead question how perceived social support can be impacted in increasingly popular online communication spaces. Previous work shows a relationship between emotional self-disclosure in semi-public online spaces and perceived social support, thus leaving a ripe space to investigate the impact of private online communication spaces (such as messaging) on self-disclosure and subsequently perceived social support.

This dissertation investigates three specific aspects that define private messaging spaces: written communication, control over ones audience, and social presence. In a survey of students during the height of the COVID-19 pandemic university lockdowns, a comparison is drawn between online communication patterns before and during the pandemic across both written, audio, and video platforms. This comparison is used to investigate the impacts of written versus other kinds of communication. A study on private messaging data collected through the Sochiatrist platform investigates how users express themselves in online spaces with a controlled audience (private) versus public spaces, showing that users' expressed sentiments in private spaces are more accurate to their actual emotional state. Finally, the emoji-based social-mood tracking app Chirp is used to investigate how social presence affects online communication, finding that the inclusion of social presence encourages more consistent communication.

Ultimately, this dissertation proposes that by leveraging these aspects of written communication, control over audience, and social presence, online communication spaces can become more supportive for undergraduate students.

1.1 Motivation

With rising rates of known or diagnosed mental health disorders, such as depression [67] and anxiety [68], mental health has recently been centered in the US national consciousness [69, 127]. This problem has been particularly prevalent among school- and college-age students [67, 68]. In response to this trend there has been a rising focus on social connection building [127], particularly in online

spaces where people around college age are increasingly spending more time [154] and looking for connection [70].

This dissertation presents 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 [52], encourage emotional self-disclosure [5], and provide a space for users to find camaraderie and support [7]. Self-disclosure has further been recognized as a fundamental human need [16], and is associated with intimacy and relationship strength [132, 157]. Conversely, social media use has been associated with negative mental health effects, such as negative body image [61], loneliness [139], 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 [73, 93, 139], Social Comparison Theory [105], Interpersonal-Connection-Behaviors Framework [37]). However, we note that previous work has shown self-disclosure to be most frequent amongst dyads [157], and that many studies investigating self-disclosure and social support in online spaces have focused on *semi-private* spaces with shared context (e.g., [4, 66, 187]). This hints that the context or aspects of the *communication space itself* may have an impact on user outcomes.

Thus this dissertation investigates 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

1.2.1 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.

1.2.2 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.

1.2.3 Not So Different After All: Demographic Differences in Emoji Interpretation

Talie Massachi*, Anika Ahluwalia*, Audrey Chou, Jeff Huang Emojis are an integral part of online communication, yet it is widely held that demographics of both the interpreter reading an emoji and the original poster of that emoji bias its interpretation. Here we investigate the degree to which demographic factors impact the interpretation of structured emoji-only strings for interpreters and posters based in the US. In particular, we focus on interpreter ability to understand emoji-only mood tracking posts. Through a survey of 163 participants, providing a total of 520 emoji interpretations, we find interpreters accurately understand on average 46% of intended meaning behind these emoji strings. Interpreters' demographics have no effect on interpretation accuracy, nor does knowledge of the original poster's demographics. Tellingly, older participants and those working full-time are significantly less *confident* in their interpretations than other participants. Our findings indicate that, in this mood tracking context, demographic factors play no role in emoji interpretation, but instead impact confidence. Therefore, emoji remain a viable form of communication for users when sharing emotional state.

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

Talie Massachi, Lauren Choi, John Roy, Gabriela Hoefer, Shaun Wallace, Jeff Huang. In proceedings of CSCW 2024.

This work seeks to better understand relationships between attributes of private messaging and emotionally beneficial outcomes such as increased perceived social support. To do so we built CHIRP, an anonymous, emoji-based 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 aspects of online private messaging spaces. In this way, CHIRP provides a way to safely investigate the effects of specific design decisions and factors influencing social media and online communication through direct experimental interventions.

Through a between-subjects study among a cohort of first-year undergraduate student users on CHIRP, we first evaluate the effect of self-disclosure both in a non-social diary space as well as a semi-private online space, allowing us to investigate how the **presence of other users** impacts emotional state and posting patterns. We find that users in social spaces post more regularly, skipping days less than half as often. We then investigate how integrating aspects of private messaging into a semi-private online space in CHIRP impacts feelings of social support. While CHIRP use does not show a significant increase in measured feelings of social support, user responses suggest that self-disclosure in CHIRP may provide more social support than typical social media use or simple mood tracking over a two-week period. Our findings indicate that even in low-bandwidth spaces, self-disclosure through written posts in spaces with **social presence** of other users may lead to feelings of social support and connection.

Chapter 2

Background

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 [152], social support, and specifically *perceived* social support, has been tied to many metrics of mental, emotional, and physical well-being [162]. This includes acting as a protective factor that improves emotional resilience in response to trauma [153]. Research has shown that feelings of social support reduce rates of depression and anxiety [162], and are tied to decreased feelings of loneliness [92].

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., [44, 66, 90, 173]), including the creation of social support networks over social media platforms (e.g., [7, 66]). Studies have also shown strong correlations between direct online communication with friends or family and feelings of social support [26, 81].

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 [93, 159]. An exception to this, Deters and Mehl [52] 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. While Deters and Mehl investigate loneliness and in this dissertation we study perceived social support, this finding shows a direct causal link between social media use and positive mental health outcomes.

2.1.1 The Impact of Audience Reciprocation in Online Spaces on Social Support

Given that social media and online communication are implicitly social, it is intuitive that reciprocation or back-and-forth communication is a necessary factor when leading to positive outcomes such as perceived social support. Indeed, some previous work has shown a relationship between reciprocal

actions and positive outcomes.

For example, when looking at the relationship between Facebook poster satisfaction and quantity and quality of responses (i.e., likes and comments), Bazarova et al. find that users feel "more satisfied overall when they receive more likes and gratifying comments (i.e., reply satisfaction), while number of comments influences overall satisfaction indirectly by enhancing reply satisfaction" [17]. In contrast, in a between-subjects longitudinal study investigating the impact of increased Facebook posting on feelings of loneliness, Deters and Mehl found that participants who posted on Facebook more frequently showed lower levels of loneliness, *regardless of whether those posts were interacted with or responded to by their network* [52]. 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 [52].

2.1.2   Measures of Social Support

Traditionally, perceived social support is measured by the Multidimensional Scale of Perceived Social Support (MSPSS) [191, 193] (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 [126] (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 dissertation, 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 [114, 190]. 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 [114, 190]. Even fewer studies on mental health have differentiated between platforms [58, 63, 82, 107], despite previous work showing different interaction types between public and private communications [25]. Of the few studies that differentiate, private platforms have emerged as more beneficial for loneliness than public platforms [122]. Their benefits are often credited to the greater perception of anonymity and privacy [190]. 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 [157].

Thus this dissertation focuses 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., [6,37,77]). 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 [132]. Studies have also found strong direct correlations between the frequency of emotional self-disclosure online and feelings of social support [52,81,103,104]. However, the specific aspects of online platforms that lead to these positive outcomes are unknown. For example, whether the presence of other users is necessary for these positive impacts has been called into question [52]. This dissertation therefore investigates the individual aspects of written communication, control over audience, and social presence and how they impact emotional self-disclosure and perceived social support.

One prominent theory of online interaction that tries to explain the relationship between online communication and mental health outcomes is the Interpersonal-Connection-Behaviors (ICB) framework [37]. 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, [124]"(citation theirs) [37].

2.3 Emoji as Communication

2.3.1 Demographic Impacts on Emoji Interpretation

Emojis— or symbols used in place of or in addition to written communication representing facial expressions, ideas, objects, and other concepts— are becoming more common as a part of written conversation. In 2017, Facebook reported that 92% of teens and 77% of those aged 55-64 used emojis [24]. Despite their common use, there has been much debate about the consistency of their interpretations and intended meaning.

Emojis are often rendered differently based on platform, something that has been shown to impact interpretation [120, 165]. Additionally, emoji have been shown to have different usage conventions among different groups [178], including gender and age [24,33,182]. With studies showing that women use emojis more often than men [34, 91], and that younger users use emojis more often than older users [24]. Older users have also been shown to interpret emojis differently from younger users [64,182], as well as less accurately than younger users [33]. Country of residence in particular has been shown to strongly influence usage and interpretation of emojis, with residents of different countries using emojis with different frequencies [74], using different sets of emojis [33,74], and using similar or the same emojis with different meanings [36,100,178].

In contrast, some studies have found no differences in emoji interpretation accuracy based on demographic factors. For example, Herring and Dainas find no difference between female and male interpreters when interpreting the function of emojis in a given message [80].

2.3.2 Interpretation in Emoji-Only Communication

While many studies examine emoji interpretation, most focus on interpretations in the context of other text. However, there also exist of a number of emoji-only apps and projects, where emoji are used without the context of text, for example, Emojli [1], G-Moji [172], and Opico [94].

There have been few investigations into emoji understanding and interpretation in emoji-only contexts. Most notably, Khandekar et al. used their system, Opico, to investigate how users interpret emoji-only "reactions" (i.e. posts about a users' emotional state, tied to a specific location and temporal marker) [94]. The authors find that 89% of reactions have at least one correct interpretation made by other users (out of a maximum of 2), indicating that given location and temporal context their emoji-only reactions are fairly understandable [94]. Additionally, Van Dam et al. developed G-Moji to test the viability of emoji as an emotional assessment tool in studies with adolescents [172]. They report qualitative responses indicating that emojis may be a useful platform for emotion reporting. However they also report a lack of confidence that the emojis would be interpreted correctly [172].

---

[1]https://emoj.li/

Chapter 3

Offline to Online Social Support during the COVID-19 Pandemic

This chapter is substantially similar to [81], where I was co-first author and contributed substantially to study design, narrative, and writing.

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 [171]. At times, the United States had the highest number of reported cases of any country [166], shutting down college campuses nationwide [32]. 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 [22]. 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," [136] 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" [101].

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., [141, 174, 176]), often attributed to psychosocial stressors associated with unforeseen increased isolation, such as life disruption (e.g., [106, 170]) and fear of illness or uncertainty (e.g., [58, 158]), therefore calling for increased prioritization of mental health (e.g., [167, 183]). 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 [176]. Furthermore, they conducted a follow-up study revealing that these symptoms persisted even a month later [177]. 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 [170]. These reports of declining mental health are replicated in studies from Australia [59], the United Kingdom [23], and among college students worldwide (e.g., [28, 58, 106, 108]). Furthermore, individuals in social isolation around the world reported lack of access to adequate mental healthcare, relying primarily on self-help techniques [46]. This matches trends in previous outbreaks, as research during SARS revealed higher levels of stress, anxiety, and depression in the general population [35], and patterns across previous pandemics reveal a lack of adequate mental health resources worldwide during or immediately following the outbreak [163].

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 [14, 186]. Moreover, many students reported a decline in feelings of social belonging and community connectedness [186]. Current studies show college students report higher levels of stress, anxiety, and depression than the general population during the COVID-19 pandemic (e.g., [28, 43, 58, 158, 179]). 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 [106] and another study found similar, persistent mental health trends in college students over multiple months of the pandemic [43]. 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" [76]. After almost every disaster we see a regional unification of communities, materials, and information, known as geographical convergence [62]. Studies have found this behavior mirrored in online interactions, as individuals gravitate towards online communities as a coping mechanism (e.g., [54, 129, 155]) 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., [63, 128, 171]). 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 [128]. Similar patterns have also been found in Italy, revealing significant increases across video calls and online gaming platforms [63], and in the United States, revealing an increase in posting across public social media platforms [171].

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 [117]. 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 [15]; and the sentiment of public-facing social media posts was found to be significantly more negative [171]. Self-disclosure, "an interaction between at least two individuals where at least one intends to deliberately divulge something personal to another," [71] is associated with increased stressful life events [189] and recognized as a fundamental human need [16]. Furthermore, self-disclosure is associated with intimacy and relationship closeness, and is most frequent between dyads [157], 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 [132] and facilitate social support [4]. Current work suggests there may

be an increased likelihood of willingness to disclose during the COVID-19 pandemic [123], 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 [153]. During the COVID-19 pandemic, increased social support was also shown to decrease crisis-related stress symptoms such as anxiety and depression [125].

Studies reveal online communities to be an effective method for receiving social support [4], 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 [179], and digital platforms were effective in truly increasing perceived social support during this time [63,108]. Despite the positive effects of online interactions, other studies have found participants tend to believe virtual platforms are inadequate substitutes for in-person conversation [137]. 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 [122], yet others reveal associations between increased public social media usage and increased feelings of loneliness [108,140]. More recent research has also found increased levels of support seeking behavior on public social media platforms [108] and have found individuals are likely to be using public platforms as a coping mechanism to compensate for existing feelings of loneliness [122, 133], rather than social media causing feelings of loneliness [171]. 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 [171]; 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) [147]. 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 [37]. 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) [37]. Active interactions are associated with greater positive outcomes, whereas passive interactions present negative associations such as increased feelings of loneliness [38]. 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 [137], 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 [21]. 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 assessing 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 [186]. Moreover, research shows this age group uses online platforms more than any other age group in the US [30], 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) [156] 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) [156] 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 [79]. This measurement, as well as the 20-item long form scale, has been widely used in the research community as a measurement of loneliness [145]. 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 [192]. 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 [27] and it is a popular tool for measurement during the COVID-19 pandemic specifically [106, 170]. 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 [30]. 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 [15, 188]. 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 [56]. 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 [32], 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 [32].

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 [116], 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 [83] 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 [86].

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

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.

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 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 ($r = -0.41$,
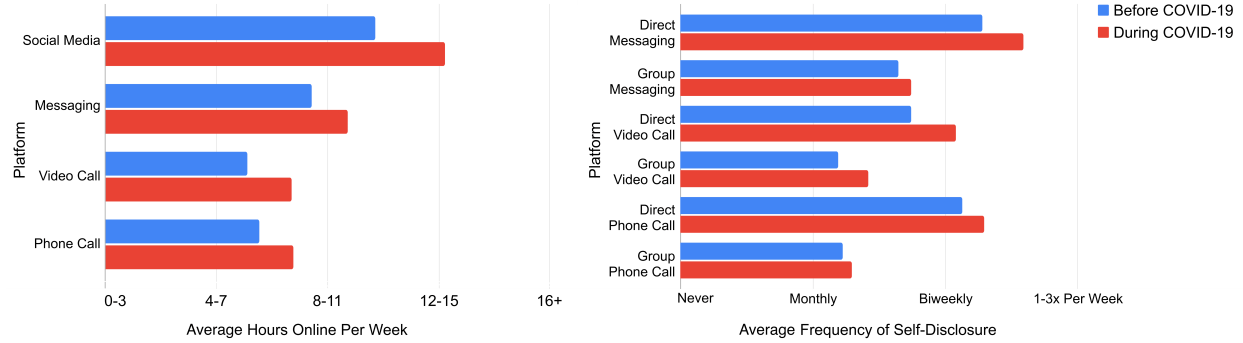
Figure 3.1: The average self-reported online interaction frequency was compared before and during the COVID-19 pandemic. The average hours spent interacting on each platform per week (left), as well as frequency of self-disclosure (right) on each platform was calculated. College students saw an increase 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.

$p < 0.001$), 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 [108]. 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.

Emotions Shared Based on Frequency of General Messaging Disclosure
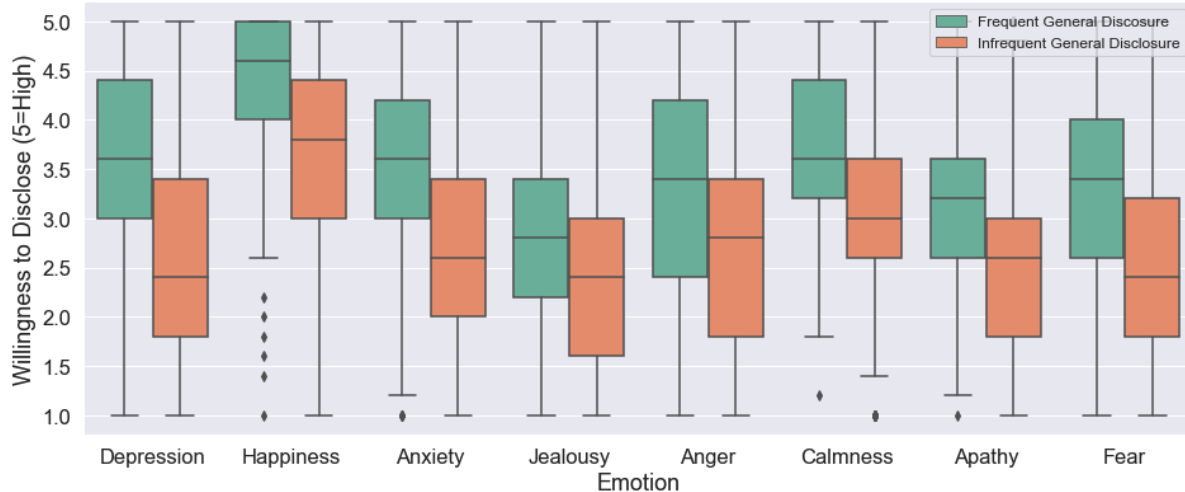
Figure 3.2: Data is grouped by those disclosing over messaging platforms at least weekly (frequent) or those disclosing monthly or less (infrequent). The specific emotions shared are plotted, revealing significant differences between groups across all emotions, with greater increases for more negative 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.

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

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 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) [156] 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

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.

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., [13, 63, 128]), 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 [171].

Given the large body of contrasting literature into the implications of social media use on various constructs related to mental health (e.g., [63, 122, 140, 147, 171]), 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 [129, 155]. 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 [122], which suggests people gravitate towards online interactions to compensate for in-person social deficits [117]. Some people may attempt to increase their popularity online to compensate for their perceived lack of offline popularity [194] or increased introversion in offline social settings [117]. 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 [37, 38].

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 [21]. 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 com-

munication. 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 [102], bringing about nonintervention known as the "bystander effect" [65]. Recent work suggests this phenomena is linked to social media behavior and online victimization [112]. 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 [53, 148], 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 [87]. 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 [87]. 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 [23] despite evidence of increased online communication during this time [63, 128, 171]. 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 [58]. 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 [30]. 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 [63, 128, 171], 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 positive 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

Signals of Affect in Messaging Data

This chapter presents Sochiatrist, a system for safe and anonymized extraction of cross-platform public and private social networking data. This chapter is substantially similar to [114], where I was co-first author and responsible for system maintenance, analysis of study data, narrative, and most of the writing.

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, 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 [53, 148], as well as evidence that users often consider self-presentation more when posting publicly, or to a wider audience [18, 95]. 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 [181]).

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 [168] and speech [40]). 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 [164, 181] 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 [119], while negative affect refers to the experiences of negative moods and emotions, such as anxiety, sadness, fear, anger, guilt, and shame [160]. 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 [181]. Common techniques used to measure positive and negative affect include the Positive And Negative Affect Schedule (PANAS) [181].

The widely used 20-item version of the PANAS has been shown to be an internally consistent self-report assessment of affect [11, 169, 181]. 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 [181].

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 [134]. Additionally, Valence Aware Dictionary for sEntiment Reasoning (VADER) is a sentiment analysis model specifically built for use in contexts such as social media posts [84].

Unlike PANAS, LIWC and VADER are indirect measures used to automatically estimate affect. LIWC is text analysis system developed in 1993 [135] 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 [135]. Updates to the scale since then have been based on cycles of expert analysis and further methods of relevant term discovery [161].

VADER is a freely available rule-based sentiment analysis tool built to improve upon existing techniques, including LIWC and human review [84]. 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 [19, 20].

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 [19,20,39,168]. 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. [168] 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. [39] 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 [19] 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 [19]. 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. [20] 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 [19]. 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 [48, 51, 55, 88, 131, 144], but some studies examined platforms such as Instagram [143] and Facebook [45]. Still others focus on two or more social media platforms [19, 20], 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) [154]. 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% [10].

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, 110, 131]. User preference for sharing emotions more openly in public posts or private messages seems to be influenced by platform [16, 131]. 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 [131]. 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 [110]. Furthermore, studies have shown that people often present themselves differently on their public social media in order to self-promote (e.g., [99]) 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 [19, 20]. In many ways, private messages are more similar to direct speech [39] than to public social media posts. So based on the success by Cohen et al., private messages may be a viable textual 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.
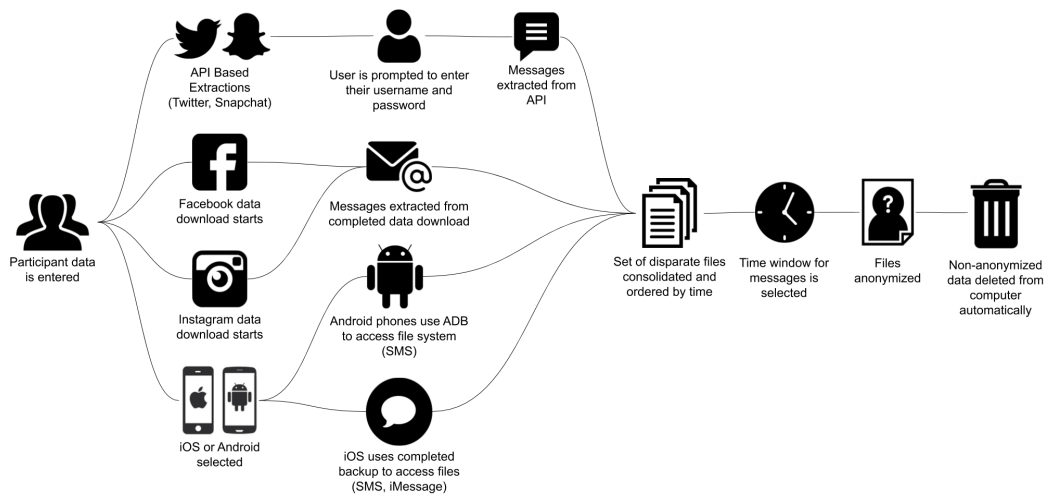
Figure 4.1: Sochiatrist extracts, consolidates, and pseudonymizes the data to develop models predicting affect based on messaging data.

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 [42, 130]. 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. [146], 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., [72, 142]).

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.

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

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.

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 [60]. 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.

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, 185]. 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 [150]. 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 [151].

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"

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.

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.

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. [181] 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.
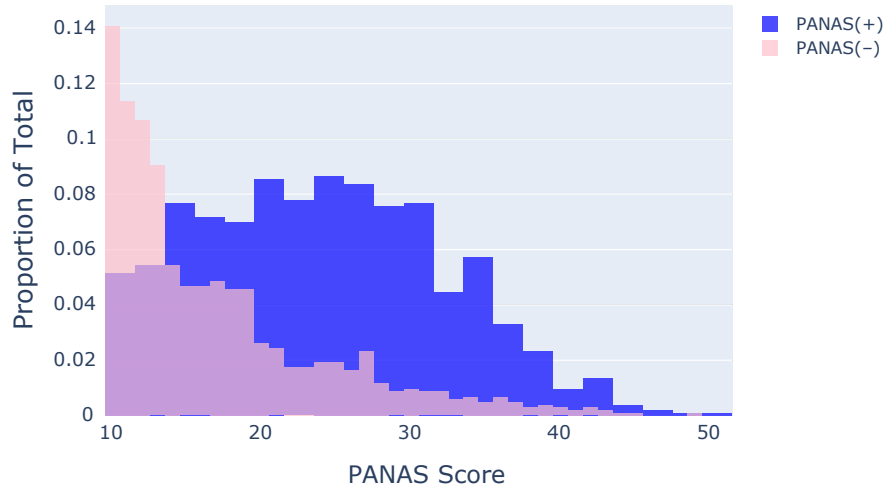
Figure 4.2: Histogram for PANAS(−) (pink) compared to the histogram for PANAS(+) (blue), with the overlap in purple. Note that PANAS(−) has a lower mean than PANAS(+) and is a more skewed distribution.

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 (±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.

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 ($\pm 2,4,8$ 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 [19, 45, 49, 50, 51, 55, 96, 97, 98, 121, 131, 144, 168]. 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 [84]. 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 (±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. [181]. 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. [181]. Before any discussion, reviewers also each individually rated a set of messages with a shorter time window (±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 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 [19, 20, 39, 168]. 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.
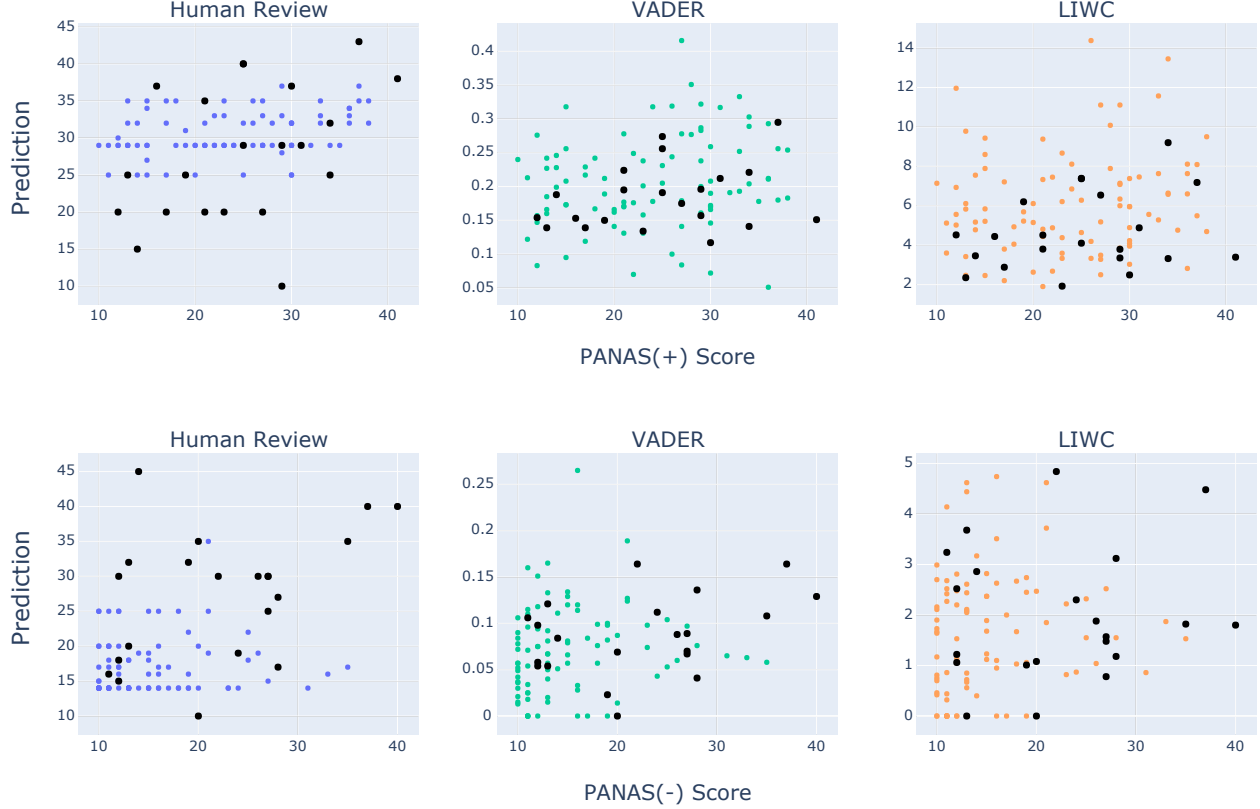
Figure 4.3: Self-reported PANAS(+) and PANAS(−) scores plotted against predicted scores. The x-axis is PANAS(+) in the top plots and PANAS(−) in the bottom plots while the y-axis is the algorithm prediction (each technique uses a different scale). The dots in black are the points labeled confident by human reviewers.

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

	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) $*p < 0.05$, $**p < 0.01$

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 [53], 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.

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 [84, 135], 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.

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".

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 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 [84], 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 [181]. 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 [19, 20], personal diary entries [168], or spoken responses [39]. 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.

| | 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(−). *$p < 0.05$, **$p < 0.01$

Across all studies discussed, LIWC analysis over natural speech performed with the highest accuracy [39]. 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) [111]. 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.

The second most successful technique found in previous studies was LIWC analysis over short daily diary entries [168]. 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 [39]. 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. [19, 20, 48, 51, 55]), others only collect data participants specifically record for the study (e.g. diary entries [168] or recorded speech [39]), while yet others ask participants to copy over messages by hand in the lab [18]. 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 [113] to 3 years [149]. 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

Integrating Features of Private Communication in Public Online Spaces

This chapter presents Chirp, a semi-public, emoji-based, social, mood tracking platform. Chirp is built as a sandbox space to better understand the impact of different design decisions on users. The app is available for use and all code is open source [1]. This chapter is substantially similar to [115], where I was first author and responsible for initial design, all aspects of the user study, analysis, narrative, and most of the writing. Additionally, this chapter includes summarized sections from a paper currently in submission [?].

5.1   Introduction

In previous chapters we discussed the potential for online private messaging spaces to support feelings of social support in undergraduate students. We discuss how the persistence of written communication and may allow for reflection, as well as how private spaces may encourage more honest emotional sharing. However, taking these factors into account, it is unclear how the impact on perceived social support differs between keeping a diary and private messaging.

In this chapter, we thus question how the presence of other users impacts perceived social support. Many studies show that, inherently social, semi-public and semi-private online support groups contain significant self-disclosure and socially-supportive communication (e.g., [4, 7, 66, 77, 187]). Similarly, Yang et al. found that users in online cancer support communities were more likely to self-disclose in public spaces, while self-disclosure and reciprocation in private (one-on-one) channels was more strongly correlated to higher levels of actual social support than self-disclosure in public channels [184]. These findings indicate that the presence of others— either in being part of the public space or in responding in a private space— have an impact on social support. However, the necessity of social interaction for positive outcomes has been called into question. Deters and Mehl found that, while increased posting frequency on Facebook decreased feelings of loneliness, loneliness was not affected by response or engagement rate on posts [52]. This leads to our first research question:

RQ1) How do responses and reciprocal emotional self-disclosures from other students impact perceived social support among students who regularly emotionally self-disclose in private or semi-private

---
[1]https://chirp.cs.brown.edu

online spaces?

Furthermore, while previous chapters have investigated individual factors of private online communication and how they may impact users (i.e. written communication and restricted audiences), these studies have focused on existing platforms with pre-existing cultures. Messaging patterns and user outcomes may be influenced on user expectations based on previous knowledge of these platforms. Here we question how perceived social support is impacted in a disclosure-based social media space where factors of private messaging are integrated, creating a semi-private space where users can interact with a cohort of similar users.

RQ2) How does the addition of regular emotional self-disclosure in a semi-private online space impact perceived social support among undergraduate students?

To test these research questions, we build a sandboxed platform called Chirp: a semi-private space that prompts cohorts (a restricted audience) of anonymous users (allowing for social presence) to emotionally self-disclose through emoji-based (a type of written communication) mood tracking posts (See Figure 5.1). Chirp is structured as a mobile application whose primary intention is to act as a safe and controlled space where we can study cohorts of users without the need to compromise participant privacy.

As Chirp uses emojis as its primary form of written communication, we further investigate the viability of emoji to transmit information between a poster and an interpreter. Previous work has found emoji to be understandable given context, thus here we confirm whether mood tracking posts provide enough context for interpretation. We additionally investigate the degree to which user demographic similarities impact interpretation.

5.2 Background and Related Work

Here we discuss existing apps, both commercial and in the CSCW and CHI community, that have been used to address questions around the impact of disclosure or support on mental health, well-being, or feelings of connection.

5.2.1 Existing Disclosure Systems

As Chirp is intended as a social media space in which to study self-disclosure by way of mood tracking, here we discuss existing alternatives. There are many commercial journaling and mood tracking apps that aim to improve users' daily routines and well-being (e.g., Daylio [31]). This also includes well-being apps with mood tracking features, such as Headspace[2] or Calm[3] [180]. 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 tracking systems similar to Chirp. The app Opico, developed by Khandekar et al., uses emojis as the basis of communication, much like we do in Chirp [94]. In the Opico app, users can select a location on a public map and leave a string of emojis as a "reaction" to it. While Opico has more similarity to public social media spaces and Chirp is more reflective of

---

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

private spaces, reactions in Opico can also encompass emotional or mood-related responses to certain locations. These reactions therefore show similarity to our "Chirp posts". Through the Opico app, the authors find that, given context, user interpretations of emojis can be accurate to the original intent.

Significant Otter [109], and IntimaSea [89] were both created as ways to communicate users' current emotional state to partners or friends that are already close with the user. Significant Otter uses biosignals detected through a smartwatch to suggest animated otter characters that users can send from their own smart watch to that of a partner [109]. By sending these animated characters, users can communicate without the use of text. Similarly, IntimaSea uses automatic stress tracking to share stress levels within a group of friends [89]. However, unlike 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 we address in Chirp, none specifically focus on emotional and mood disclosure and social support in semi-private social media settings. Additionally, popular social media platforms (such as Facebook and Twitter) have existing norms and relationships built between current users, which can impact the outcomes of users engaging with the platforms.

In this paper, we aim to isolate the impact of self-disclosure and reciprocation in semi-private spaces on perceived social support while avoiding confounding factors such as pre-existing relationships (such as those that exist in popular social media platforms). Doing so will allow us to better understand the individual factors in online communication that impact well-being, informing future development of systems that support user mental health. By leveraging Chirp, we can isolate individual factors while investigating their effects in a manner that would be difficult or impossible in existing systems.

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

Chirp is an open source[4] mood tracker[5] and social media mobile application that uses emojis as the only form of user expression. It is built as a 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 previous work around emotional self-disclosure and feelings of social support (e.g., [81,104]), Chirp is built to **encourage emotional self-disclosure**. 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**. This includes preventing

---

[4] https://Chirp.cs.brown.edu

[5] 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.

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 [9].

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.

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 the timeline feature) to 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 [75].

Across two separate phases of early user testing, testers were placed into a Discord server[6] 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 for research purposes, 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 [57]; as well as the six moods tested in the Profile of Mood States (POMS): tension, depression, anger, vigor, fatigue, and confusion [118]. 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?" as 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.

---

[6]https://discord.com/

5.3.1.2 Anonymity

Interactions over social media and other online communication are not solely positive [9]. Sharing personal information can lead to negative online interactions or bullying, for example, doxxing [8]. 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.

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) [75]. 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[7]. 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., [175]).

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 [94]. 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 is also a trend seen on many social media platforms. For example, Twitter's character limit, or Instagram's requirement that every post includes an image or video. We additionally note that not every social network relies on text, for example, the social networks Yo[8] or Emojili[9]. During user testing, testers expressed that they found the emoji-only requirement 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 used emojis from a third-party library called JoyPixels[10]. Thus, all emojis appeared identical for all app users.

5.3.1.4 Limiting Responses to Reactions

The ICB framework describes how active social media use involves connection promoting behaviors [37]. 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 (consider a user posting about a pet's death and the comment response "😂😂😛").

---

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

A different, very common feature in social media platforms is a system of reactions or likes [75]. Examples include reactions on Facebook or in Twitter DMs, likes/favorites on Instagram posts or Tweets, and upvotes/points systems such as on 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 [29, 78, 90]. 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. For these, we chose common reactions among 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. We identified that the angry react could be used as a negative response, and thus, it was not included as an option. 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 (as shown in Figure 5.1e). 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 appear 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
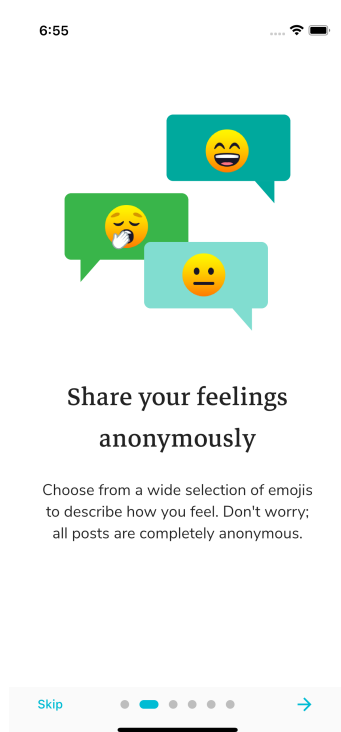
(a) Onboarding Screen
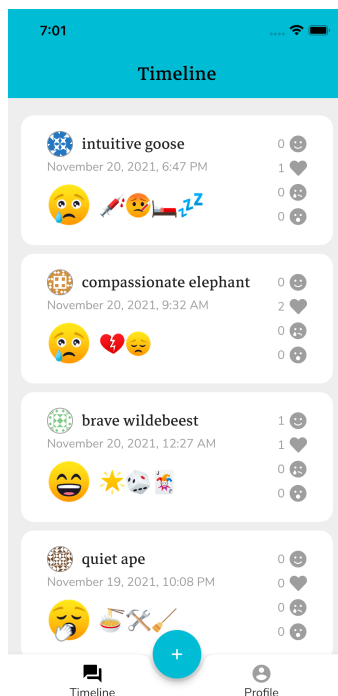
(b) Main User Timeline (Social Group Only)
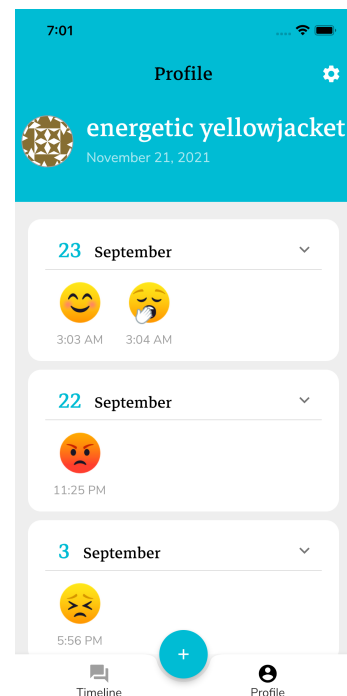
(c) User Profile (Social Group Only)
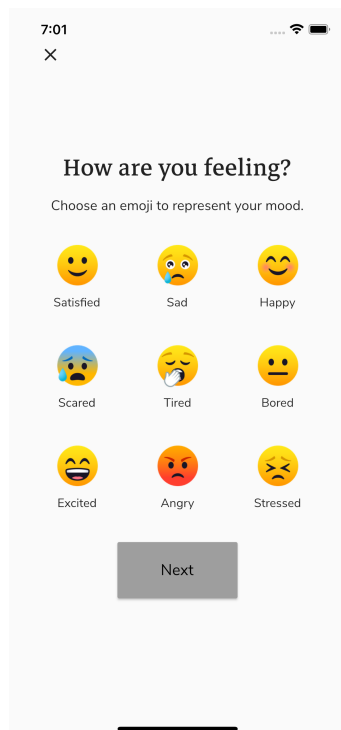
(d) Main Emoji Selection
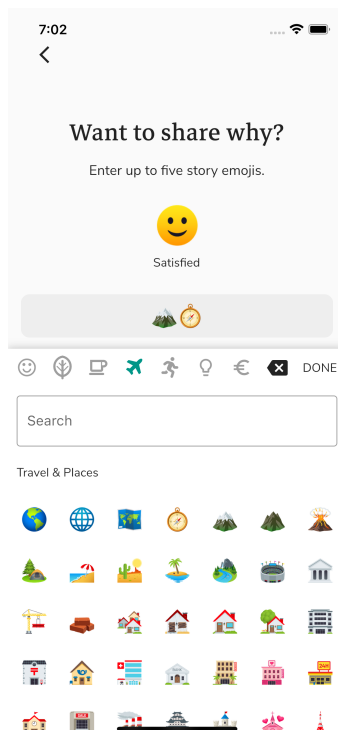
(e) Story Emoji Selection
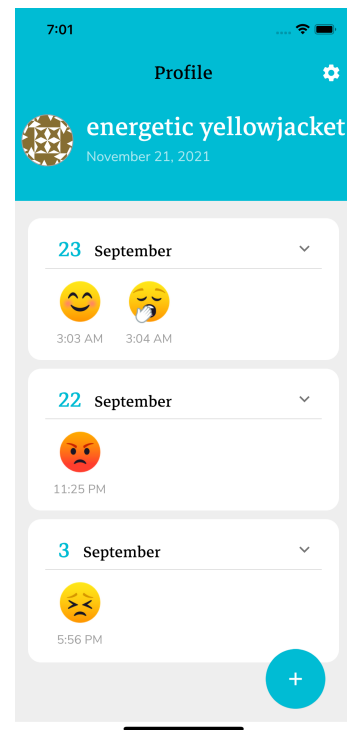
(f) Alternate User Profile (Individual Group Only)

Figure 5.1: Chirp application main pages as shown to study participants. Note that the main user timeline (b) is the default home page of the app for general users and participants in the Social group in this study. Participants in the Individual group in this study are unable to view (b), instead are presented with (f) as a replacement for (c) and as the default home page of the app.

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 this 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.

5.4 Methods: Impact of Chirp on Perceived Social Support

5.4.1 Participants

We recruited participants from a pool of first-year undergraduate students at the author's university. Physical posters were placed around our institution's campus, including in and around first-year dormitories, and in popular areas of congregation near campus. We also posted advertisements online in social media groups created as conversational spaces for first-year undergraduates from the authors' institution (i.e., Facebook groups, Instagram, Reddit). We required that all participants be 18 or older to be eligible to participate in the study.

While, in general, our research questions focus on university students, we limit to only first-year university students in this study as a control. Once they start college, many students are no longer physically near their primary social support network (e.g., parents, friends) and, therefore, are likely to benefit from an online source of social support. This is particularly true in the first year before they are able to build strong new connections with other students at their institution [12]. We restricted recruitment to a single university to ensure that all participants were under the same university-mandated COVID safety policy. Furthermore, as previously mentioned, previous studies show that emojis are understandable *given a shared context* [94], for example, within a specific Facebook group, shared friend group, or known location. By using a cohort of students from the same university and class year, we can ensure an appropriate level of shared context that would ensure understandable communication between participants.

In total, we recruited 77 consenting participants (43 female, 32 male, and 2 non-binary) based on the eligibility criteria. All participants were over the age of 18. This study was approved by our institution's IRB as a human subjects study.

5.4.2 Study Procedure

The study lasted two weeks in March 2022. Participants were randomly separated into one of three groups, each with differing levels of access and features within Chirp.

1. **Social (27 total participants)**: using the fully-featured version of Chirp as described above (Figure 5.1(a-e)). Participants were in an isolated study-only group. These participants only shared posts and interacted with users in this group.

2. **Individual (25 total participants)**: using only Chirp's profile page as a mood tracker. These participants had no access to the social timeline or other social features (Figure 5.1(a,d-f)).

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

Our three-group study design allows us to individually test our two research questions. We can test the overall impact of emotional self-disclosure in a semi-private online space on perceived social support, as described in RQ1, by comparing the Social and Control groups. We can also isolate the impact of the potential for reciprocal interactions (i.e. emoji reactions) described in RQ2 by comparing the Social and Individual groups.

Since Chirp is a relatively unknown system that is not broadly used outside of this study, we use a between-subjects study design. This allows us to avoid the influence of familiarity with the platform over time on our study conditions. A between-subjects study also allows us to collect data on all three groups over the *same period of time*, which prevents temporal influences on participant feelings of social support. As we are studying students, there are common outside factors that are likely to impact feelings of social support, for example, changes to university COVID policy, school breaks, or (as we encountered in this study) periods of mid-terms or finals.

Pre-Study Eligibility and Onboarding   Before the study period began, all participants completed 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 were additionally asked to install and join the correct versions of the Chirp app as part of the survey. They were asked not to begin posting in the app until the study start date and were instructed to post in the Chirp app at least once a day once the study began. Participants in the Social group were also informed 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 were sent 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 were asked to fill out the surveys to the best of their ability within two days of receiving the email request to complete the relevant survey.

We sent emails containing personalized survey links to participants at around 10 am local time every week for three weeks. All participants were contacted at the same times on the same days. Up to three additional reminder emails were sent over the following two days reminding participants to complete the surveys until they did so.

All three study surveys asked participants to report on their 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 included 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 our comparative measure
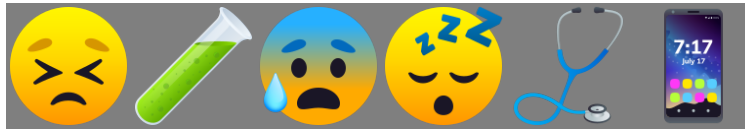
Figure 5.2: An example of a post made in Chirp. The first emoji shown is the "main" emoji, and all following emojis are the "story" emojis. This image was shown to participants as part of the final study survey, and participants were asked the following question: What do you think the author of the following 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?"

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 were also asked if they had gone to someone for support in the past week, and if so who and when. These surveys also asked participants in the Social and Individual groups whether they had shared their Chirp username with anyone else, and to explain their intended message from a provided recent post they had made in the Chirp app. Additionally, two of these provided posts were 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 were asked to describe what they thought the original poster meant to express in the given post. An example of one of these chosen posts can be seen in Figure 5.2.

The final survey additionally asked 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 were given a space for any final thoughts. In this final survey, Control group participants were also shown the same two selected posts as the other participants and asked to describe what they thought the original poster meant to express. See Appendix 6.3 for survey questions.

Compensation Participants were 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 were consistently engaging with the app, with the option to engage more often than we required.

Data Cleaning for Analysis Following data collection, we removed a total of 12 participants from our analysis. These 12 participants were removed for the following reasons: 7 participants were removed for failing to complete at least one survey, one participant in the Individual group failed to use the app at any time during the study, 3 participants had unusually low MSPSS scores outside the normal distribution computed using the interquartile range method (IQR), and one participant was removed as they had responded 2 to all questions of a 40-question Likert scale, and thus were likely answering at random. This left 65 total participants in the study, with 25, 22, and 18 participants in the final Social, Individual, and Control groups respectively.

5.5 Results: Impact of Chirp on Perceived SOcial Support

This study primarily asks two research questions:

RQ1) How do responses and reciprocal emotional self-disclosures from other students impact perceived social support among students who regularly emotionally self-disclose in private or semi-private online spaces?

RQ2) How does the addition of regular emotional disclosure in a semi-private online space impact perceived social support among undergraduate students?

Here we look to answer these questions through a comparison in OSSS scores between our three study groups. We then investigate potential confounding factors that may be have an influence the results. As this study investigates multiple participant groups each with measurements at three equally spaced time points, we use mixed repeated measures ANOVA.

Metrics for Social Support: Using OSSS Over MSPSS

The MSPSS is an older measure of social support, first developed in 1988 [191]. As a result, it primarily focuses on in-person interaction and social support, which is reflected in the questions it asks (See Appendix 6.2). As this study focuses on online interactions and specifically tries to exclude in-person interaction, we use a more recently developed measure, the Online Social Support Scale (OSSS) [126], to measure perceived social support from online sources (See Appendix 6.1).

We further note that OSSS is a particularly long questionnaire, with 40 individual items that sum to a single score. With 40 Likert scale questions in a row, a participant is likely to skip a response accidentally. Since the scale is summed, a missing response can significantly impact scores between weeks. To account for this, we normalize it in two ways to account for between-participant and question ordering biases.

First, for each participant, we find any questions they skipped across the three surveys (observed min: 0, max: 2). These questions were removed from the scale across all weeks for that participant. This ensures that for each participant, the set of measured and responded questions is consistent across all weeks of the study. However, this means the maximum OSSS scores *between* different participants are inconsistent. We account for this in the next step.

Second, we measure the *change* in each participant's OSSS over time. This allows us to fairly compare participants across different groups, despite the used OSSS scale being slightly different per participant due to skipped questions.

5.5.1 OSSS scores Decline, With the Exception of Social Users

As we measure perceived social support in this study using the OSSS, we first check if the Social group shows a positive change in OSSS over the course of the study when compared to the Social and Individual groups. By comparing OSSS changes between the Social and Control groups we can see the overall impact of regular emotional self-disclosure in a semi-private space among undergraduate students (RQ1), and by comparing OSSS changes between the Social and Individual groups we can see the impact of response and potential for reciprocation (RQ2). The results are not statistically

Figure 5.3: Mean change in OSSS scores over the course of the study. Error bars denote 95% confidence intervals. Note that while the Social group stays approximately constant, the Individual and Control groups show a decrease in OSSS over the course of the study. The Control group shows a faster and more pronounced decrease than the Individual group. Two weeks after the study start indicates the end of the study.

| (a) Esteem | (b) Companionship | (c) Instrumental | (d) Informational | (e) Key |
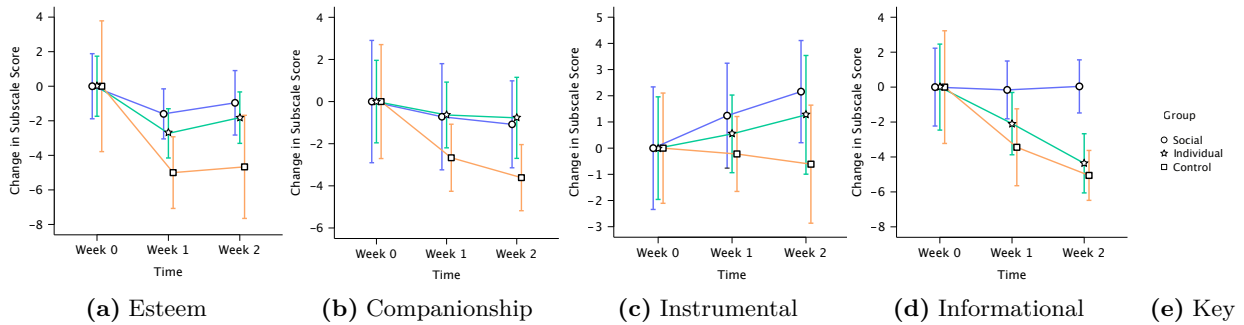
Figure 5.4: Mean change in OSSS subscale scores (Esteem, Companisonship, Instrumental, Informational) over the course of the study. Error bars denote 95% confidence intervals. Note that all subscales show similar trends to the full scale (Figure 5.3). In all cases, the Social group shows approximately constant or increasing scores. With the exception of the companionship subscale, where the Individual group matches Social, in all cases the Individual group shows less positive results than the Social group, and the Control group shows the most pronounced decrease in scores. Two weeks after the study start indicates the end of the study.

significant (2-way repeated measures ANOVA Time × Group interaction effect[11], $F(3.3, 102.3) = 1.3$, $p = 0.27$, $\eta_p^2 = 0.06$), indicating no provable difference between study groups. We do see a moderate effect size where participants in the Social group seem to have maintained a consistent level of perceived social support, while OSSS scores for the Individual and Control groups dropped (See Figure 5.3).

5.5.1.1 OSSS Subscales Are Consistent With Full Scale

As mentioned previously, the OSSS as a measure is built from four subscales, each corresponding to a different "type" of social support (i.e., esteem, companionship, informational, and instrumen-

---

[11] This test violates the assumption of sphericity and therefore in reported with Greenhouse-Geisser correction

tal) [126]. If the trend we see in the OSSS scores is a result of random variation, we might expect to see more variance in results among the subscales. Alternatively, if the pattern we see is only impacted by one specific subscale, then we would expect to see random variance among some subscales and a very strong relationship in one specific subscale.

As shown in Figure 5.4, we see a continued trend among all subscales of non-significance, yet with a consistent drop in scores in the Control group. The consistency of this same pattern across all subscales of the OSSS may increase the likelihood that these are true trends. However, again we do not find statistically significant differences when we perform a two-way repeated measures ANOVA for each case, for the interaction effect between Time and Group: esteem[11] ($F(3.4, 104.0) = 1.1$, $p = 0.36$), companionship[11] ($F(3.5, 108.6) = 0.5$, $p = 0.7$), informational[11] ($F(2.9, 91.0) = 2.2$, $p = 0.10$), and instrumental ($F(3.7, 115.6) = 0.5$, $p = 0.73$)

5.5.1.2 MSPSS Impact on OSSS

There is also the possibility that factors outside of the Chirp app had an impact on the results. To test this, we check for changes or differences in MSPSS scores between participant groups over the course of the study.

Since the MSPSS is a separate measure of (in-person) social support, we expect to observe one of two patterns if it has an impact on OSSS scores. 1) If OSSS scores reflect online conversations that participants are having with in-person friends then we would expect the MSPSS to have a similar pattern to what we see in OSSS; that is, that participants in the Social group have the highest scores or increase and those in the Control group have the lowest or decrease. 2) If OSSS scores reflect that participants with weak offline ties build stronger online support networks, as predicted by Self Determination Theory, then we would expect the MSPSS to have an opposite pattern to the OSSS; that is, participants in the Control group have the highest or increasing MSPSS scores and those in the Social group have the lowest or decreasing.

While we find a significant difference between the three groups, we do not see either of these patterns reflected (as shown in Figure 5.5). Furthermore, we do not see a change over time in MSPSS scores in any group, thus we would not expect a change in outside social support to have an impact on OSSS scores. A two-way repeated measures ANOVA shows no statistically significant interaction effect on MSPSS score between time and study group ($F(4, 116) = 0.29$, $p = 0.88$). Simple main effects analysis also shows that time did not have a significant effect on MSPSS ($F(2, 116) = 1.44$, $p = 0.24$), though the study group did ($F(2, 58) = 5.17$, $p = 0.01$).

5.5.2 Social Group Participants Posted More Consistently, but with similar Timing and Emotional Valence as Those in the Individual Group

As we determined that MSPSS scores are likely not the cause of the trends we see in OSSS scores, we next question whether differences in frequency, timing, or emotional valence (positivity/negativity) of Chirp posts between the Social and Individual groups had an impact on OSSS scores. In both groups, participants are asked to post at least once a day in the same format.

As suggested by Deters and Mehl, it is possible that the act of posting about oneself is enough for positive outcomes. Thus, we question whether participants in the Social group simply posted in the Chirp app more than those in the Individual group. We find that the Social and Individual groups
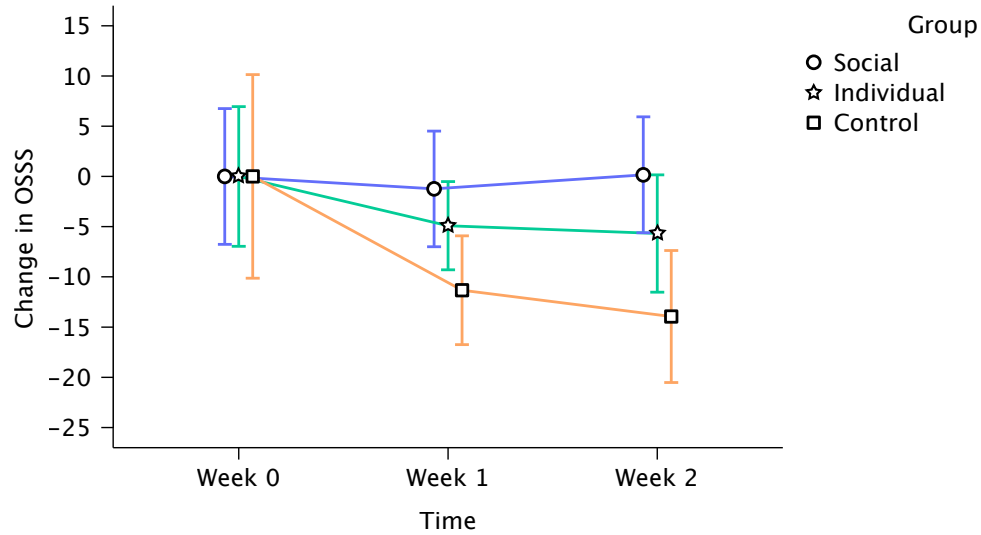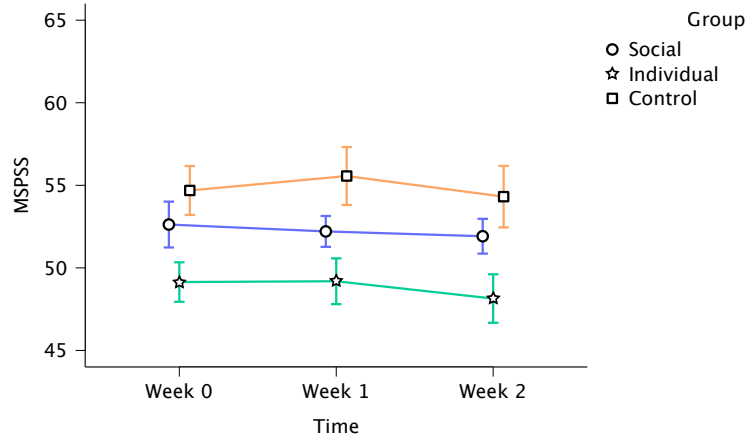
Figure 5.5: Mean MSPSS scores over the course of the study. Error bars denote 95% confidence interval. Note that each group stays at a consistent MSPSS over the course of the study. Two weeks after the study start indicates the end of the study.

of Chirp users posted with similar frequency and at similar times, which suggests that the results were unlikely to be impacted by one group posting significantly more often than the other. 356 posts were made in the Social group over the study period, with each participant posting at minimum 8 and at maximum 18 times throughout the 14 day study (mean: 14.1, median: 15). Individual group participants made a total of 290 posts throughout the study, with each participant posting a minimum of 2 and a maximum of 27 times throughout the study (mean: 13, median: 14). Users in both Social and Individual groups posted once per day a majority of the time (84.8% and 70.3% of the time, respectively). However users in the Individual group were twice as likely to not post on a given day than Social users— 22.1% and 10.4% of the time, respectively ($\chi^2(1, N = 702) = 20.6452, p < 0.001$). This may indicate that the presence of other users in the Social group motivated users to post more regularly. Most posts across both participant groups were made between 6–7 PM local time (Social: 24%, Individual: 23%), while the vast majority were made between 10AM and 9PM (Social: 79%, Individual: 83%). This emphasis on the 6–7 PM time period is likely because Chirp sent a daily push notification at 6:05 PM to those who had not posted yet that day.

Yang et al. suggest that the valence of shared posts may have an impact on reciprocation and support [184], leading us to the question of whether posts in the Social group were more negative than those in the Individual group. Here we mark each of the nine "main emojis" in Chirp as either positive (satisfied, happy, excited), negative (sad, scared, angry, stressed), or neutral (tired, bored)[12] and compare the frequency of positive or negative emoji choices in the Social and Individual groups. We find the Social and Individual groups to have similar levels of positive valence main emojis over the study period (45.5% of Social posts and 49.7% of Individual posts). We also find the Social and Individual groups to have similar levels of negative valence main emojis (27.5% of Social posts and 26.5% of Individual posts). This suggests that differences in post valence did not lead to the differences in perceived social support between the Social and Individual groups.

---

[12]emoji labels as described in the Chirp app, shown in Figure 5.1d

5.5.2.1 Interactions Beyond Posting in the Social Group

As described previously, the experience of a participant in the Individual group was similar to the Social group, with the removal of social features as the only difference (see Figure 5.1). We question whether simply posting in a social space versus non-social space accounts for the differences in OSSS between these groups or if interaction from other users, represented in this study by reactions left on posts, had an impact on OSSS (RQ2).

Most participants in the Social group (72%) did interact with another post at some point in the study. Excluding reactions that were later removed, 18 of 25 participants in the Social group reacted to at least one post during the study, with 16.7 reactions on average (total: 320, min: 1, max: 121, median: 5.5). Though not every participant in the Social group reacted to another post, every user received reactions on at least one of their own posts (number of reactions min: 4, max: 28, mean: 12.6). This indicates that there was interaction and reciprocation between participants in the Social group, even though it was not incentivized in the study.

Based on previous work that suggests that interaction and reciprocation leads to more positive outcomes [17, 184], we would expect that reacts as visible indicators of support would lead to higher levels of perceived social support and that more reacts would lead to higher levels of perceived social support. However, we see no correlation between the number of reactions to a participant's posts and OSSS at the end of the study (*Pearson's r*$=0.16$, $p=0.46$).

5.5.3 Participants Were able to Connect Through Posting in Chirp

To better understand whether and how Social group participants built relationships through Chirp, in the final survey we asked Social users whether there was "any other user in Chirp that you felt particularly positive about or felt like you connected with?". Two participants pointed out specific other users by username that they felt close with (P170 mentioned P189, and P208 mentioned P201). Other participants expressed that focusing on posts by specific users was less impactful than general trends in posts. For example, P145 mentioned that seeing other posts helped them feel less isolated:

> I honestly did not pay much attention to usernames so I can't say that a specific user connected with me but there were a few posts that I remember resonating with and it helped me feel a little less isolated knowing that other people aren't constantly joyful and happy 24/7. - P145 (Social)

Other participants noted that common shared experiences among Social group users led to reassurance and supportive responses. A few participants noted a chemistry midterm that seemed to be a common thread among posts:

> Whenever anyone posted the chemistry test tube with green liquid inside, I knew they were talking about the upcoming chemistry exams, so I felt a bit reassured knowing that there were some individuals out there stressed about the exam as well. - P197 (Social)

While not a shared experience in the moment, common experiences such as relationship struggles also elicited support. One participant noted an effort to support another user who they believe posted about a breakup:

> I remember specifically one post with a crying emoji and a heartbroken/teddy bear emoji, which instantly read as a breakup or a fight in a relationship. It had a lot of sad reacts from people empathizing with the original poster. - P147 (Social)

Overall, survey responses seemed to indicate that users were able to and actively did build connections and attempt to provide support over the Chirp app.

5.5.4 Participants Understood Emoji Posts

As the Chirp platform uses emoji-only communication, there is a possibility that participants were unable to interpret others' posts. If this were the case, then participants may assume that their own posts are not understandable and the "social" aspect of the app would be irrelevant. Given evidence that users felt supported in the Social group, we expect there to be a high level of understanding or perceived understanding on the part of the participants.

5.5.4.1 Participants Were Divided on Perceived Interpretability of Emoji Posts

We first find common themes of participants' perceptions of their understanding of emojis through a thematic analysis conducted by the first author. The analysis focused on participants' perceptions of emoji communication and understanding, thus we used open coding and focused on responses to the following questions from the survey (as shown in Appendix 6.3):

- In general, how did you feel when using emoji-only communication?

- Were there any specific times where you felt you had a very good understanding of what a post on Chirp was trying to convey? Can you give an example and how you interpreted it?

- Were there any specific times where you felt you had a very bad understanding of what a post on Chirp was trying to convey? Can you give an example and your best guess at interpreting it?

- This is the end of the survey, is there anything else you would like to tell us?

Some example codes include: "deciding which emojis to choose is a creative outlet", "unsure if others understood intended meaning", "knowledge of popular courses informs post interpretation". These codes were then consolidated into broader themes, including "use as a space for self-reflection", "impact of anonymity on comfort", "ability to interpret given context of the cohort", and "emotional response to emoji-only restriction". Here we focus on the theme of "ability to interpret given context of the cohort". Overall we find that while some types of posts felt very difficult to interpret, there were many posts with a clear narrative or outside context that made them highly interpretable.

As mentioned previously, Social group participants were asked if "there [were] any specific times where [they] felt [they] had a very *bad* understanding of what a post on Chirp was trying to convey". Most participants pointed out that certain types of posts were hard to interpret, or "seemed random". As one participant shared,

> when individuals posted emojis that seemingly had no correlation, it was difficult to interpret what the user meant. For example, if someone posted an emoji of them being scared with random emojis of a fruit, a sun, a dolphin, and a chemistry test tube, I would probably misinterpret why they are scared. - P189 (Social)

However, many participants also noted that there were patterns and common cases that they felt were easy to interpret when asked if "there [were] any specific times where [they] felt [they] had a very *good* understanding of what a post on Chirp was trying to convey". One participant pointed out the importance of time to their interpretation,

> usually on Sundays most people posted the tired emoji with work related emojis so I usually figure people were tired and had lots of work after doing fun things the night before.
> - P208 (Social)

5.5.4.2 Participant Interpretations of Emoji Posts Were Accurate

With the inconsistent nature of participant perceptions of their ability to interpret posts, we investigate how consistently and accurately participants are actually able to interpret posts. We perform a content analysis on one post with a known "ground truth" meaning to see actual interpretation accuracy among participants. We further used content analysis to identify themes across all participants' interpretations of that same post, described in more detail below. We find that interpretations were often accurate to the author's intent, but even when not accurate, interpretations overlapped significantly between users.

As mentioned previously, participants were shown a Chirp post from the day before the survey was distributed, which was posted by another participant in the Social group (shown in Figure 5.2). We collected both the ground truth intent of what the original poster meant to convey in the post, as well as interpretations from all other participants. The original poster described their post,

> I was stressed because I had a chem exam, hadn't been getting much sleep, and had to do a CPR certification module - P189 (Social)

From this post, we extracted 5 topics that encompass the intended meaning of the post:

- Stress
- Exam
- Lack of Sleep
- Chemistry
- CPR

After these topics were identified in the ground truth post, a codebook was created to code each topic across all participant interpretations of this post. A subset of participant interpretations was used to identify synonymous concepts and terms, which were used to update the codebook. Ending in the following, more generalized topics:

- Stress/Anxiety/Worry
- Exam/Test/Midterm
- Lack of Sleep/Tired
- Chemistry
- CPR

All participant interpretations were coded based on these topics. We then similarly identified and coded concepts from every participant's response to find common concepts between participants that were not included in the ground truth.

| | Described by Author | | | | | Other Common Topics | | | | | |
	Stressed / Anxious / Worried	Lack of Sleep / Tired	Chemistry	Exam	CPR	Pre-Med / Med. Student	STEM	Lab	Phone Use	Doctor	Sick
Social	70.83%	41.67%	41.67%	29.17%	0%	41.67%	20.83%	16.67%	12.50%	12.50%	4.17%
Individual	77.27%	54.55%	31.82%	31.82%	0%	31.82%	22.73%	18.18%	27.27%	18.18%	18.18%
Control	22.22%	55.56%	11.11%	5.56%	0%	22.22%	5.56%	5.56%	11.11%	27.78%	44.44%

Topic Rate of Prediction per Group for Example Post

Table 5.1: Participants who had been using Chirp in any form (in either experimental group) were able to successfully identify that the post's author was communicating stress and a chemistry exam much more accurately than the Control group. However no participants were able to correctly interpret that the author had a CPR module. Despite it not being described by the author, many participants also said that the author was a pre-med or medical student, and nearly half of the Control group thought the author was sick.

As shown in Table 5.1, participant interpretations from the Social group are at least 29% accurate across all concepts, with the exception of CPR, which no participants mentioned. Additionally, even for concepts not included in the ground truth, participants are consistent between themselves. This implies that there is a broader shared understanding of a post within the group, even if it is not true to the original poster's intent.

We, therefore, assume that the participants in the Social group are able to understand each other's posts and that the context provided within our chosen cohort is sufficient for successful interpretation, as is consistent with previous work [94]. This further implies that participants in the Social group were able to understand each other and connect in this space.

5.6   Methods: Emoji Understanding and Interpretation

In this section we describe our study process. We first drafted and piloted a survey, fine tuning it for participant understanding and reducing survey fatigue. This survey was then distributed over online and in-person channels, receiving 644 total responses. We cleaned the data, removing invalid and duplicate responses, resulting in 163 valid responses (520 emoji post interpretations). We additionally report on our participant population. The final distributed survey can be found in Appendix 6.4. This study was approved by our institution's IRB as a human subjects study.

Our final distributed survey includes four displayed emoji posts, alongside demographic information about the poster (either accurate or inaccurate). Participants are asked to interpret the meaning of each post, as well as rate their confidence in their interpretation both before and after seeing "another interpretation of the same post" (the ground truth).

5.6.1   Pilot Testing

We conducted a pilot test to inform the design of this survey. The pilot test was used to assess an appropriate survey length that maximizes response collection while minimizing survey fatigue and to ensure that participants would notice provided information about poster demographic information in the emoji post.

All pilot sessions were conducted virtually over video call, though testers used different devices for testing (i.e., laptop, desktop, iPhone, Android). Upon entering the call, testers were sent a link to the online survey. Testers were asked to share their screen and were prompted to use the think aloud method as they completed the survey. The researchers did not give context or answer questions during the test. At the end of each pilot session, researchers asked the participant several questions

related to their understanding of the survey questions, the length of the survey, and their overall impression and thoughts. The survey was then adjusted before the next pilot based on researcher observations and tester feedback.

Based on these pilot sessions, the following changes were made: demographic information of the original poster was moved directly above the emoji post, a progress bar was added to the survey, the wording of the request for an interpretation was adjusted, confidence rating was switched from a numerical scale to a Likert scale, a question about why interpreter confidence did or did not change after seeing another interpretation was added, and survey length was limited to four emoji posts.

5.6.2 Emoji Post Structure and Selection

In this survey, we investigated how accurate participants are when interpreting a structured emoji string (or "post"), in order to better understand user interpretations of emoji-only strings. Emoji posts were drawn from an existing emoji-only mood tracking app. Each displayed post starts with one emoji that describes the overall "mood" represented in the post (in response to the prompt "How are you feeling"), followed by exactly five emojis describing the reason for that mood (in response to a follow-up prompt "Want to share why?"). Each post included in this survey additionally has a written ground truth description of the meaning of the post, written by the original poster. We note that all original posters in this study are young adult students at [anon institution]. We purposefully chose a homogeneous population to control for actual demographic or cultural differences between posters that could act as a confounding variable.

(a) The selected "happy" post

(b) The selected "angry" post

(c) The selected "satisfied" post

(d) The selected "tired" post
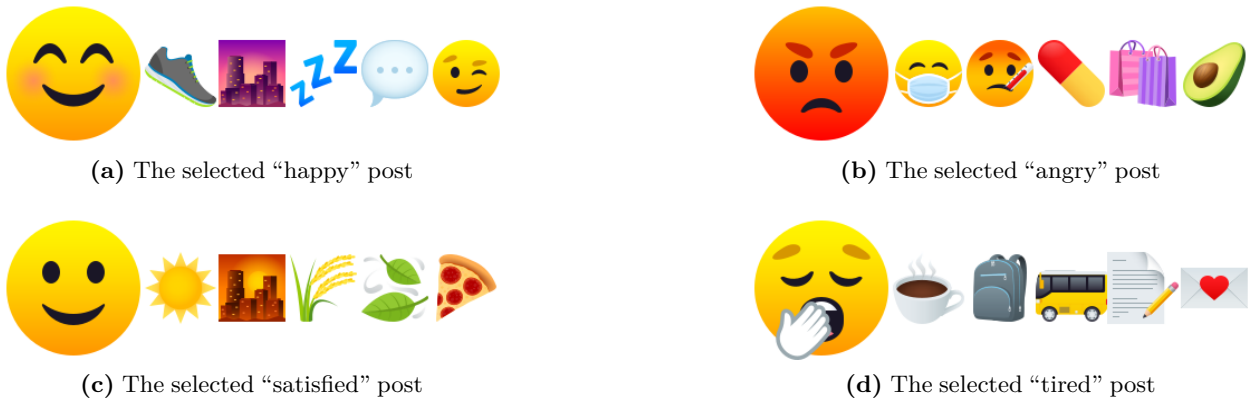
Figure 5.6: The four selected emoji posts. (a) is the selected positive post, (b) the negative post, (c) the ambiguously positive, and (d) ambiguous meaning.

We selected four emoji posts for use in this survey, selected randomly through stratified sampling. Posts were categorized into four groups based on the main emoji of the post: definitively positive, definitively negative, ambiguous positivity, and ambiguous meaning. One post was randomly selected from each of these groups. Valence and ambiguity of the main emojis were determined through analysis of perceptions in a survey in a previous study. The final selected emoji posts can be seen in Figure 5.6.

5.6.3 Study Procedure

Within the survey, participants were first asked to complete a consent form and a set of eligibility questions ensuring that all participants were over the age of 18, residing in the US, and had not participated in a previous study using the emoji-based mood tracking application from which the posts were selected. The survey then included a set of demographic questions around their age, gender, student status, work status and area, and time residing in the US. Participants also reported their familiarity with emoji, as well as their mood at the time.

Finally, participants were presented with all four emoji posts, each with the following context:

> The following post was made on the [ANONYMOUS] app. The first of these emojis is a response to the prompt "How are you feeling?", and the following emojis are responding to the prompt "Want to share why?". This post was written by a [male/female] [young adult/adult] [who is currently a student/who is not currently a student].

All participants saw all four emoji posts in a random order, each with either correct, incorrect, or blank demographic descriptions of the original poster. For each emoji post, participants were asked "What is your best guess of what the author of the post meant to convey?". Following this interpretation, participants were shown the ground truth interpretation and told that it was an interpretation provided by "another user". Participants were asked to rate their confidence in their interpretation after seeing this alternate interpretation, as well as to describe why. Participants were compensated with the opportunity to enter a raffle for a $25 gift card with a 1/10 chance to win.

5.6.4 Participants and Recruitment

This survey was distributed over 3 and a half weeks in March 2023. To sample a diverse group of users, participants were recruited through both online and in-person platforms. X (formerly Twitter) and Mastodon advertisements were posted through lab and personal accounts and subsequently shared or forwarded by other users. Recruitment posts were also shared over survey completion Reddit subreddits, specifically /r/SurveyExchange, /r/SampleSize, /r/PaidStudies, and /r/SurveyCircle. Advertising flyers were distributed in public spaces in the city of [redacted for anonymity]. Since this survey is likely to be impacted by technical knowledge or internet use, we wanted to allow for non-technical users to find and respond to the survey. Thus flyers were distributed in high-traffic areas that might attract a broad audience (e.g., restaurants, cafes, train stations, bus stops).

This survey received 644 responses from both online and in-person sources. Of these, the vast majority were responses from Reddit ($N = 448$), followed by posts on Mastodon or X ($N = 130$), from in-person fliers ($N = 13$), and from other or unknown sources ($N = 53$). However, many of these responses were found to be invalid (See Section 5.6.5). After cleaning for these invalid responses, we perform our analysis on the remaining 163 responses. Our data cleaning methods are described in subsection 5.6.5.

Our final cleaned dataset contains 163 responses (49% from Reddit, 39% from X and Mastodon, 7% from fliers). 36% of participants were female, 56% male, and 9% non-binary. Participant age averaged at 27.3 (*median* $= 25$, *sd* $= 8.2$), and participants had lived in the US for an average of 25.8 years (*median* $= 24$, *sd* $= 9.6$). 58% of participants reported that they were students.

5.6.5 Data Cleaning

A prevalent problem in responses gathered through online platforms is that the surveys may not be completed in good faith. To ensure our analyses are based on good quality data, or good-faith responses, we filter out bad-faith responses in the following ways.

First, of the 644 responses, we removed 287 participants who did not fill out or did not meet the eligibility criteria. Our eligibility criteria stated that participants must be at least 18 years of age, residing the US, and have never participated in a study using the [ANON] system. We also removed 3 participants that reported residing in the US longer than they had been alive. In addition, we used the short answer responses of the survey to filter for bad-faith responses or bots. We removed 13 participants who responded to the short answer questions with a single letter response.

Next, we removed participants who provided the exact same short response multiple times within the same survey. This is to prevent bad-faith users who were copying and pasting the answers multiple times without critically answering the question.

Finally, we removed survey entries with duplicated responses seen in *other* survey entries. Because surveys were anonymous, the same participant could respond to the survey multiple times. Thus, we check for duplicate responses across all 644 survey responses from participants. If survey responses were duplicated, we take the first timestamped instance of the duplicated survey and removed all subsequent survey responses.

To ensure quality of data, removal of duplicates was done manually. Two of the paper authors read through every survey response. The final "cleaned" data was then read by a third author who verified that all remaining survey response were unique. This resulted in a total number of 163 good-faith responses from participants.

5.7 Results: Emoji Understanding and Interpretation

In this section we evaluate how demographics affect interpretation of emoji posts. We question how the demographics of the *interpreter* (i.e. the participant) and how the demographics of the *poster* impact accuracy of interpretation. We additionally investigate how outside factors including emoji familiarity, current mood, and confidence affect accuracy of interpretation.

5.7.1 Defining Interpretation Accuracy

As both our ground truth interpretations provided by the original poster and our participant interpretations are provided as short answer responses, we use content analysis to identify how many themes from the ground truth are accurately identified in each interpretation.

For each emoji post, two authors identified major themes present in the description. Using these themes, a code book was generated for each post, outlining what descriptors fit within the theme and which did not. Using a subset of survey responses, the code book was iterated and refined until both authors felt satisfied with its specificity. The final code book can be found in Appendix 6.5.

All participant responses were then hand-coded based on the code book by one author and a subset of responses were coded by two other authors. Thus all responses were coded by at least two authors ($irr : 0.92 - 0.97$). We note that based on our coding scheme, only identification of intended

themes by the poster affect accuracy; the addition of unrelated themes in an interpretation do not affect accuracy scores.

5.7.2 Interpreter Demographics have no Impact on Accuracy, but Do Affect Confidence

In this chapter we question whether interpreter demographics impact interpretation accuracy or interpreter confidence. We find that participant demographics have no effect on interpretation accuracy, but that many do have an impact on confidence. In this study, we collected gender, student status, work status, and age which we use to define demographic status. We see that none of these factors significantly impact accuracy: gender (F$(2, 517) = 2.9$, p $= 0.058$), student status (F$(2, 518) = 0.33$, p $= 0.57$), work status (F$(2, 517) = 0.70$, p $= 0.5$), and age (Pearson's $r = 0.04, p = 0.3$). However, we do find that, when measured individually, all of these factors have an impact on interpreters confidence, with non-binary participants being significantly less confident $(F(2,517) = 6.16, p < 0.01)$[13], as well as younger (Pearson's $r = -0.25, p < 0.001$), student $(F(2,518) = 25, p < 0.001)$ interpreters feeling more confident, and full-time working interpreters feeling less confident $(F(2,517) = 11.6, p < 0.001)$. However, we note that younger participants are also more likely to be students and less likely to be working full time. Thus when we treat age as a covariate, we find that student status does not independently impact confidence, though **gender**, **age**, and **work status** all continue to show a significant impact on confidence. Overall, while confidence levels varied among different demographics, we find that demographics, ultimately, had no effect on the accuracy of interpretation.

5.7.3 Confidence, Emoji Familiarity, and Emotional State Do Not Impact Accuracy

As we do not find any relationship between demographic factors and emoji interpretation accuracy, we question how other factors might have led to different interpretations. As we see that confidence between interpretations does vary, even for the same interpreter looking at different posts, we first investigate whether confidence correlates with interpretation accuracy. However, we find no relationship between interpreter confidence and accuracy (Pearson's $r = 0.02, p = 0.71$). Similarly, we would intuitively assume that interpreters more familiar with emojis should be more accurate, but we find no relationship between emoji usage frequency and accuracy (Pearson's $r = 0.06, p = 0.21$). We do find that more frequent emoji usage correlates with confidence (Pearson's $r = 0.18, p < 0.001$). Finally, we find that the current emotional state of the interpreter also does not influence interpretation accuracy $(F(8,511) = 1.50, p = 0.16)$.

5.7.4 Overall Interpretations Were Highly Accurate

Looking at all interpretations, we see that participants on average understood about 46% of topics intended by the original posters ($mean\,accuracy = 0.46, sd = 0.27$). This holds true across all demographic categories we measured (age group, gender, student status, work status). Participant mood at the time of interpretation and self-report emoji familiarity (with the exception of "very unfamiliar" ($n = 4$)) also had no effect on interpretation accuracy. We note that this 46% is in indication

---

[13]please note that there were significantly fewer non-binary participants than men or women which may have influenced these results

of how much each interpreter understood of each post, meaning that about half of the topics were understandable to most users in any given post.

We note that in this study we were purposefully quite restrictive in our coding scheme, so 46% accuracy may be lower than the effective level of understanding. For example, in the "happy" post (Figure 5.6a), the original poster described that they "went to the gym", in our code book we marked interpretations as correct when they mentioned going to the gym or doing exercise, but not when describing specific types of exercise such as running or playing sports.

5.8 Discussion

Numerous studies have shown evidence of socially supportive actions over online platforms. These have included semi-public [66], semi-private [184], private [81, 184], and anonymous [6] online communities. Previous work has also shown evidence of communication and support through emoji and other non-textual communication [90, 94, 109]. In Chirp, we leverage these design choices in a sandbox space but fail to find a significant relationship between regular self-disclosure in a semi-private space among students and perceived social support (RQ1), nor do we find a significant difference between disclosure in a space with potential reciprocation versus one without (RQ2). Despite this, qualitative data in this study shows that participants in the Social group were able to build connections in the space and expressed that they felt they were able to relate to and felt comforted by posts made by other participants. This suggests that emotional self-disclosure in Chirp's semi-private social space did facilitate supportive behavior as has been seen in previous studies.

Here we discuss these findings and the potential reasons behind these discrepancies.

5.8.1 Social Spaces Encourage Activity Which May Lead to Disclosure and Support

5.8.2 Participant Intent May Impact Outcomes

Qualitative responses from participants in this study show a diverse trend. While some participants felt the emoji posts were too random to interpret, others recalled specific instances and users with whom they either built connections or gave or received support. We propose that this may reflect user intent while using the system.

Previous work has pointed out intent as an important factor in receiving support. Yang et al. identify negative self-disclosure as support-seeking behavior, which then leads to support [184]. The Interpersonal-Connection-Behaviors (ICB) Framework similarly suggests that connection-promoting (or "active") behaviors lead to more positive outcomes [37]. As this study incentivizes daily posts containing emotional self-disclosure, the act of self-disclosing—whether negative or positive—is no longer on its own indicative of any specific intent or support-seeking behavior. This may lead to a separation between participants who posted in the app simply to fulfill the requirements of the study and those who were also genuinely interested in using the app—and therefore engaging with the intent to connect. This would align with previous work, and if this were the case, it could explain the variations in participant experiences observed in this study.

It would be prudent for future work to concurrently study both naturalistic user experiences as well as outcomes from direct experimental interventions like the study described here in order to

disentangle this factor.

5.8.3 Differentiating Supportive Actions and Perceived Social Support

It is notable that previous work has primarily looked at the existence and frequency of supportive responses rather than the metric of perceived social support as in this study (e.g., [7, 17, 90, 184]). While these two metrics are closely related, they are different measures. There is a possibility that while online interactions do lead to socially supportive behavior, online platforms do not effectively facilitate perceived social support. This would align with Media Richness Theory, which proposes that the richer platforms, or platforms that provide more information to the user (e.g., tone, body language, facial expression), lead to more effective communication and, therefore, better outcomes [47]. Supporting this suggestion, recent work has found that the addition of richer information in messaging interfaces, such as more informative typing indicators, can increase feelings of connection [85].

However, both in previous work and the current study, participant responses in interviews and surveys indicate that users do feel socially supported by these online conversations (e.g., [5, 187]). This seems to support the idea that users can both provide and perceive support over online platforms.

We suggest that future studies look not only at metrics of actual social support but also include measurements of perceived social support. Particularly given evidence that received social support has less of an impact on mental health outcomes than perceived social support [138].

5.8.4 Ability to Relate May Be More Important than Direct Response

When discussing connection and support, participants in this study point out specific posts and users with whom they felt they were able to connect or empathize. Participants also point out cases where they or others used reactions in an attempt to provide support to others. However, participants make no mention of feeling supported based on reactions that they received. This may imply that responses made on posts are less important than feelings of similarity or empathy with independent posts made by other users.

This would align with previous work by Deters and Mehl, where the authors found that when participants were asked to post more frequently on Facebook, the number of received likes or comments on posts did not significantly impact outcomes [52]. Deters and Mehl propose that the act of disclosure with a target audience in mind may act as a reminder of existing bonds, helping users maintain a positive emotional state without social interaction [52]. However, unlike the Facebook users studied by Deters and Mehl, participants in the current study are pseudo-anonymous. While participants were aware users visible to them in Chirp were all in a similar cohort (i.e., first-year students at the same institution), they were unaware of who the other users in the app were (when asked, only four Social group participants said they had shared their username with anyone in or out of the study group). This lack of previous connection means that disclosure in Chirp is not made with existing bonds in mind. Yet participants describe feeling supported based on posts made by other users in the app.

This may indicate that support is found in feelings of similarity and empathy between users. In their study of cancer support communities, Yang et al. describe that supportive users showed support through reciprocity to negative self-disclosures by the support-seeking user [184]. Viewed from a different perspective, this supportive reciprocal disclosure also provides an opportunity for the support-seeking user to build a feeling of similarity and empathy with the supportive user.

5.8.5 Mood Tracking Context Provides Enough Structure for Understandable Emojis

Previous work around emoji-only communication has emphasized the need for context in understanding emoji strings [94]. In Chirp, we provide three forms of context: the mood posting framework, posting time (temporal), and knowledge that all posters are first year undergraduates at the same institution (demographic). We find that within this context, participants are able to interpret emoji posts quite consistently— correctly identifying the three (of 5) most identifiable concepts at least 41% of the time. Even when removing temporal and demographic context and providing potentially false demographic information about the poster, we still find relatively high interpretation accuracy across all users, regardless of their familiarity with emojis or confidence in their interpretations. This aligns with our findings among Chirp users, where participants showed mixed confidence in their ability to interpret emojis, but their actual interpretation ability was relatively high. Thus, we are confident that even the lightweight framework provided by the mood past setup is enough context for accurate interpretations.

Previous work has noted that when interpreting long strings of emojis, interpreters have a tendency to attribute literal meaning [3, 41]. Less familiar users may also be prone to attach more literal meaning to emojis, as opposed to the more abstract meanings we sometimes see [41]. This might explain why knowledge about the poster's demographics did not affect interpretation accuracy. If all interpreters were simply interpreting the literal meaning of the provided emojis, then the more nuanced differences in emoji use between different demographic groups wouldn't be taken into consideration in their interpretations.

However, interpreters using literal meanings does not explain why their interpretations were highly accurate. We speculate that, since the original posters of these emoji posts were writing them for a semi-public audience, those posters may have used emojis in a more literal fashion in order to purposefully make them more understandable. This departure from the usual, more nuanced, meanings behind emoji in some users' posts could lead to more consistent interpretations, as the shift to literal meaning on the part of the posters may coincide with the shift to literal meanings among the interpreters. This raises questions on the nature of the posting space and how it can affect the intent behind emoji. It would be interesting to investigate whether posts made in spaces with a broad intended audience do indeed use emojis in a more literal way than in other contexts.

Chapter 6

Conclusion and Future Work

This dissertation explores undergraduate student online communication patterns, investigating how individual design choices in private messaging can be integrated into social media spaces to increase perceived social support among undergraduate users. In Chapter 3 we presented findings from a survey of undergraduate students about their messaging habits before and during university COVID-19 lock-downs. This survey focused specifically on the relationship between different online communication spaces and user perceived social support. In Chapter 4 we discussed Sochiatrist, a unique data-extraction platform that collects and anonymizes private messaging data. Using data from Brown undergraduate students, we investigated the similarity between displayed sentiment in their private messages and ground truth measures of those students' emotional state. In Chapter 5 we presented Chirp, an anonymous, emoji-based sandbox platform used to investigate individual design choices in social media spaces. We validated the viability of emoji for communication in this context, and then discussed how the addition of a social space affected messaging patterns. Finally, in Chapter 5 we used Chirp to evaluate how and whether participants were able to build connections over the low-bandwidth space of Chirp.

Overall, this dissertation aims to provide actionable suggestions to improve user outcomes when using online social media platforms. Each of the described projects identify and evaluate different aspects of private online communication spaces and how they relate to supportive outcomes. In Chapter 5, we combined all of these different aspects, finding positive indicators in participant responses. Participants expressed that they were able to relate to other users in their communication spaces and that they felt supported. However, the study described in Chapter 5 only exists in a lab setting. Thus it would be prudent to additionally investigate naturalistic users.

Going forward, we can apply the lessons learned from this dissertation to other, existing systems. While Chirp may act as a proof of concept for the ideas explored here, student online communication will not be confined to one tool or system. Thus, we can leverage our findings around integrating new features into public communication spaces in platforms that students are already using.

Furthermore, in this dissertation we focus specifically on design choices that private messaging. However, students communicate in a wide variety of ways that are not reflected in private messaging. For example, video calls, voice calls, or online forum spaces. While Chirp and Sochiatrist provide an example of how private messaging can be directly investigated, future work should look into other understudied spaces.

While this dissertation studies student outcomes, struggles with mental health are not a student-

specific issue. Future work can use systems like Sochiatrist or Chirp to examine how these findings apply to other populations.

The systems presented in this dissertation focus specifically on social support as a positive mental health outcome, however there are many other factors that contribute to our mental health. Chirp and Sochiatrist are general systems that can be used much more broadly, to investigate other interaction types and outcomes beyond percieved social support. Further work can leverage these systems to investigate the impacts of online communication on other mental health outcomes and on

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] Dmitriy Agranovskiy and Alina Avilova. Emojis as a language of their own. *Technology and the Media Environment of the Information Society*, 2021.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

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

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

[11] 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.

[12] 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.

[13] 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.

[14] 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].

[15] 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.

[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, Yoon Hyung Choi, Victoria Schwanda Sosik, Dan Cosley, and Janis Whitlock. Social sharing of emotions on facebook: Channel differences, satisfaction, and replies. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*, pages 154–164, 2015.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] Jeremy Burge. 77% of 56-64 year olds use emojis on messenger, Nov 2017.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

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

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

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

[33] Yihua Chen, Xingchen Yang, Hannah Howman, and Ruth Filik. Individual differences in emoji comprehension: Gender, age, and culture. *Plos one*, 19(2):e0297379, 2024.

[34] Zhenpeng Chen, Xuan Lu, Wei Ai, Huoran Li, Qiaozhu Mei, and Xuanzhe Liu. Through a gender lens: Learning usage patterns of emojis from large-scale android users. In *Proceedings of the 2018 world wide web conference*, pages 763–772, 2018.

[35] 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.

[36] Sherman Chui. A comparative study of the interpretations of emojis in between us and chinese users. *International Journal of Literature and Arts*, 12(2):108–118, 2020.

[37] 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.

[38] 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.

[39] 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.

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

[41] Neil Cohn, Jan Engelen, and Joost Schilperoord. The grammar of emoji? constraints on communicative pictorial sequencing. *Cognitive research: principles and implications*, 4:1–18, 2019.

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

[43] 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.

[44] 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.

[45] 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.

[46] 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.

[47] Richard L Daft and Robert H Lengel. Organizational information requirements, media richness and structural design. *Management science*, 32(5):554–571, 1986.

[48] 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.

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

[50] 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.

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

[52] 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.

[53] 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.

[54] 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.

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

[56] 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`.

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

[58] 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.

[59] 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.

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

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

[62] 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.

[63] 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.

[64] Jose A Gallud, Habib M Fardoun, Fernando Andres, and Nehme Safa. A study on how older people use emojis. In *Proceedings of the XIX international conference on human computer interaction*, pages 1–4, 2018.

[65] 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.

[66] 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.

[67] 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.

[68] 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.

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

[70] 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.

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

[72] 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.

[73] 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.

[74] Sharath Chandra Guntuku, Mingyang Li, Louis Tay, and Lyle H Ungar. Studying cultural differences in emoji usage across the east and the west. In *Proceedings of the international AAAI conference on web and social media*, volume 13, pages 226–235, 2019.

[75] 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.

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

[77] 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.

[78] 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.

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

[80] Susan C Herring and Ashley R Dainas. Receiver interpretations of emoji functions: A gender perspective. In *Proceedings of the 1st International Workshop on Emoji Understanding and Applications in Social Media (Emoji2018). Stanford, CA*, 2018.

[81] 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.

[82] 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.

[83] 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.

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

[85] Zainab Iftikhar, Yumeng Ma, and Jeff Huang. "together but not together": Evaluating typing indicators for interaction-rich communication. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2023.

[86] 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.

[87] 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.

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

[89] 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.

[90] 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.

[91] Lara L Jones, Lee H Wurm, Gregory A Norville, and Kate L Mullins. Sex differences in emoji use, familiarity, and valence. *Computers in Human Behavior*, 108:106305, 2020.

[92] 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.

[93] 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.

[94] 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.

[95] 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.

[96] 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.

[97] 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.

[98] 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.

[99] 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.

[100] Andrey Krekhov, Katharina Emmerich, Johannes Fuchs, and Jens Harald Krueger. Interpolating happiness: understanding the intensity gradations of face emojis across cultures. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–17, 2022.

[101] 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.

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

[103] 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.

[104] 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.

[105] 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.

[106] 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.

[107] 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.

[108] 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.

[109] 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.

[110] 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.

[111] 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.

[112] 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.

[113] 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.

[114] 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.

[115] Talie Massachi, John Roy, Lauren Choi, Gabriela Hoefer, Wallace Shaun, and Jeff Huang. Chirp: The impact of private online self-disclosure on perceived social support. *Proceedings of the ACM on Human-Computer Interaction*, (CSCW), 2024.

[116] 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.

[117] 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.

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

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

[120] Hannah Miller Hillberg, Zachary Levonian, Daniel Kluver, Loren Terveen, and Brent Hecht. What i see is what you don't get: the effects of (not) seeing emoji rendering differences across platforms. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):1–24, 2018.

[121] 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.

[122] 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.

[123] 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.

[124] 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.

[125] 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.

[126] 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.

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

[128] 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.

[129] 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.

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

[131] 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.

[132] 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.

[133] 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.

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

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

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

[137] 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.

[138] Gabriele Prati and Luca Pietrantoni. The relation of perceived and received social support to mental health among first responders: a meta-analytic review. *Journal of Community Psychology*, 38(3):403–417, 2010.

[139] 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.

[140] 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.

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

[142] 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.

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

[144] 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.

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

[146] 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.

[147] 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.

[148] 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.

[149] 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.

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

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

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

[153] 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.

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

[155] 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.

[156] 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.

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

[158] 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.

[159] 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.

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

[161] 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.

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

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

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

[165] Garreth W Tigwell and David R Flatla. Oh that's what you meant! reducing emoji misunderstanding. In *Proceedings of the 18th international conference on human-computer interaction with mobile devices and services adjunct*, pages 859–866, 2016.

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

[167] 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.

[168] 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.

[169] 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.

[170] 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.

[171] 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.

[172] Levi Van Dam, Sianne Rietstra, Eva Van der Drift, Geert Jan JM Stams, Rob Van der Mei, Maria Mahfoud, Arne Popma, Eric Schlossberg, Alex Pentland, and Todd G Reid. Can an emoji a day keep the doctor away? an explorative mixed-methods feasibility study to develop a self-help app for youth with mental health problems. *Frontiers in Psychiatry*, 10:593, 2019.

[173] 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.

[174] 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.

[175] 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.

[176] 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.

[177] 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.

[178] Shiwei Wang. Sarcastic meaning of the slightly smiling face emoji from chinese twitter users: When a smiling face does not show friendliness. *International Journal of Languages, Literature and Linguistics*, 8(2):65–73, 2022.

[179] 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.

[180] 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.

[181] 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.

[182] Martin Weiß, Dariana Bille, Johannes Rodrigues, and Johannes Hewig. Age-related differences in emoji evaluation. *Experimental Aging Research*, 46(5):416–432, 2020.

[183] 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.

[184] Diyi Yang, Zheng Yao, Joseph Seering, and Robert Kraut. The channel matters: Self-disclosure, reciprocity and social support in online cancer support groups. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pages 1–15, 2019.

[185] 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.

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

[187] 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.

[188] 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.

[189] 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.

[190] 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.

[191] 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.

[192] 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.

[193] 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.

[194] Jolene Zywica and James Danowski. The faces of facebookers: Investigating social enhance-ment 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.

Appendix

6.1 OSSS

Table 6.1: The OSSS as it was presented in this dissertation (Chapter 5. Note that OSSS score is calculated as a sum of all responses, where a "Never" is coded as a 0 up to "A lot" which is coded as a 4. Thus a higher OSSS indicates more perceived social support.

Please rate **how often** the following things have happened for you **while you interacted with others** online over the last week. Use the following scale:	Never	Rarely	Sometimes	Pretty Often	A lot
People show that they care about me online.					
Online, people say or do things that make me feel good about myself.					
People encourage me when I'm online.					
People pay attention to me online.					
I get likes, favorites, upvotes, views, etc. online.					
I get positive comments online.					
When I'm online, people tell me they like the things I say or do.					
Online, people are interested in me as a person.					
People support me online.					
When I'm online, people make me feel good about myself.					
When I'm online, I talk or do things with other people.					
People spend time with me online.					
People hang out and do fun things with me online.					
Online, I belong to groups of people with similar interests.					
People talk with me online about things we have in common.					
Online, I connect with people who like the same things I do.					
I am part of groups online.					
When I'm online, people joke and kid around with me.					
People relate to me through things I say or do online.					
Online, people make me feel like I belong.					
When I'm online, people give me useful advice.					
Online, people provide me with helpful information.					
If I had a problem, people would help me online by saying what they would do.					
Online, people would tell me where to find help if I needed it.					
People help me learn new things when I'm online.					
People offer suggestions to me online.					
People tell me things I want to know online.					
When I'm online, people help me understand my situation better.					
If I had a problem, people would share their point of view online.					
People help me see things in new ways when I'm online.					
People online would help me with money or other things if I needed it.					
When I'm online, people help me with school or work.					
Online, people help me get things done.					
If I needed a hand doing something, I go online to find people who will help out.					
Online, people offer to do things for me.					
Online, people help me with causes or events that I think are important.					
When I'm online, people have offered me things I need.					
When I need something, I go online to find someone who might lend it to me.					
When I need a hand with school or work things, I get help from others online.					
I contact people online to get help or raise money for things I think are important.					

6.2 MSPSS

Table 6.2: The MSPSS as it was presented in this dissertation (Chapter 5). Note that MSPSS score is calculated as a sum of all responses, where a "Strongly disagree" is coded as a 1 up to "Strongly agree" which is coded as a 5. Thus a higher MSPSS indicates more perceived social support.

Please rate how much you agree with the following statements with regard to **this current point in time**.

	Strongly Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
There is a special person who is around when I am in need.					
There is a special person with whom I can share my joys and sorrows.					
My family really tries to help me.					
I get the emotional help and support I need from my family.					
I have a special person who is a real source of comfort to me.					
My friends really try to help me.					
I can count on my friends when things go wrong.					
I can talk about my problems with my family.					
I have friends with whom I can share my joys and sorrows.					
There is a special person in my life who cares about my feelings.					
My family is willing to help me make decisions.					
I can talk about my problems with my friends.					

6.3 Chirp Study Survey Questions (Chapter 5)

- In the past week, was there a point where you felt like you needed to go to someone for support? If so, who did you go to, and when was it?

- MSPSS (See Appendix 6.2)

- OSSS (See Appendix 6.1)

- What kinds of feelings does the following emoji represent?: 🙂

- What kinds of feelings does the following emoji represent?: 🥺

- What kinds of feelings does the following emoji represent?: 😊

- What kinds of feelings does the following emoji represent?: 😰

- What kinds of feelings does the following emoji represent?: 🤔

- What kinds of feelings does the following emoji represent?: 😐

- What kinds of feelings does the following emoji represent?: 😄

- What kinds of feelings does the following emoji represent?: 😡

- What kinds of feelings does the following emoji represent?: 😖

- What do you think the author of the following post in Chirp meant to convey?

  As a reminder, the first of these emojis was chosen in response to the prompt "How are you feeling?", while the others were chosen in response to followup prompt "Want to share why?"

- This is the end of the survey, is there anything else you would like to tell us?

  Included only for Social and Individual groups:

- Did you share your Chirp username with anyone? If so, who? (ex. a friend, parent, spouse)

- You made the following post in Chirp. Please explain below what you meant to convey in this post.

  As a reminder, the first of these emojis was chosen in response to the prompt "How are you feeling?", while the others were chosen in response to followup prompt "Want to share why?"

- In general, how did you feel when using emoji-only communication?

- At any point during your use of Chirp, did you use the Profile page as a mood tracking log? If so, what did you use it for?

  Included only for Social group:

- Were there any specific times where you felt you had a very good understanding of what a post on Chirp was trying to convey? Can you give an example and how you interpreted it?

- Were there any specific times where you felt you had a very bad understanding of what a post on Chirp was trying to convey? Can you give an example and your best guess at interpreting it?

- Is there any other user in Chirp that you felt particularly positive about or felt like you connected with? If so, what is their username?

- Is there any other user in Chirp that you felt particularly negative about? If so, what is their username?

6.4   Emoji Understanding Survey (Chapter 5)

- Are you at least 18 years of age?

- Are you currently residing in the United States?

- Have you ever used the Chirp social mood tracking application before OR have you ever participated in a previous study held at Brown University involving the Chirp social mood tracking application?

  Some images of the Chirp social mood tracking application shown below:

- How do you identify yourself? *(This question provides a list of gender identities to select from)*

- How old are you?

- Which of the following describes your current student status?

- Which year of undergraduate study are you in? *(If indicated that they are an undergraduate student)*

- How long have you lived in the US?

- Which of the following describes your current work status?

- What area do you work in? (ex. higher education, technology, construction, healthcare) *(If indicated that they are working*

- How often do you use emojis in your day to day life?

- How are you feeling? Choose an emoji to represent your mood.

- The following post was made on the Chirp app. The first of these emojis is a response to the prompt "How are you feeling?", and the following emojis are responding to the prompt "Want to share why?". This post was written by a [male/female] [young adult/adult] [who is currently a student/who is not currently a student]. *IMAGE OF EMOJI POST* What is your best guess of what the author of the post meant to convey?

- How confident are you that you interpreted the intended message of this post accurately?

- Another user provided the following interpretation of the post: *GROUND TRUTH DESCRIP-TION OF EMOJI POST* How confident are you that you correctly interpreted the original poster's intended message compared to before you saw this other user's interpretation? Please briefly explain why.

6.5   Emoji Understanding Content Analysis Code Book (Chapter 5)

Emoji Post & Description		They were feeling good and happy because a lot of good things happened. They went to the gym and saw a very nice sunset. They shared photos with a friend and they talked about it. They also got enough sleep and got some things done during the day

Themes	Inclusion Criteria	Exclusion Criteria
Happy	Description of a positive feeling (e.g., feeling positive, feeling good) Inclusion of a positive emotion word (e.g., happy)	Does not mention an emotion word, even if the description is positive (e.g., "having a good day")
Exercise	General mention of exercise or gym going (e.g., went to gym, did some exercise)	Mentions of specific types of exercise (went on run, lifted weights, sports)
Sunset	Specific mention of the sunset	No mention of the sunset, even if there is a mention of another time of day (e.g., sunrise, dusk, nighttime)
Social/Communication	Mentions any kind of communication or socializing (ex. hanging out with, seeing a friend)	No mention of communication or socializing
Enough Sleep	Mentions of sleeping or resting enough, statement that the poster went to sleep, description of "well rested"	Mentions of needing rest or sleep, naps, or that the poster was tired
Productive/Getting Things Done	Mentions of being accomplished, "got work done", or productivity	

Emoji Post & Description		They came down with a cold but they went shopping and also got guacamole and chips!

Themes	Inclusion Criteria	Exclusion Criteria
Cold/Illness	Mentions of any symptoms of being sick, description of being sick or illness, or specific mention of a cold	mention of a specific type of illness that is not a cold (e.g., flu, covid)
Shopping	Mentions anything to do with a purchase or shopping (e.g., bought medicine, went shopping)	No mentions of purchases or shopping, even if there are mentions of going out or spending money
Guacamole and Chips	Mention of avocado, or guacamole, "chips", or food as a general category	Mentions a specific food item other than avocado, guacamole, or chips without a mention of avocado, guacamole, or chips

Emoji Post & Description		They are happy because it is sunny today and they got watch the sunset. They spend a lot of the day outside, in nature where the wind was blowing and that made them feel good and refreshed. They ended the day with a yummy slice of pizza for dinner

Themes	Inclusion Criteria	Exclusion Criteria
Happy	Description of a positive feeling (e.g., feeling positive, feeling good) Inclusion of a positive emotion word (e.g., happy)	Does not mention an emotion word, even if the description is positive (e.g., "having a good day")
Sunny Weather	Reference to the weather and a positive descriptor (e.g., nice weather, warm weather, pleasant out)	Weather described as bad or not at all
Sunset	Specific mention of the sunset	No mention of the sunset, even if there is a mention of another time of day (e.g., sunrise, dusk, nighttime)
Nature	Mention of nature or seeing nature or being in nature	Does not include the word "nature" specifically
Outdoors	Mentions "outdoors," "outside," or "out" or anything that implies going outside (e.g., went to farmlands, in the forest)	No mention or implication of going outside, even if there is a mention of plants or other things found both outdoors and indoors
Wind	Mention of wind, or descriptors that imply wind (e.g., breeze, windy, gusts)	Description of other weather phenomena without reference to wind (e.g., rain, air, leaves falling)
Pizza	Mention of pizza or food as a general category	Mentions a specific food item other than pizza without a mention of pizza

Emoji Post & Description		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

Themes	Inclusion Criteria	Exclusion Criteria
Tired	Mentions feeling tired, exhausted, sleepy, drowsy	Reference to sleep that's not a feeling (e.g., "slept enough") Mentions of feeling lazy or unmotivated without mention of being tired
Coffee	Mentions coffee, caffeine, "hot drink", or types of coffee-based beverages (e.g., americano, grabbed caffeine)	Mentions a specific drink that is not coffee or coffee-based
Backpack	Mention backpack or a general word describing a bag	Mentions a specific type of bag that is not a backpack (e.g., purse, luggage, tote)
Bus	Mentions "bus," "schoolbus," or "public transit"	Mentions specific a form of transit that is not a bus (e.g., car, train, bike)
Work (location)	Mentions work as a location, "job," "workplace," "office"	Mentions "school", or uses "work" as a verb as opposed to a location
Doing Work	References being accomplished, "got work done," productivity, or doing homework	Mentions "work" as a noun rather than a verb, or mentions a specific type of work (e.g., writing a paper, writing a letter)
Significant Other	References spending time with or seeing a love interest or boyfriend/girlfriend (e.g., spending time with a crush, called their boyfriend)	No reference to spending time together or seeing each other, references asking someone out