# I'LL SEND ALL POSITIVE THOUGHTS OUT TO YOU: DETECTING HYPERPERSONAL RELATIONSHIPS THROUGH SELF-DISCLOSURE
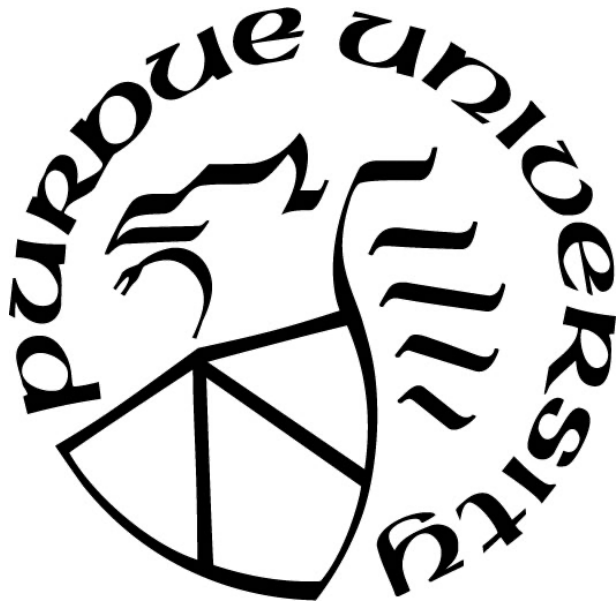
by

**Elise Taylor**

**A Dissertation**

*Submitted to the Faculty of Purdue University*

*In Partial Fulfillment of the Requirements for the degree of*

**Doctor of Philosophy**

Brian Lamb School of Communication

West Lafayette, Indiana

August 2022

# THE PURDUE UNIVERSITY GRADUATE SCHOOL
# STATEMENT OF COMMITTEE APPROVAL

**Dr. Sorin A. Matei, Chair**

Brian Lamb School of Communication

**Dr. Stacey Connaughton**

Brian Lamb School of Communication

**Dr. Emily Buehler**

Brian Lamb School of Communication

**Dr. Matthew Hannah**

Libraries and School of Information Studies

**Approved by:**

Dr. Marifran Mattson

*To MeeMaw, who read every word of the first one and has assured me that she will read every word of this. I'm so glad you're in my life.*

*Except for Appendix B. That goes to Nathan, to replace Appendix A.*

# ACKNOWLEDGEMENTS

Finally, every dissertation has a cast of characters who work behind the scenes to help ensure the end document is clear, coherent, and functional.

Dominic Edison – for helping me figure out various programming things in Python

Jonathan "The Eagle" Claypool – both for doing the heavy Python lifting and helping me learn to code my way out of def paper_bag(elise):

Special gratitude to Drs. Mimi Brinberg, Fritz Günther, and Robin Mukherjee for their methodological and statistical help.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# ABSTRACT

The hyperpersonal model predicts computer-mediated communication can allow for relationships formed in that medium to become more intimate than their offline counterparts. Specifically, it combines ideas first presented in social information processing (i.e., that the volume of information exchanged over time within computer-mediated relationship is more important than how long it takes to exchange that volume) with the technological affordances given to the sender, receiver, message, and channel in order to create a feedback loop of assumed good intentions within the CMC medium that allows for online relationships to surpass face-to-face relationships in terms of their emotional intimacy. Existing research has shown that a variety of factors influence how people feel about an online friend, including the richness of the medium, personality, and the amount of emotional self-disclosure that had been exchanged within the relationship. However, studies to date have inconsistently measured self-disclosure and largely rely on survey or experimental methods rather than the examination of existing text-based datasets. This study proposes and tests a model that the relationship between an initial person's emotional self-disclosure and the reciprocal self-disclosure the friend responds with in a CMC medium is mediated by the degree to which their language converges, or the degree to which they empathize with each other, and is moderated by the volume of text exchanged by the pairs during their relationship. The study uses a corpus of the text messages exchanged between 2,174 pairs of people. The results indicate that there is an indirect relationship between initial self-disclosure and reciprocal self-disclosure which is mediated by is empathetic convergence. Furthermore, the volume of information exchanged may also play a role in some of these interactions. This study offers implications and suggestions for refining the hyperpersonal model to be applicable in the current digital zeitgeist.

# CHAPTER 1: INTRODUCTION

I came of age in the early days of the Internet's popularity, when the rallying cries of online culture were "Information wants to be free" and we were just beginning to learn the power of developing relationships online. I had my first significant romantic relationship online, culminating in a trip to Toronto, a school dance, and – soon after – a breakup. My father had an affair with a woman online before such things became more fashionable than scandalous. Later, I would make dear friends online, some of whom are still friends two decades later. Now, as a parent, I see similar patterns happen in my own child; being a so-called "digital native" has not exempted them from the vagaries of extreme emotion online. Developing friendships and falling in love happen at a pace that cannot be explained by mere teenage hormones. For both of us, and many others, computer-mediated relationships and communication more broadly were an inexorable aspect of adolescent socioemotional development.

Although some visionaries recognized the power of online relationships, many were initially skeptical, perhaps mirroring a skepticism present in larger society that "somehow people develop trust over computers and often they're dead wrong" (Swarns, 1996, para 11). Some of the earliest research doubted that computer-mediated communication (CMC) could allow its participants to be "present" at all since it is a medium with inherently less paralanguage than a face-to-face (FtF) encounter (Short et al., 1976). Media richness theory argues that communicators add depth (i.e., richness) to their media interactions through four different mechanisms: the number of individual mediums, or cues, that can be transmitted within the medium; how quickly feedback can be transmitted and received; the variety of language used in it (i.e., whether or not the language is natural or invented; and the personal focus of the medium, or how intimately people can use the medium (Daft & Lengel, 1986; Daft & Wiginton, 1979). For example, a FtF interaction would be the proverbial gold standard since communicators can fully utilize all aspects. In contrast, a letter written would have a great variety of potential language and a very personal focus, but individual cues possible would be few and the response time slow. Due to the reduced cues, the conventional wisdom suggested that CMC communicators would have less intimacy than their FtF counterparts.

Despite the initial reticence, research began to emerge that cast doubt on this assumption. For example, experiments showed relatively little difference in the *quality* of communication in leaner mediums (Dennis & Kinney, 1998). Within the context of friendships specifically, differences between FtF and CMC friendships diminished as the length of the relationship increased, and some results even suggested that CMC relationships could even provide unique opportunities for relational development, challenging the dominant perception of the medium (Chan & Cheng, 2004; Walther & Burgoon, 1992). An early study surveyed users of Multi-User-Dungeons, Object Oriented (MOOs), which are online, text-based worlds where users can converse and interact with others in real-time, to determine what kinds of long-term relationships they had experienced in that platform (Parks & Roberts, 1998). This approach stood in contrast to previous studies that had used strangers in laboratory settings. Their survey indicated that the overwhelming majority of participants, 93.6%, had at least one ongoing friendship that they had developed while on MOOs, and the average person had developed five new, close relationships. The duration, breadth, depth, and intensity of CMC relationships were comparable to those of FtF relationships. Scholars have noted that within these entirely online communities "there is nothing about virtual worlds that make them *uniquely* well suited for friendship formation" (Welles et al., 2014, p. 11, emphasis in the original). This also means that there is nothing about virtual worlds that make them uniquely *ill-suited* for friendship formation, thus casting further shadow on previous research that implied a weakness or paucity in virtual friendships.

In the intervening decades, the world has become increasingly computer-mediated. Nearly 40% of romantic couples meet first online (Rosenfeld et al., 2019). 37% of the American population can work entirely from home (Dingel & Neiman, 2020). Significant parts of our lives, whether professional or intimate, take place in an online setting, and thus the ability to develop and sustain relationships in that environment has become clearer and more obvious. As we move away from being predominantly centered within our local communities and become broadly networked entities who are connected to the world at large, we have increasingly shifted from *homo sapiens* to *homo conexus*, or the human characterized by technologically driven connection (Bay, 2009). Our jobs, our friends, and our lovers interact with us through a computer-mediated lens.

Despite this, we do not yet fully understand the relational mechanisms behind the formation of relationships in general, and specifically friendship online. Walther (1992)

proposed that, given enough time, online relationships could be equally as deep and significant as offline relationships. Because CMC is a low-bandwidth medium, the amount of time required to type the same information that can be exchanged in a FtF interaction is considerably longer. He labeled this phenomenon "Social Information Processing," or SIP (also creating a clever mnemonic: people take *sips* of information over time). So long as the total volume of information exchanged, according to SIP, matches that which was exchanged in an FtF encounter, the pair is capable of the same depth of relational intimacy. Within decision-making scenarios, FtF groups were certainly the fastest at reaching consensus, but CMC groups, if given unlimited time, reached the same decision at approximately the same volume of total information exchanged. Simply put, "CMC and face-to-face groups operate at different rates" (Walther, 1992, p. 61).

Later, Walther (1996) built on the ideas first presented in SIP to suggest that the online medium affords the ability to become "hyperpersonal," in that the participants can feel even more intimacy in an CMC setting than FtF through an ever-increasing spiral of intimacy, bonding, and connection. The most important difference between CMC and FtF communication is not the amount of communication, but the rate at which it is exchanged; we type more slowly than we speak. Therefore, any "accrual of interpersonal effects is expected to be slower in time and develop in proportion to the accumulation of message exchanges" (Walther, 1996, p. 10). However, he noted that there was anecdotal evidence that people in CMC relationships could feel even closer to their online friends than their FtF friends and proposed that when two people with friendly relations send messages, they are incentivized to create messages that will be positively received by the recipient. Absent other information, any ambiguities in the sender's intent are filtered through the receiver's pre-existing positive opinion, thus having a net-positive effect on their opinion of the sender. The desire to make and leave a good impression also creates an active incentive to select what information to disclose in order to achieve their ultimate goal (whether that goal be a romantic relationship, general friendship, or simply continuing the conversation). With a FtF pair, even with the equivalent volume of information exchanged, they are not able to selectively self-present to the same degree that a CMC pair can, and therefore the degree to which intimacy forms is reduced compared to the more intimate (i.e., hyperpersonal) relationship that a CMC couple can achieve. The process creates a spiral of positive perception that serves to reinforce and escalate intimacy, and it creates the positive equivalent of online

incivility spirals in which negative and hurtful messages beget more of the same (Welch, 2020). Patterns within incivility spirals also suggest that the rapidity and intensity at which these relationships escalate is due to disproportional escalation (Felson & Steadman, 1983). In the context of incivility, this escalation is the result of topic sensitivity and an individual's desire to incite others to anger (i.e., "trolling"). However, the hyperpersonal model was developed to explain and predict a dyad's *internal* emotional state and feelings towards each other that they achieve through various externally observable behavior, such as selective self-presentation. Emotional spirals focus on the observable phenomenon of incivility. This project makes use of the external clues that can be discerned from language patterns that reveal internal emotional states.

Another key question is how to determine the reciprocity element of intimacy. Many scholars point to self-disclosure as a specific mechanism in the development of offline relationships, and therefore suggest that increasingly intimate self-disclosures further the formation of online relationships as well. The present study explores this phenomenon in depth. Furthermore, it explores the possibility that hyperpersonal relationships, which are discussed below as a particular type of high-intensity, high-engagement connections that surpass normal relationships, emerge based on self-disclosure. Altman and Taylor (1973) compared self-disclosure to peeling an onion, in which each individual self-disclosure removes an additional layer until the parties know the most intimate details about one another. The obverse may be also true: as people build relationships, the relationship grows around self-disclosures like an onion growing larger throughout the season. Yet, self-disclosures are not necessarily easy to manage or initiate; demographic variables, such as age, gender, and culture (Antheunis et al., 2007; Gupta et al., 2010), often affect the frequency and depth of self-disclosure (Vijayakumar & Pfeifer, 2020), as well as the comparative rates online versus offline. Furthermore, to engage in self-disclosure the individual must have a positive opinion of online relationship formation (Attrill & Jalil, 2011). If they have a negative opinion of online relationship formation, they will not engage in the self-disclosure behaviors required to develop relationships.

Qualitatively, we know that online relationships can become hyperpersonal. Henderson and Gilding (2004) found that participants saw their relationships having little to do with the medium in which they were formed, but instead focused on the interpersonal exchanges they had had with the other person. Absent of other nonverbal clues, they assigned particular importance

to text-based communication, as the friendship is built on communication, rather than a shared history of activities. Interestingly, they found that participants were more likely to elaborate and provide greater detail when doing an interview in an online setting than in FtF interviews, thus confirming the tendency to engage in hyperpersonal connections, even when the connection was with a researcher, not a friend. Netnographic examinations of online social groups suggest that some hyperpersonal connection may be due to being unable to experience certain kinds of connection offline (Workman & Coleman, 2014). As a result, people who feel isolated (either geographically, from a larger community, or due to other circumstances) may engage in hyperpersonal relationships more often than those who are not isolated. We must also examine the situations and processes by which these relationships develop.

Quantitative evidence of hyperpersonal relational development has remained largely elusive. A 2011 meta-analysis found that online relationships have similar levels of intimacy, but not the greater intimacy that the hyperpersonal model predicts (Kim & Dindia, 2011). Although the authors acknowledge issues that could lead to errors within this meta-analysis (as will be discussed later), it points to a disconnect. It is also possible that the existing research simply is using methods that cannot detect these relationships. Kim and Dindia note that few of the studies examined actual self-disclosure, versus self-reported disclosure, and that 11 studies are not enough to make any conclusive statements. Furthermore, the studies used different methodologies, which may have limited their ability to consistently measure and identify hyperpersonal relationships. Of the 11 studies, only one examined an existing friendship, and that was limited to the self-reported information the participants provided. The other ten experiments largely used people who had been strangers before the experiment (six), and most often used self-reported data (eight), and only two of them allowed a longer time for the CMC communicators to interact than their FtF counterparts.

Rather than conclude that these relationships do not exist, we must consider that these relationships, driven by intimacy and self-disclosure over a long period of time, are simply difficult to uncover using survey-based and experimental methods. Past studies using these methods may not have been able to fully and completely capture the amount of communication necessary for hyperpersonal effects to emerge. Therefore, we have an incomplete picture of how interpersonal relationships develop in online settings. We know they do exist, but if we want to create the environments in online communication that allow for these relationships to flourish,

we must also be able to identify exactly how they develop so we can encourage them. As our world becomes increasingly mediated by online communication, we are faced with the fact that if we want our world to be friendlier, we need to understand the most environments in which positive interpersonal relationships blossoms. This paper looks to achieve that goal and fill the gap in research by first developing a model by which we can analyze and predict the deepening intimacy that is the hallmark of hyperpersonal relationships. Then, I will use this model to analyze variables extracted from a chat dataset which is comprised of chat logs of dyads communicating in a non-laboratory environment. In Chapter 2, I will review the relevant literature regarding the nature of computer-mediated communication, the hyperpersonal model, self-disclosure as it relates to online communication, and propose hypotheses based in this existing literature. In Chapter 3, I discuss the challenges in finding a comprehensive dataset of conversations amongst pairs of communicants, the specifics of the BOLT chat dataset and how I transformed it to be appropriate for statistical analysis, and the statistical methods I used. In Chapter 4, I share the results of the analysis and compare the results against the initial hypotheses proposed. Finally, in Chapter 5, I analyze these findings in detail, summarize the implications, and discuss both limitations and future research opportunities.

# CHAPTER 2: LITERATURE REVIEW

This chapter serves as a guidepost for our examination of hyperpersonal relationships. First, we explore the key assumptions of computer-mediated relationships, which allows us to understand the landscape of computer-mediated communication from which the hyperpersonal model grew. Next, we discuss the hyperpersonal model and how that model conceives of the development of online relationships. Then, we will discuss social penetration theory, which will allow us to theorize the mechanics of hyperpersonal relationships more deeply. Finally, I will establish the hypotheses that guide this study.

## Assumptions of Computer-Mediated Communication

Academic examination of computer-mediated communication gained steam in the 1990s as computers became increasingly common as household items, and three dominant ontological assumptions have emerged that will guide this paper, namely that although text-based computer-mediated communication (CMC) is a lean medium, that deficit can be overcome and users are, indeed, driven to overcome it. All three rely on the underlying assumption that human nature is both communicative and social, and therefore online communication, too, has a social nature. With its typed syllables flowing across a computer screen, CMC differs from face-to-face communication, where thoughts are transmitted by phonemes and body language, but no matter the medium, people are driven to connect and communicate. Next, I will explore these three assumptions and discuss the implications for identifying hyperpersonal relationships in these environments.

### *Assumption 1: Computer-mediated communication as a lean environment*

As networked computers began to increasingly infiltrate the homes and office spaces of the 1970s and 1980s, moving from novelty to a core part of the work experience, researchers turned to this new medium to discover how it affected communication. Early research pointed to classic CMC as a lean medium with reduced social cues, which led users to feel like their co-communicator was less present when engaging with that person through a computer-mediated

channel. It was envisioned as if a person were staring into the distance during a conversation; if one cannot see indicators of verbal immediacy in their conversational partner, such as active listening, eye contact, etc., then they will view the person as less present. In this view, the "richest" communication channel is face-to-face (FtF) communication as it can encompass the full range of verbal, nonverbal, and paraverbal messages, and all other communication channels are compared against this proverbial gold standard. As a result, the "social presence" felt by dyads using leaner channels of communication should be lessened, and thus the relationships built in the environment would be less intimate, if they existed at all (Short et al., 1976). As such, when users want to form emotional connections and communicate deep emotions and ideas, they would have to choose FtF communication (or other richer communication modalities, such as synchronous video conferencing) to convey the same depth of nuance.

Furthermore, media richness theory acknowledges that different communication tasks require different inherent bandwidths for information transmittal and suggests that people find the most suitable channel to match the bandwidth of their task (Daft & Lengel, 1986). The more information that needs to be exchanged, the "richer" a medium one would choose. In a CMC setting, people would choose to remove the conversation from a digitally-mediated format for conversations that are emotionally complex, based on the amount of information capacity of that channel. In theory, media richness theory is intuitive; many people have had the experience of choosing to respond to a text message via phone call, rather than continue the text exchange, and therefore we believe that if we are not communicating with enough socioemotional depth, we would choose a richer medium. In practice, media richness theory fails to consider two elements. First, people do not always make the most rational decision, and thus even if a richer medium might be appropriate and we acknowledge it as such, that does not mean that we make that choice in the moment (Walther & Parks, 2002). It also does not account for other factors that may influence a person's decision to utilize a lean media, such as personality (Hertel et al., 2008), hierarchy and social standing (Sheer & Chen, 2004), and threats to face or loss of privacy (O'Sullivan, 2000). We have also seen in practice that people regularly choose leaner environments over richer options and still manage to have thriving interpersonal relationships within them. This both provides a prelude to hyperpersonal communication and leads to the next core assumption of computer-mediated communication: people enrich lean environments.

***Assumption 2: People enrich lean environments through the volume of information
exchanged and technical affordances of the medium.***

The second core assumption of CMC is that although a given medium may lend itself to
leaner communication, users of CMC media channels find ways to enrich the channel. The first,
and perhaps simplest, way that people enrich lean environments is through how much they use
the platform and the amount of information they use it to convey. As in FtF communication,
relationships are built over time through dyadic communication, but unlike in FtF
communication, because it takes longer to communicate the same number of words, CMC
relationships take longer to develop (Walther, 1992). Past experiences affect current
relationships as well, since individuals build databases of experiences and understandings within
a given computer-mediated medium that result in increased perceptions of how rich a given
communication channel is (Carlson & Zmud, 1999). This so-called database is influenced by
how much experience the person has communicating with the specific person on that specific
medium and the specific situation being communicated. Furthermore, these factors can ebb and
flow over time, as more experienced users of a platform reach an equilibrium after which
additional experience does not significantly affect their perceptions. Users of lean
communication channels are able, through the culmination of their experiences with a particular
communicative partner, to experience the channel as a less-lean setting.

The second way that people enrich lean environments is through the technical
affordances of those environment that allow a certain degree of robustness to develop. For
example, people may use emoticons to convey more nuanced emotions than text generally
allows. Emoticons, emojis, and their successors are recognized as indicators of emotion (Byron,
2008), illocutionary force (Dresner & Herring, 2010), and sarcasm (Thompson & Filik, 2016),
and are explicitly and deliberately used to deepen the channel depth of CMC and to supplement
the lack of social cues in CMC. Not only do these additions result in enriched lean environments
between two communicative participants, but they can also lead to more widespread meanings
that remain consistent over time unless they become co-opted by social movements (Robertson
et al., 2021). As with traditional oral and written communication, computer-mediated
communication also evolves in response to the wider society that uses it. Therefore, much as
language evolved to accommodate emotional nuance and other technological advancements

(ranging from written communication to the telephone), it also continues to evolve to deepen "lean" CMC avenues.

As a result of using channel affordances to enrich and deepen lean communication channels, the resultant communication is not a hobbled, haphazard form, but instead is fully capable of conveying the same level of complex, subtle messages as we find in FtF communication. This leads to the third core assumption of CMC: the relationships formed in CMC environments can be equally intimate as those formed in offline environments.

***Assumption 3: People can form equally intimate relationships within computer-mediated contexts as in face-to-face contexts.***

Social information processing, or SIP, comes from an entirely communicative background (Walther, 1992). SIP consisted of five assumptions and three propositions that formed the base of the theory and proposed that humans, by their nature as social creatures, affiliate and engage in relational communication. In a computer-mediated environment, this process of encoding and decoding takes longer since individuals must adapt their verbal and nonverbal communication cues. Therefore, if we give a sender/receiver communicating in a CMC environment enough time to exchange the same volume of information as they would otherwise do with FtF communication, the "relational valences…will be the same" (Walther, 1992, p. 69). In other words, regardless of whether you fill the communicative bucket with a cup or a teaspoon, the bucket is still filled; the only change is how long the filling takes.

The specific mechanisms by which people form lasting online relationships equal to their FtF counterparts are not as clear. In the next section, I will discuss two ways that this study will explore and explain these mechanisms. First, I will discuss the hyperpersonal model which suggests that not only can people form equal relationships in a CMC environment, but people can form relationships that are *more* intimate than FtF conversations. Then, I will discuss how self-disclosure is a necessary precursor to the development of this intimacy that the hyperpersonal model relies upon.

## Hyperpersonal Model

As relationships online began to become more ubiquitous, the hyperpersonal model grew from the belief that the immediacy, affection, similarity, and depth that two people experienced

in CMC relationships could surpass, rather than equal, that of FtF relationships (Walther, 1996). It was originally proposed as an extension of social information processing (SIP) through adapting to criticism levied at SIP by proponents of the social identity model of deindividuation effects (Walther et al., 2015). Deepening online relationships were explained by the CMC environment creating an exaggerated, larger-than-life relationship based on the four components of communication: Sender, receiver, channel, and feedback.

Within a text-based medium, senders have a unique ability to custom tailor their messages, resulting in *selective self-presentation*. Unlike FtF communication, senders have a backspace key and the time necessary to perfectly craft the exact intention they want to convey. Examples of this can be seen in a variety of contexts, including online dating (Toma et al., 2008), winning and losing in video games (Downs & Sundar, 2011), and disclosure of disabilities (Furr et al., 2016). Online mediums also frequently allow the post-hoc editing of one's message or profile, allowing further customization of their online images to ideally result in a more positive perception than reality is capable of inducing, but there is a point of diminishing returns; Toma and Hancock (2011) found, perhaps unsurprisingly, that when people intuit lies, or come to believe that a person is misrepresenting themselves, they view it as unacceptable and discontinue the relationship. Simply put, with the ability to selectively self-present yourself, your online friend need never to know you forgot to take out the metaphorical trash, but if you imply you have *no* trash, they may find it unbelievable.

If an individual remains within the realm of believability, the hyperpersonal model posits that the receiver will accept the information. The social identity model of deindividuation effects (SIDE) proposes that within small group communication, anonymization of individual participants results in deindividuation and the polarization of the individual opinion towards the group mean (Lea & Spears, 1991). Walther (1996) applied this thought to the individual and proposed that when lacking other social cues, people engage in overattribution based on their existing opinions of the person. For example, if an individual communicating in a CMC environment received an ambiguous message from a close confidant, they would be more likely to make assumptions based on their idealized assumptions, and thus assume positive intentions. However, the same text message from a more negative party would assume negative intentions. Within the context of relationship building, this implies that when a person who you view positively (e.g., a potential job contact, budding new relationship, etc.) makes a statement, the

receiver is likely to guess the statement is more positive, which furthers their existing positive perception of the sender.

The third dimension to the hyperpersonal model is the channel by which the message is sent, including how one interacts with this channel. The hyperpersonal model predicts, in line with the assumptions of CMC, that users will take advantage of the technical affordances of a given channel to customize messages, using editing and deletion tools to ensure that their message has the desired effect. When a user's goal is to seek friendship and affinity within a CMC environment, they tend to do more advanced preparatory work to uncover their conversational partner's likes and dislikes (Walther et al., 2010). They also spend greater time editing messages when they are motivated to have the other party have a positive opinion of them (Walther, 2007). Users deliberately and knowingly utilize the technical abilities of CMC channels to influence their communicative partner's positive perception of them.
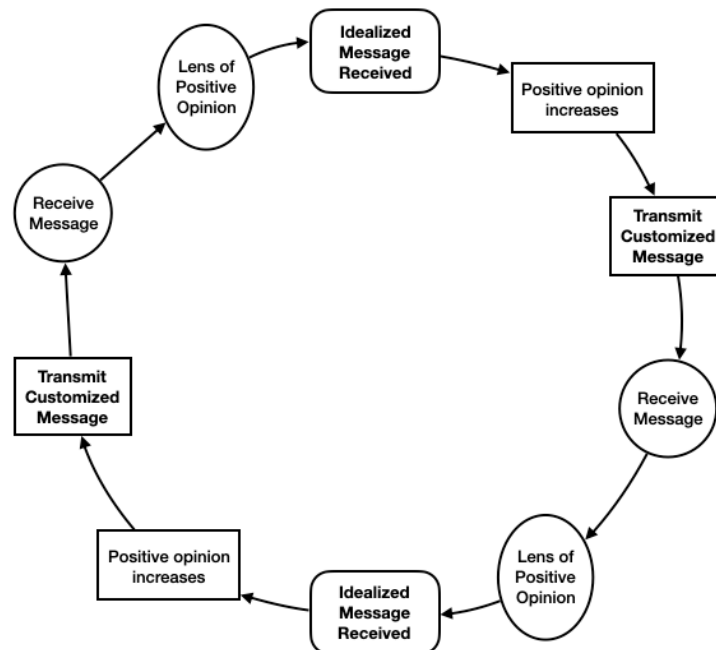


Figure 1. The hyperpersonal model

Finally, the hyperpersonal model proposed that due to the confluence of these three factors – the channel allowing the perfecting of self-presentation, the receiver filling in the gaps in their knowledge of the other party, and the sender being deliberate and mindful of their

communication – a feedback loop is created. As a result, CMC interactions can become more intimate and elicit a greater response than a similar FtF interaction.

> That is, when a receiver comes upon a selectively self-presented message and idealizes its source, that individual may respond in such a way that reciprocates and reinforces the partially modified personae, reproducing, enhancing, and potentially exaggerating them. The manner by which the dynamics of these reciprocated expectations modify the participant's character may reflect the process of behavioral confirmation. (Walther et al., 2015, p. 6)

This system of ever-increasing perceptions explains a variety of interpersonal interactions in online environments. Some users report feeling that they have never felt such a close camaraderie with their offline friends, possibly the result of being able and willing to disclose secrets that they have not shared with even their closest friends (Freeman et al., 2016; Henderson & Gilding, 2004). As a result, the theorized hyperpersonal relationship feels more intimate while having a level of self-disclosure that is deeper or more frequent than in their offline relationships. This lends further evidence to the idea that there is a causal link between self-disclosure and feelings of intimacy in online spaces as well as offline (Jiang et al., 2011). Walther's early work largely focuses on positive relationships and community formation and implies that the very name "hyperpersonal" references a positive and supportive type of intimacy. Despite these optimistic origins, not all online relationships are destined for an intimate, hyperpersonal-positive level; the hyperpersonal feedback loop can be negative as well if someone is treated poorly online. If someone responds to hostility in kind, this creates an identical, "hypernegative" feedback loop (Sillars & Zorn, 2021; Vossen et al., 2017). Much like positive hyperpersonal reactions can result in a level of idealization and intimacy unrealistic in FtF communication, negative hyperpersonal reactions can produce a similarly significant amount of negativity. Whether it be the result of flaming, cyberbullying, or simply very negative conversations present in a CMC environment, the feelings that result from those negative interactions are disproportionate to what someone might feel if a similar conversation happened offline. Interestingly, the message sender's anonymity *decreases* the amount of trauma/harm that a message sender feels due to the trolling/flame war experiences (Staude-Müller et al., 2012). In short, it is the betrayal of confidences that hurts the most.

These deep emotional connections – both those that lift and those that injure – would not be possible if not for the development of emotional intimacy and connection through the disclosure of personal information. To understand this mechanism, we must now explore the role that interpersonal intimacy has in the growth of relationships in general, and hyperpersonal relationships specifically. We will first examine the importance of intimacy and self-disclosure within relationships, and then more deeply examine the mechanisms of self-disclosure and intimacy through the lens of Altman and Taylor's social penetration theory. The goal is to use these two theoretical strands to further enhance the hyperpersonal model and show potential ways by which CMC relationships can become more intimate than their FtF counterparts.

## The Development of Intimacy

Whether the relationship is based in an online chat room, an offline office, or a combination of the two, we understand what being intimate with another person in a relationship feels like, but defining it outside of a dictionary is difficult, and – as with many emotions – the best colloquialism to describe it is, "we know it when we see it." Although the hyperpersonal model speaks to the development of relational intimacy, other research outlines the individual attributes of what it means to feel intimate within a relationship. Three of the most significant aspects of relational intimacy are immediacy/affection, similarity/depth, and receptivity/trust. (Burgoon & Hale, 1987). By these measures, within an intimate relationship, people feel like their communicative partner is interested in and enthusiastic about their interactions, they behaved like a friend, and they each are perceived to have a level of trust in each other. Parks and Floyd (1996) examined the same question with relationships that began online and found that online dyads tend to have a great deal of breadth and depth of interaction. This means that as the intimacy felt increases, they discuss more topics, engage in more activities together, reveal more sensitive information, and increase the number of communication channels they use to communicate. They also found that dyads had moderate levels of interdependence, commitment, and predictability/understanding. This suggests that the dyads depended on one another for relational needs (interdependence), had an expectation that the relationship would continue (commitment), and believed that they understood their partner's personality and how they were likely to react given a specific situation (predictability/understanding). Interestingly, unlike in offline relationships, they found little support for the dyads' social circles becoming increasingly

overlapped and merging over time. The lack of network convergence may be an effect of the lack of widespread Internet access since relatively few people regularly used the internet in 1996. The strong presence of breadth and depth of interaction is of particular interest, since they are measures of the number of different topics discussed and the degree of self-disclosure, respectively. When researchers compared online/offline friendships throughout the relationship lifecycle, they determined that although FtF relationships maximize the breadth and depth of the relationship more quickly than online relationships do, the difference fades as the online friendships mature (Chan & Cheng, 2004). This lends itself to examining social penetration theory and disclosure as a necessary antecedent to deepening friendships and the ever-increasing intimacy predicted to happen within hyperpersonal relationships.

More recently, researchers examined donated text messages from couples in romantic relationships to analyze their behavior over time and found that the average number of messages exchanged per day (approximately 161) grew during relationship formation, but the average number of characters in text message and the speed at which each person responds were static (Brinberg et al., 2021). The regularity in communication suggests that interdependence is as important an aspect of CMC as it is in FtF relationships, as the pairs grow to depend on the other person to respond to text messages. Although the researchers did not investigate the data to see the role of breath and depth of self-disclosure, they did determine that during the relationship formation stages of a relationship, couples' usage of language became exponentially more similar until it plateaued after they declared themselves an official couple (Brinberg & Ram, 2021). Although their data was collected from "real life" dyads with no comparison for dyads who communicate exclusively online, it does speak to elements of empathy and linguistic homophily that develop throughout a relationship. Specifically, they found that the couples' use of language tended to converge over time, but they have not yet followed up with the participants to determine if this convergence was correlated to relationship satisfaction or feelings of intimacy; the relationships could have dissolved as soon as the couple left the research lab (Brinberg & Ram, 2021). Furthermore, although all the text messages from each couple were analyzed, they were not able to use *all* communicative avenues, so if the couple communicated via avenues other than Apple iMessage, that data was not included.

# Social Penetration Theory

We have now established that online communication, although the base format may be lean compared with its FtF counterpart, relies on its users to take advantage of the technical affordances of the medium to make the medium less lean. When the medium is perceived less lean, it is capable of conveying more emotional content than otherwise might be expected. Therefore, relationships in this medium are capable of at least the same level of emotional intimacy as offline relationships, if not even more, which justifies the assumption that it might foster hyperpersonal relationships. I have also posed that this deepening of emotional intimacy happens through the combined force of dyads discussing a wide variety of topics (i.e., breadth) in a deeply personal way (i.e., depth). Through breadth and depth, pairs engage in a significant amount of self-disclosure about the goings-on in their lives, their thoughts and emotions related to certain topics, and so forth. Furthermore, because they can tailor their messages to ensure they present only the best version of themselves, the hyperpersonal model suggests that these relationships can become more emotionally intimate than offline counterparts.

To imagine a framework that will allow us to conceptualize this process, I will use Altman and Taylor's (1973) social penetration theory (SPT) which allows us to consider the self-disclosure patterns that people engage in while forming and maintaining relationships. SPT compares individuals' personalities to an onion that has multiple layers that are peeled back through the process of self-disclosure. The first of these layers, or the outmost part of the metaphorical onion, include biographical details such as those that would be gleaned by anyone (e.g., age, gender, ethnicity, etc.). The second layer consists of attitudes, opinions, and so forth. The most intimate layer consists of beliefs, fears, and needs. Altman and Taylor's conception of social penetration is that over time, through communication, people engage in increased self-disclosure, and thus gain access to "deeper" levels of a person. The movement of a relationship from "stranger" to "deepest, most intimate" relationship consists of four stages (Altman et al., 1981). First, the individuals have infrequent communications and characterize their relationship as somewhat superficial. The second stage, once the pair has engaged in a more rewarding exchange, expands the communication to "broad and shallow" communication, in which the pair has frequent contacts, but the depth of the conversation is still superficial. Next, the relationship becomes narrow and deep, where the communication is solidly comfortable in the second layer and broaches the most intimate disclosures. Finally, relationships can become both broad and

deep; this is often compared to a "best friend" or romantic partner-level relationship. These developments are not coincidental, as the disclosure of intimate information is inextricably linked to relationship development. Without self-disclosure, relationships can find no fertile ground from which to grow.

The further implication of SPT is that self-disclosure causes a person to "like" the person who disclosed personal information. Known as the "disclosure-liking hypothesis," meta-analysis found in general that self-disclosure causes one to "like" the other (Collins & Miller, 1994). They also found that the degree to which participants liked each other increased regardless of whether they were on the sending or receiving end of the disclosure and that gender does not seem to impact the overall amount of "liking." They also noted that the depth of self-disclosures (or, *how personal* the information was disclosed) had a stronger correlation to liking than the breadth of self-disclosures (or *how much* information was disclosed). Finally, they found that reciprocal self-disclosure creates a feedback loop of increasingly frequent disclosures. Although this study did not specifically examine online uses, later research did so in an experimental setting and found that participants do like those who disclose to them but did not find a reciprocal feedback loop (Kashian et al., 2017). They suggested that this unexpected result may have been due to the idea that friends need to disclose less to each other than to people with whom they do not already have close relationships. They also pointed to the experimental nature of their study as a potentially limiting factor; since the parties knew that the study was centered on self-disclosure, they may have reported responses that would not necessarily reflect their actual experience. To date, there are no studies that examine these phenomenon using data collected in natural settings, such as unobserved chat rooms and SMS. Therefore, we do not yet know if initial self-disclosure causes reciprocal self-disclosure in these settings.

Although SPT pre-dates the widespread use of the Internet as a communicative venue, it has been frequently applied successfully to CMC contexts. In experiments, researchers have found that CMC communicators disclose at a greater depth (Antheunis et al., 2007; Coleman et al., 1999) and a greater frequency (Tidwell & Walther, 2002), but systematic review shows that surveys tend to either find no difference or report deeper levels of FtF disclosures (Nguyen et al., 2012). Again, this points to a disconnect between individuals' perceptions of their activities (within surveys), and their actual observed outputs (in experiments). Relying on experiments often does not allow researchers to see how people react, disclose, and bond over extended

periods of time in natural settings. They also found a great deal of variability in terms of how self-disclosure was defined and measured and found that the surveys tended to measure the dimensions of self-disclosure (i.e., depth, breadth, etc.) inconsistently, and sometimes didn't even measure self-disclosure as a construct, instead measuring friendship quality.

## A New Model for the Emergence of Hyperpersonal Relationships: Rationale and Variables

Despite the robustness of research surrounding the hyperpersonal model, self-disclosure, and social penetration theory in the context of computer-mediated communication, there is a lack of consensus on how these three ideas intersect. Existing research lacks a clear, consistent operationalization and measurement of self-disclosure, and largely relies on surveys or experiments collected over a relatively brief period. The lack of research that incorporates large datasets collected over time is particularly egregious, given the opportunities that those datasets provide to act as longitudinal datasets without the cost associated with them. This study looks to fill these two gaps. First, I propose a model that contributes to a better specification of the hyperpersonal model by analyzing both factors that are currently frequently applied in existing research (such as self-disclosure) as well as offering complementary ones (such as empathetic convergence). Second, this paper makes use of large datasets gathered over extended lengths of time to examine the volume of communication that will allow us to see if there is an upper limit to intimacy formed in CMC interpersonal relationships. It is important to note that this study does *not* confirm the existence of hyperpersonal relationships since that requires a comparison to a FtF counterpart. However, it does attempt to add to our knowledge of the circumstances under which hyperpersonal relationships might emerge and create a replicable model that can be tested in a comparative manner.

### *Disclosure as a means to hyperpersonal relationships*

SPT and the hyperpersonal model are well-established within their respective domains and have been used in conjunction with one another to establish that reciprocal disclosures online led to greater intimacy (Jiang et al., 2011; Walther et al., 2016), more time spent communicating online leads to higher relationship satisfaction measures such as perceived commitment (Anderson & Emmers-Sommer, 2006), and to examine the mechanisms of online support groups

28

(VanLear et al., 2005). It is pertinent to discuss self-disclosure as it relates to online relationships, and hyperpersonal relationships specifically. My goal is to show that the conditions for hyperpersonal relationships to grow emerge from the confluence of self-disclosure, the volume of information exchanged, experience with the platform, and the degree to which the communicants empathize with each other.

We must also carefully examine how – and indeed, if – these hyperpersonal relationships emerge, and whether there is a link between volume of communication, the quality/quantity of self-disclosure, and increased intimacy. Neither the presence of self-disclosure nor increased communication alone causes increased intimacy within online relationships; self-disclosure is far from a simple input/output system. Researchers have found that the volume of self-disclosure, as well as the degree of positively-valenced self-disclosure affected feelings of connection and intimacy in social media environments (Park et al., 2011). The process of self-disclosure in online forums is reminiscent of the hyperpersonal feedback loop; one self-discloses online, which encourages another to do the same, which encourages the first person to do so again, and so forth. Both processes may be affected by the total amount or volume of communication exchanged. Although SPT predicts that this intimacy is a result of increased self-disclosure, it does not necessarily tie this to the ratio of self-disclosure utterances compared to the total amount of communication. It remains to be seen if this has a direct impact intimacy within a solely online relationship, and how that compares to an online/offline relationship.

This paper tests a model of the necessary and sufficient conditions for hyperpersonal relationships to emerge. In the model that I propose, the relationship between each communicant's self-disclosure is mediated by empathy, and all relationships are moderated by the volume of information exchanged. In other words, the relationship between both Communicant 1's self-disclosure to Communicant 2 is mediated by empathy. Furthermore, the volume of information exchanged positively moderates the mediated relationships.
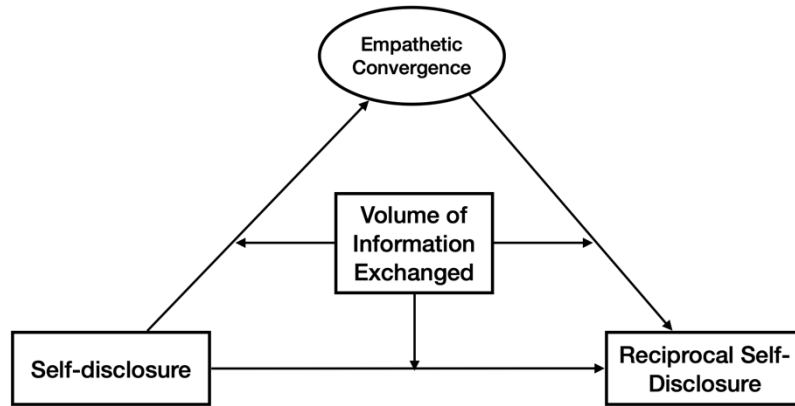
Figure 2: Hyperpersonal relationship emergence model

### *Intimacy versus self-disclosure*

The hyperpersonal model centers on the feelings of people as it relates to their conversational partners; people can sometimes feel more closely connected to and more emotionally intimate with their online conversational partners than their offline partners over time. Measuring how intimate people feel within an interpersonal relationship is most often accomplished via survey measures, such as by using the Emotional Self-Disclosure Scale, in which participates answer questions related to hypothetical self-disclosure events and what would/would not be disclosed, or by directly interviewing people about their relationship experiences (Locke & Gilbert, 1995; Snell et al., 1988). These methods have an appropriate time and place for use, but none can be easily adapted to examining interpersonal behavior in an impersonal way, making examining CMC in its naturally-occurring environments via large datasets a challenge. However, we can note that the disclosure-liking hypothesis notes that for intimacy to form, there must be self-disclosure present (Collins & Miller, 1994). Therefore, by measuring the amount of self-disclosure, we can infer a fruitful environment for emotional intimacy to grow. Measuring the number of self-disclosure acts over time in a large dataset allows researchers to both examine CMC in a naturalistic environment and to compare communicative dyads that use CMC channels at different rates.

To measure emotional self-disclosure (ESD) in a given text, one must measure the ratio of words related to ESD compared to the total number of words. Although this can be done manually, doing so would be quite onerous in large corpora. One common solution to this burden is using the LIWC – or the Linguistic Inquiry and Word Count – software package to analyze the

text (Pennebaker, Booth, et al., 2015). The software operates by comparing the words in a text against an array of dictionaries of various word types (e.g., function words, emotional valence words, verbs, etc.) and generating the percentages of each given type of word in the text as a whole. This software has been repeatedly tested and confirmed to have both internal and external validity as compared to manual coders (Pennebaker, Boyd, et al., 2015; Tausczik & Pennebaker, 2010). Although some studies for specialized purposes create their own customized dictionaries (see Jackson, 2019, as an example), because these texts are more generalized, this step was not deemed necessary. By analyzing texts from two people communicating, researchers have shown that a variety of language patterns are indicative of relationship status. For example, people in lower-quality relationships use more second-person pronouns ("you"), and the usage of the first person singular ("I") indicates a higher quality relationship (Tausczik & Pennebaker, 2010). Interestingly, first-person plural ("we") is not related to relationship quality.

For the purposes of this study, I define *emotional self-disclosure* as the ratio of sensitive disclosure words to all other words in a conversational turn. The "self-disclosure" construct consists of the eight categories that research has found to be indicative of emotional self-disclosure in online environments (Houghton & Joinson, 2012). Three of these categories consisted of process-word related (verbs, prepositions, and third-person pronouns), four were related to psychological categories (family, sexual, discrepancies, and negative emotions), and one was related to personal concerns (death)[1].
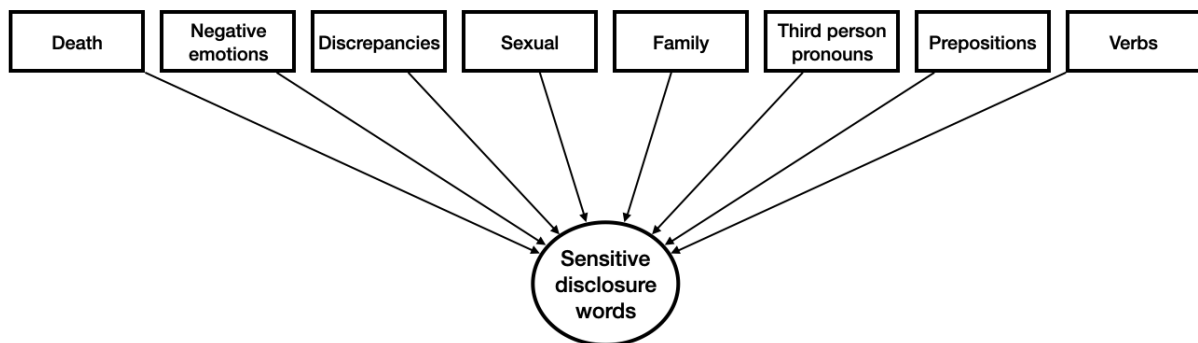


Figure 3. Illustration of factors contributing to the sensitive disclosure words construct.

---

[1] See Appendix A for an example list of words in each of these final eight categories.

### *Empathy through linguistic convergence*

The model proposes that we can further detect these relationships not simply through the presence of self-disclosure, but also through other attributes of intimate relationships. Previous research has shown that couples who form emotionally intimate relationships also have empathetic, converging language patterns. We understand this intuitively; couples have an array of references, phrases, and sounds that harken back to earlier points in their relationships but would be nonsensical to others, but the relationship goes deeper than simply private jokes, as research has shown that the structure of couples' language becomes more similar over time. Malloch and Taylor (2019) found that when one person engaged in emotional self-disclosure, it did not result in the other interactant self-disclosing. Instead, empathy, as measured through empathetic convergence, may mediate the relationship between the Communicant 1's ESD and Communicant 2's reciprocal ESD. In other words, when a person reads a message in which the message writer is being vulnerable, they are more likely to empathize and then share a similar experience.

Landaeuer and Dumais (1997) asked, "How do people know as much as they do with as little information as they get," and proposed *linguistic semantic analysis* (LSA) as a way to understand how language converges between two parties (p. 211). This method uses latent semantic analysis (Landauer et al., 1998) to compare two texts by "placing the words in each block in a high-dimensional semantic space, [and computing] the cosine of the angles between the two vectors to estimate the overall degree of latent semantic similarity" (Babcock et al., 2014, p. 79). As a result, words that are used in similar context have a higher score than those that are not. For example, the cosine of the words "communication" and "talking" would be higher than the words "communication" and "shark," indicating that the former pair is more often used in similar contexts to the latter pair. This analysis produces a score -1 (least similar) to +1 (most similar) that compares the two texts and has been shown to both correlate to the exchange of large amounts of information during a conversation and to relationship development similarly to LSM (Brinberg & Ram, 2021). LSA decreased when interactants had to repeat themselves, suggesting that higher LSA scores indicate that the conversants feel heard and understood. As a result, if a person felt heard and understood, presumably they would also be more willing to engage in ESD.

Some studies have also used linguistic style matching (LSM) to measure empathetic convergence, which is the degree to which the function words in two corpora of texts align (Ireland et al., 2011). LSM measures communication at the level of each pair of individuals communicating and uses the mirroring of function words to create a quantitative "measure of the degree to which two people…subtly match each others' speaking or writing style" (p. 39). After determining internal reliability (Cronbach's $\alpha$ = .61, $SD$ = 0.12) for the nine function word categories[2], they found that one's style matching tends to be a trait of the individual relationship as well as the situation, and that the LSM of couples in harmonious relationships tend to synchronize over time. This relationship also can be seen in the text-message conversations of people who are dating (Brinberg & Ram, 2021). Data also suggests that when one person sees their communication partner disclose emotions, it engenders empathy and mutual understanding (Malloch & Taylor, 2019). It is also possible that LSM is simply a response to a communication partner sharing strong emotions by simply reflecting the words back to the person who uttered them (Babcock et al., 2014). Therefore, it may not necessarily measure actual empathy.

Although linguistic style matching is a powerful measure, I chose to use linguistic semantic analysis to measure empathy in this study for three reasons. First, several of the function word categories within the LSM are duplicated within the measures for self-disclosure. As a result, the variables would not be distinct. Second, LSA provides a more robust measurement than LSM, as LSM is only the measure of function words, whereas LSA compares the language used more broadly. Finally, although LSA is predicted to have a similar mediation relationship as LSM, because it measures the text in a different way, it has the potential to reveal additional connections and processes by which relationships deepen.

### Testing the emergent model: The hyperpersonal loop

The presence of self-disclosure and empathy alone does not imply the feedback loop of intimacy within the hyperpersonal model. This feedback loop is created when an individual shares a sensitive detail about themselves and the recipient, idealizing the first communicant, assumes positive intentions. Thus, they should be more likely to engage in reciprocal self-

---

[2] In this study, the function word categories used were personal and impersonal pronouns, articles, conjunctions, prepositions, auxiliary verbs, common adverbs, negations, and quantifiers.

disclosure and the total amount of text exchanged over time would affect the willingness to self-disclose with one's conversational partner. Furthermore, as pairs exchange more information and get to know one another better, the degree to which they empathize with one another and relate to each other may increase. Therefore, the *volume of information exchanged* is predicted to moderate all paths in this model. This is defined as the cumulative number of words that the dyad has exchanged within the dataset.

### *Control variables*

Many variables that are traditionally controlled for online are not possible to determine with datasets that are collected from many online settings, as people have little incentive to be transparent about their age, gender, beliefs, and so forth. *Communicative competence* was first proposed by Chomsky and then expanded by Hymes to indicate the broad overview of language and cultural behaviors that a person must master to be accepted within a society and is particularly important to acknowledge (Wiemann & Backlund, 1980). Although generally conceptualized as an idea that applies most strongly to native speakers who grew up in that culture, it can be applied to any subculture where rules and norms are present and can be developed over time with practice, exposure, and study (Agbatogun, 2014). Since no human is born with CMC as their primary language modality, competency in a CMC environment is developed over time as the person uses and observes norms of the platform. Given that the users within the data sets often communicate with multiple other users, their communicative competence on the platform is likely to affect how efficiently and effectively they use the platform, as well as how often they use it. To measure communicative competence, I have defined it as *the number of elapsed days since the person sent their first message in the dataset.*

### *Hypotheses*

This study will test several central hypotheses with the overall goal of probing the relationship of the volume of information exchanged on amount of self-disclosure exchanged between two individuals engaging in computer-mediated communication. As such, this paper will test the following hypotheses:

34

Hypothesis 1: Empathetic convergence mediates the relationship between self-disclosure and reciprocal self-disclosure.

Hypothesis 2: The volume of information exchanged moderates the relationship between self-disclosure and reciprocal self-disclosure.

Hypothesis 3: The volume of information exchanged moderates the relationship between self-disclosure and empathetic convergence.

Hypothesis 4: The volume of information exchanged moderates the relationship between empathetic convergence and reciprocal self-disclosure.

Hypothesis 5: The volume of information exchanged moderates the indirect effect of initial self-disclosure on reciprocal self-disclosure through empathetic convergence.

# CHAPTER 3: METHODS

This chapter begins with an overview of the data collection challenges that make this sort of work difficult. Then, I provide a detailed description of the dataset being used in this study, including how it was acquired, ethical oversight for its use, and modifications made to enable analysis. I justify the use of the dataset within the greater context of data collection challenges that are present in this sort of project. Then, I describe how the data was analyzed. Finally, I discuss the preliminary findings and descriptive statistics.

## Data Acquisition

As previously mentioned, a significant gap in previous literature is that relatively little observes how people behave in "natural" computer-mediated environments (i.e., how they behave in unobserved online settings versus in experimental/survey settings). In part, this is due to the challenge of data collection from naturalistic settings. Much of this communication happens in areas of the Internet that are not easily accessible to researchers, such as private message boards, chat rooms, and text messages. Brinberg et al. (2021) suggest using mobile data collection to download – with permission – archives of communication data for analysis, but notes that this mechanism is difficult due to the ethical implications and the practical difficulty in anonymizing large amounts of data; she estimated that it took approximately 500 hours to de-identify her dataset of 1 million text messages from 41 couples. This approach was determined to be time- and cost-prohibitive for the purposes of this project.

### *DARPA BOLT dataset*

The DARPA BOLT (Broad Operational Language Translation) dataset is a public corpus that negates the need to manually collect data and is used in this paper because it gathered user-generated content within one-on-one conversations over extended periods of time. The BOLT program itself began as a way to improve natural language processing through user-generated content (Song et al., 2018). The Linguistic Data Consortium (LDC) worked with DARPA to gather more general data sources like text messages and chat dialogues, as well as evaluate the

content for machine-learning-specific tasks. This specific corpus was published in 2018 and contains both text message (SMS) and chat exchanges from 18,000+ conversations that were either donated by the users or collected by BOLT for this purpose (Chen et al., 2018). Although the data were collected via several different means, the data used were donated by the conversational participants.

Although this dataset has limitations, such as a lack of demographic data and only having larger amounts of data via single platform (i.e., the text messages), it still consisted of a robust post-hoc collection of conversations that made it suitable for this paper, since it gathered user-generated content within one-on-one conversations in a naturally-occurring environment. The users are all native speakers of English, and informed consent was obtained before donation. A second benefit to this dataset was that the data had already been "cleaned" of personally identifying information. Therefore, this study is classified as exempt from ongoing oversight by the Institutional Review Board under Category 4, which outlines use for secondary research. The use of this dataset allowed testing of the model in a simple and direct manner that can be easily adapted to future, more robust datasets that can examine other factors more deeply, such as native language, platform, demographic data, and so forth.

The data was initially received as a series of 18,438 XML files consisting of 1,073 files gathered through a BOLT chat server and 17,365 donated SMS conversations. Due to wanting consistency in the platform, I proceeded with only the donated SMS conversations and used Python to extract the relevant information into a spreadsheet consisting of 301,163 individual messages and 2,176 unique users. The data was then organized by the 4,417 unique conversational dyads. Several dyads were determined to be unsuitable for analysis as they were one person sending texts to another with no response; those were removed, leaving 4,337 total cases for analysis.

When working with large text-based corpora, it is important to consider how to pre-process the data before beginning analysis. For example, some methods require the stemming of words (i.e., changing "making" to "make") or removal of "stop" words (such as "for" and "the"). For this analysis, neither of these changes was made as the methodological approach did not necessitate it. All changes and transformations are discussed in the next section.

<center>**Operationalization of Variables**</center>

As discussed in the literature review, there are four primary variables of concern. I will provide a brief description of them, as well as discuss how they are specifically operationalized in this work and discuss their transformations as applicable. These variables are also summarized in Table 1.

### *Self-disclosure and reciprocal self-disclosure*

To analyze the corpus for self-disclosure, I first randomly assigned each individual from each dyad as either Person 1 or Person 2. In practical terms, this split the dataset so that half of the individuals were "senders" of self-disclosure (variable X, Self-Disclosure), and the other half were "respondents" of self-disclosure (variable Y, Reciprocal Self-Disclosure). Then, I used the text analysis software Linguistic Inquiry and Word Count (LIWC) to analyze each individual's amount of emotional self-disclosure. This software was first developed in 1993 and has been refined and updated multiple times in the intervening years, most recently in 2022. Due to its long life in the increasingly popular field of computer-aided textual analysis, it has been tested and validated in a variety of contexts.

LIWC's power comes from its internal dictionaries, which are updated with each version to reflect trends in meaning. These internal dictionaries allow users to compare a given word (or words) against a database of words, word-stems, and classifications, then produce a percentage per category to reflect what percentage of total words are in each category. For example, the first sentence of this paragraph consists of 11% pronouns (its), 22% prepositions (from, with), and 5.6% motion words (reflects). This short sentence merely indicates the mechanics; the full tool comes with over 100 dictionaries that classify thousands of individual words.

LIWC has several benefits as compared to human coders (Pennebaker et al., 2022). First, it can classify and define a text considerably faster than human coders possibly can; the full dataset consisting of 4000+ pairs of communicators contained 301,000,000 words and took less than a minute to process. There is some concern that a given word may be miscategorized in any given scenario, as often words are used in contexts to imply the opposite ("I'm mad at you," with a wink emoji, likely would mean the opposite), but these incorrect classifications are offset, as

<center>38</center>

writers tend to use additional language indicating their *actual* intent. In larger datasets, such as the one used in this study, the effects of incorrect classifications are negligible.

To create the self-disclosure construct, I used existing research to determine the eight word categories that were indicative of self-disclosure (Houghton & Joinson, 2012). Although LIWC returns the results separately for each category, calculating the overall construct is a simple matter of calculating the number of words present, then determining the percentage of total words. For example, the message "*I Just watched the most aweful movie ever. Doomsday machine. You can watch it on youtube. Its like they ran out of money so they just ended it. What the fuck.*" consists of 31 words, 41.93% of which fall into one of the self-disclosure categories. I did not run a spell-check; anyone who has seen a helpful spellchecker change words incorrectly has observed the false-positives that these tools tend to produce; methodologically, this is rarely a concern in corpora (Pennebaker et al., 2015). LIWC also automatically transforms text to lower case, and thus that step was not necessary for this analysis. Two variables were produced, one each for each individual dyad (Person 1 and Person 2), based on the total amount of text each pair sent to one another within the corpus.

### *Empathetic convergence*

As pairs of conversants engage with one another, they begin to develop a repository of in-jokes, slang, turns of phrases, and other linguistic peculiarities that are the hallmark of close interpersonal relationships. Moreover, even the word choices that they use and the patterns of words that emerge correlate to intimacy within relationships (Brinberg & Ram, 2021; Landauer et al., 1998). To measure this empathetic convergence, or the degree to which the language that the dyad used is similar, I used latent semantic analysis (LSA) to compare the texts generated by each dyad within the dataset. LSA[3] was calculated using the LSAfun package in R and a pre-existing semantic space, baroni, that was created through an analysis of the British National Corpus, a 2009 Wikipedia dump, and a corpus gathered from websites in the .uk domain (Baroni et al., 2014; Günther et al., 2014; R Core Team, 2018). The semantic space is 400 dimensions and categorizes 100,000 words, making a robust tool by which to analyze the data. LSAfun requires light pre-processing of the text consisting of removal of punctuation and converting all

---

[3] See Appendix B for the full command in R. Special gratitude to Fritz Günther for helping me with the specifics.

text to lowercase letters. Finally, the two texts from each dyad were compared, resulting in a single LSA calculation for each pair which was then exported for analysis in SPSS.

### *Volume of information exchanged*

Social information processing notes the importance of the volume of information exchanged. Specifically, CMC communicators take longer to exchange the same amount of information. Although the hyperpersonal model builds on SIP and does not directly address the volume of information, it is reasonable to predict that the volume of information exchanged would affect one's willingness to disclose personal information, and that only past a certain point of intimacy would one be willing to engage in more intimate conversation. As such, the volume of information is the total number of words a pair exchanged in this dataset. It is calculated by adding together the total number of words the sender sent and receiver sent.

### *Communicative competence*

In order to measure the level of communicative competence, I used the total volume of communication in the dataset is an additional covariate. Although this is not explicitly a measure of communicative competence, Lev-On (2015) found that the length of time an individual had spent on a platform affected feelings of insecurity, their level of participation, and their overall sense of being a part of the online community. Research also suggests that a certain element of communication competence is required to fully use technological affordances of a particular medium in order to achieve one's objectives; although that study specifically examined political expression online, a similar communication competence would likely influence one's ability to use a medium to achieve interpersonal objectives, such as endearing oneself to another, as well (Velasquez & Rojas, 2017). Within organizational settings, studies have found that one's tenure within the organizations affects the sense of belonging that is necessary to build trust and personal satisfaction since those emotional factors can only be developed over time (Haines, 2014). Since one's usage of a site or platform can influence both one's own ability to use the medium to communicate fully, as well as the emotional factors that may shape one's perception of others using the platform, *tenure on site* was developed as a stand-in covariate to account for communicative competence and was calculated by subtracting the user's first activity within the

dataset and subtracting that from the date of the last message sent to their conversational partner. For example, user 130622's first message in the BOLT dataset was sent July 1, 2012, and their last included message was sent on July 16, 2013. Therefore, the user's tenure when the last message was sent was 380 days.

Table 1. Conceptualization and operationalization of variables

| Variable (Label) | Conceptual definition | Operationalization |
|---|---|---|
| Emotional Self-disclosure (Person 1: P1_ESD; Person 2: P2_ESD) | How often an individual self-discloses personal information | Percentage derived from LIWC categorization of emotional self-disclosure language to total language. Both the variables Person 1 and Person 2 are calculated in the same way using their respective collected texts. |
| Empathetic Convergence (LSA) | How much people express their empathy with one another by using similar words and phrases | Measured by using latent semantic analysis (LSA) to produce a calculation of semantic similarity between each text on a continuous scale of 0-1, with 0 being least convergent and 1 being the most. |
| Volume of Information Exchanged (LN_VolXC) | How much two individuals talk to one another | The total wordcount exchanged between two individuals. |
| Communicative Competence (Person 1: P1_Tenre; Person 2: P2_Tenre) | How experienced an individual is with the communication medium. | The difference in time (measured in days) between the first message sent in the corpus (to any user) and the last message sent (to the particular other individual). |

**Preliminary Analysis**

The purpose for this study was to determine if there is any kind of increasing self-disclosure over time that could imply a fruitful ground for hyperpersonal relationships to emerge. Specifically, it sought to determine if the total volume of information exchanged between each pair played a moderating role in the indirect relationship between the reciprocal emotional self-disclosure, mediated by empathetic convergence. After defining the variables, the data was uploaded into SPSS version 28 and examined for unusual outliers. The only one found had an impossible tenure (specifically, it had calculated as 41,092 days, which would impossible, given that the Internet has not existed for 112 years) and was removed. Then, the descriptive statistics for each variable were examined to determine if the data met the fundamental assumptions of regression analysis by skewness and kurtosis. As shown in

Table 2Table 2, this examination showed that the volume of information exchanged variable was outside of the acceptable skewness and kurtosis ranges due to the large number of people exchanging relatively little information (<100 words). Thus, a natural-log transformation was applied to this variable and explored. Log transformations were found to produce both the most normal skewness and kurtosis measurements (.20 and -4.9, respectively), and most acceptable P-P plots (see Table 2, Figure 4, and Figure 5) . Thus, the log-transformed variable for volume of information exchanged was used. The Person 2 tenure variable was also skewed right due to many individuals with "0" tenure and thus causing zero-inflation within the dataset, implying that a large number ($n = 550$) of conversations only happened on a single day. I examined these individuals and found that these most often were the result of one person sending a text message to another person and having only a single day of interactions with them. Transformation of these variables did not significantly improve their skewness, kurtosis, or P-P plots, and thus were not used in analysis.

Table 2. Descriptive statistics of variables

| Variable | Mean | Median | St. Dev. | Variance | Skewness | Kurtosis |
|---|---|---|---|---|---|---|
| Person 1 ESD | 0.38 | 0.39 | 0.10 | 0.01 | -0.74 | 6.44 |
| Person 2 ESD | 0.37 | 0.38 | 0.10 | 0.01 | -0.42 | 7.15 |
| Empathetic Convergence | 0.83 | 0.91 | 0.19 | 0.04 | -1.97 | 3.76 |
| Volume of Information Exchanged | 700.56 | 190.00 | 1,395.95 | 1,948,686.92 | 3.98 | 19.33 |
| Volume of Information Exchanged (Natural Log Transformed) | 5.35 | 5.25 | 1.57 | 2.47 | 0.20 | -.49 |
| *Control Variables (Days)* | | | | | | |
| Person 1 Tenure | 296.11 | 218.00 | 256.91 | 66,002.95 | 1.23 | 1.20 |
| Person 2 Tenure | 117.24 | 39.00 | 180.47 | 32,568.00 | 2.29 | 5.90 |
| *Note: N = 2174* | | | | | | |
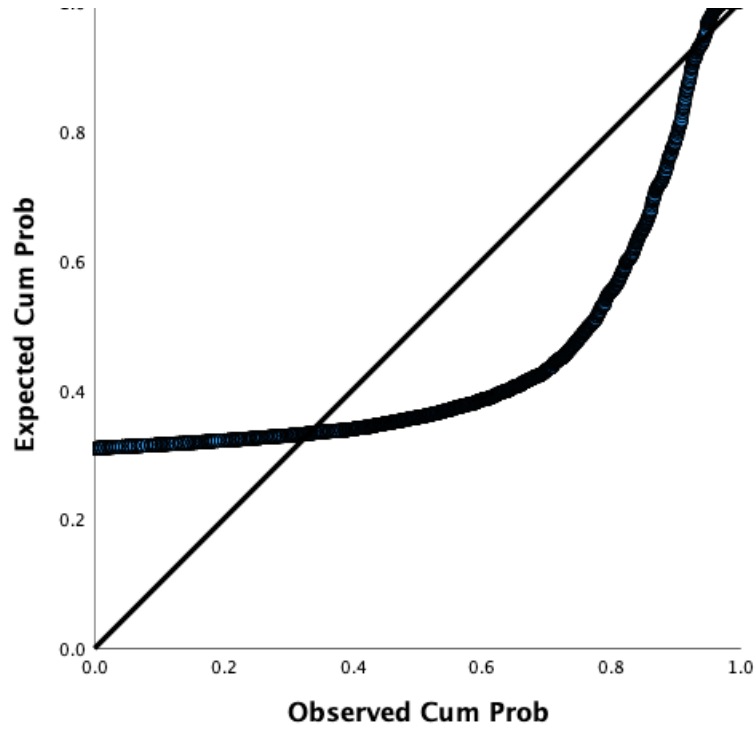
Figure 4. P-P plot of the total volume of information
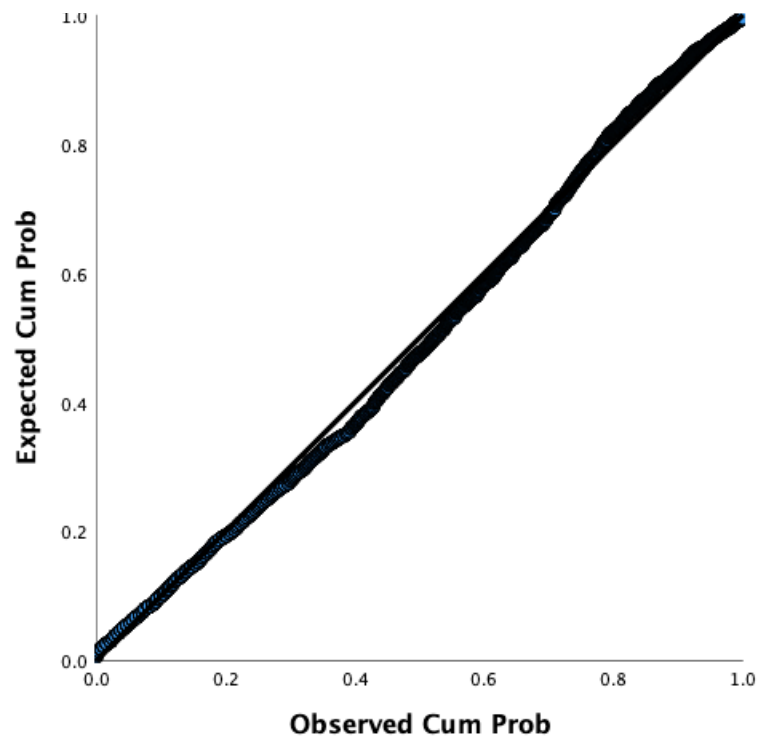


Figure 5. P-P plot of the natural-log transformed total volume of information

# CHAPTER 4: RESULTS

This chapter analyzes the results in light of the hypotheses proposed. To test the proposed model, I used Hayes's PROCESS macro (Models 4 and 59) to test the predicted relationships among variables, controlling for the tenure of the conversants, and the bootstrapping approach to more closely estimate the indirect effects (Hayes, 2018). Correlations for all study variables and covariates are presented in Table 3.

Table 3. Correlations of Variables

| Model Variables | Person 1 ESD (X) | Person 2 ESD (Y) | Empathetic Convergence (M) | Volume of Information Exchanged (W)^ | Person 1 Tenure | Person 2 Tenure |
|---|---|---|---|---|---|---|
| Person 1 ESD (X) | 1 | | | | | |
| Person 2 ESD (Y) | .19** | 1 | | | | |
| Empathetic Convergence (M) | .32** | .28** | 1 | | | |
| Volume of Information Exchanged^ | .22** | .21** | .78** | 1 | | |
| Covariates | | | | | | |
| Person 1 Tenure | .08** | .05 | .13** | .21** | 1 | |
| Person 2 Tenure | .10** | .09** | .34** | .51** | .58** | 1 |

*Note:* ** $p < .01$ ^ Log transformed

## Hypothesis 1: Testing for Mediation Effects

Hypothesis 1 tested that empathetic convergence would mediate the relationship between initial and reciprocal self-disclosure. I analyzed this through Model 4 of Hayes's PROCESS macro in SPSS with 5,000 bootstrap models to estimate the indirect effect as well as through the four linear models criteria set forth by Baron and Kenny (1986):

1. The independent variable must affect the mediator [Empathetic Convergence (M) ~ Person 1 ESD (X) (Direct effect)]

2. The independent variable must affect the dependent variable [Person 2 ESD (Y) ~ Person 1 ESD (X) (Direct effect)]

3. The mediator must affect the dependent variable [Person 2 ESD (Y)~ Person 1 ESD (X), Empathetic Convergence (M), LSA affects Person 2 ESD, accounting for Person 1 ESD (Direct effect)]

4. There must be a significant effect on the dependent variable for the indirect path from emotional self-disclosure to reciprocal self-disclosure through empathetic convergence. (Indirect effect)

The results of these tests are shown in Figure 6 and indicate that this model accounts for approximately 4% of the variation in the reciprocal self-disclosure ($R^2 = .04$). Emotional self-disclosure of the first party positively predicted empathetic convergence ($b = .56$, $t = 15.39**$, 95% CI [.49, .63], $p < 0.001$), which satisfies condition 1. Next, the emotional self-disclosure of the first party affects reciprocal self-disclosure ($b = .11$, $t = 4.87$, 95% CI [.14, .22], $t = 8.47$, $p < 0.001$), satisfying the second condition. Next, empathetic convergence significantly predicted reciprocal emotional self-disclosure ($b = .13$, $t = 10.80$, 95% CI [.11, .16], $p < 0.001$), satisfying the third condition. Finally, there is a small, but significant effect on the indirect path which satisfies the final condition and confirms that partial mediation is occurring ($b = .08$, 95% Bootstrap CI [.04, .11], $p < 0.001$). The effects for establishing mediation were satisfied, indicating support for H1.

As alluded to earlier, although the Baron and Kenny approach had been the standard for many years, its popularity is waning in favor of Preacher and Hayes's (2004) approach. Hayes (2018) notes that Baron and Kenny's steps were popular because they were simple to teach and understand, but this does not make the statistics correct. First, their approach fails to quantify the indirect effect, and instead makes inferences based on unrelated hypotheses. As a result, the conclusions are not tested directly (is there, or is there not an indirect relationship?), but instead are assumed (there's an indirect relationship if these three hypotheses are true). Furthermore, if these hypotheses are in error, one may incorrectly assume there is no mediation present. It also assumes that an indirect effect requires a significant total effect, which is often not true. Finally, the confidence intervals around the indirect effect are more informative than simply knowing

46

whether or not a variable mediates a relationship between the independent and dependent variables because it allows the quantification of effects versus a simple dichotomous yes or no. Baron and Kenny's approach makes it difficult for researchers to proceed with more complex methods as it discourages one from thinking quantitatively, which is required for understanding the nuanced relationships that are often investigated.

Instead, the better approach to establish mediation is to determine whether the indirect *a* x *b* path is significant using PROCESS, then looking at the direct and total effects output of those results. Hayes also suggests using a bootstrapping method, in which sub-samples of the dataset are taken with replacement and generates a new sample distribution which can then be analyzed. An additional benefit of bootstrapping is that it is more resilient to non-normality. Whereas the normal theory approach generally requires either a normal sample or transforming the sample, bootstrap samples allow for more accurate results when samples are not normal. Accordingly, bootstrapping was chosen in this paper due to both its statistical robustness, and the fact that many of the variables are not normal and were resistant to transformations designed to increase normality (i.e., natural-log transformations, square root transformations, etc.). In this paper, a bootstrap sample of 5,000 is consistently used throughout.

Accordingly, to establish mediation through Hayes's (2018) criteria, I used Model 4 of PROCESS to test the potential mediation and found that there was an indirect effect of initial self-disclosure on reciprocal self-disclosure through empathetic convergence (Indirect effect: *B* =.07, BootSE = .12, BootCI [.04, .11]). This analysis also provides support for hypothesis 1, which predicted a mediation relationship.
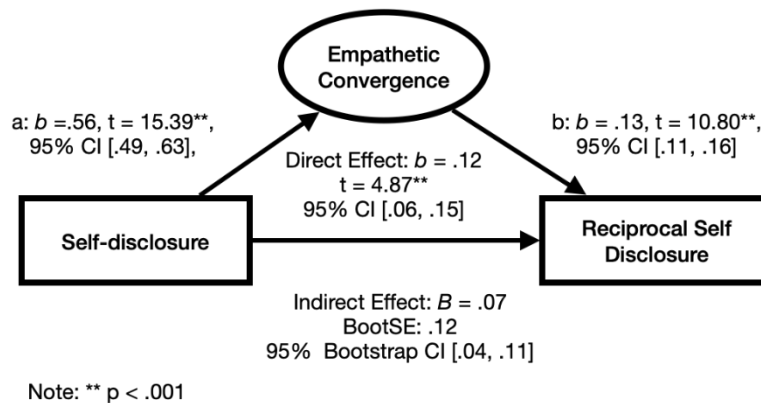


Figure 6. Visualization of the mediation effects predicted by hypothesis 1

47

**Hypotheses 2-4: Testing for Moderation Effects**

In Hypotheses 2-4, this study assumed that the volume of information exchanged would enhance the indirect relationship between initial emotional self-disclosure and reciprocal self-disclosures from their communicative partner. Broadly, moderated mediation would be established if either the path between initial self-disclosure and empathetic convergence or the path between empathetic convergence and reciprocal self-disclosure were moderated by the volume of information exchanged. Specifically, I hypothesized that the volume of information exchanged would positively moderate all three paths within the model: the relationships between self-disclosure and reciprocal self-disclosure (Hypothesis 2), between self-disclosure and empathetic convergence (Hypothesis 3), and between empathetic convergence and reciprocal self-disclosure (Hypothesis 4). To examine these hypotheses, I used Model 59 of the PROCESS macro, which allowed for the examination of any moderating effects of volume of information exchanged on the direct paths in the larger model while controlling for the effects of other variables in the model. The results of this test are shown in Table 5 and reported below.

First, I examined the moderation effects of the volume of information exchanged on the relationship between initial self-disclosure and reciprocal self-disclosure. The interaction is significant and positive ($b = .06$, $SE = .02$, $t = 3.29$, $p = .001$, CI [.02, .10]). Furthermore, as Figure 7 illustrates, examination of the conditional effects shows that, as

Hypothesis 2 predicted, higher volumes of information exchanged increasingly moderate the relationship; Johnson-Neyman output indicates that 95% of the regions are significant. In short, this analysis provides evidence for Hypothesis 2.

Figure 7. Moderation effect of volume of information exchanged on the relationship between initial self-disclosure and reciprocal self-disclosure

Next, I will examine the moderating effect of the volume of information exchanged on the relationship between initial self-disclosure and empathetic convergence, as predicted in H3. First, I investigated the direct paths involved in H3. The relationship between initial emotional self-disclosure and empathetic convergence is significant and positive ($b$ = .64, $SE$ =.08, $t$ = 8.42, $p$ < .001), as is the relationship between the volume of information exchanged and empathetic convergence ($b$ = .13, $SE$ = .01 $t$ = 16.79, $p$ < .001). Turning to the relationship predicted in H3, the interaction between initial self-disclosure and volume had small-but-significant negative effects upon empathetic convergence ($b$ = -.09, $SE$ = .02, $t$ = -4.68, $p$ < .001, CI [-.13, -.06]). This supports Hypothesis 3, which predicted that the volume of information exchanged over time would moderate the relationship between initial emotional self-disclosure and empathetic convergence. As Figure 8 shows, a simple test of slopes showed that lower volumes of information have significant positive effects on the relationship between self-disclosure and empathetic convergence (1 SD below [LnVLXC=3.78]: $b$ = .29, $SE$ = .03, $t$ = 11.25, $p$ < .001, 95% CI [.24, .34], but by the mid-range, the relationship is beginning to fade (Mean: [LnVLXC=5.35]: $b$ = .14, $SE$ = .04, $t$ = 3.14.25, $p$ = .002, 95% CI [.05, .22]), until it disappears

49

altogether in the high-range volumes (1 SD above [LnVLXC=6.92]: $b = -.01$, $SE = .07$, $t = -.15$, $p = .88$, 95% CI [-.15, .131]).
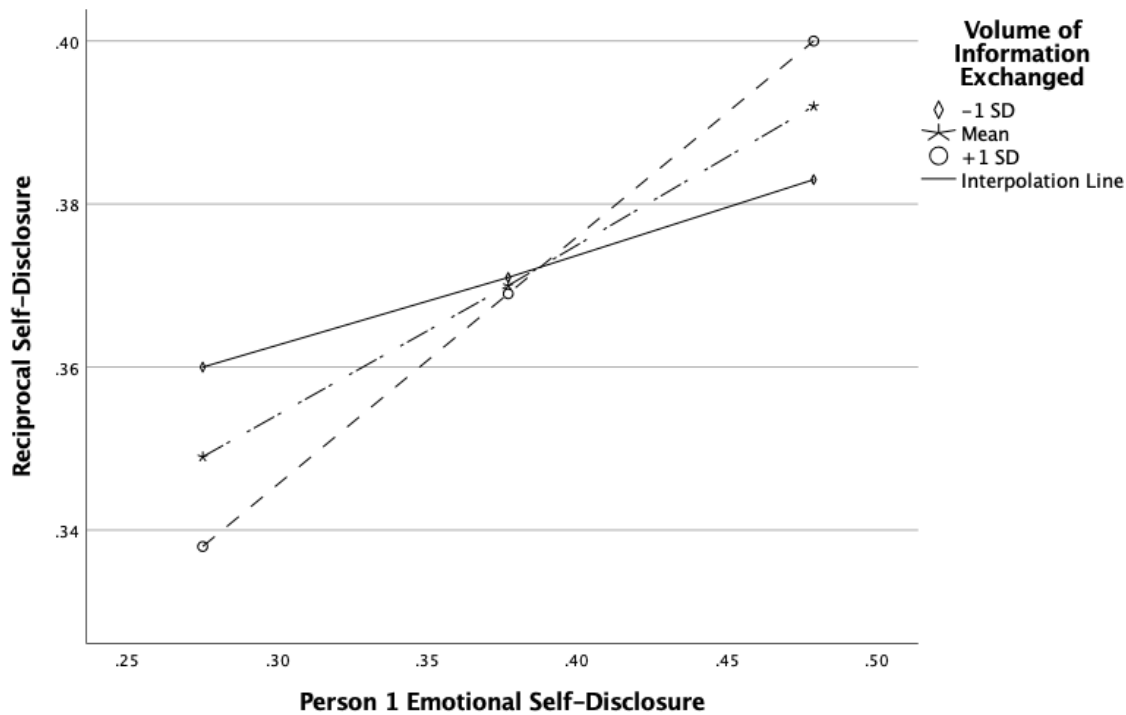
Table 4 details the Johnson-Neyman regions of significance.
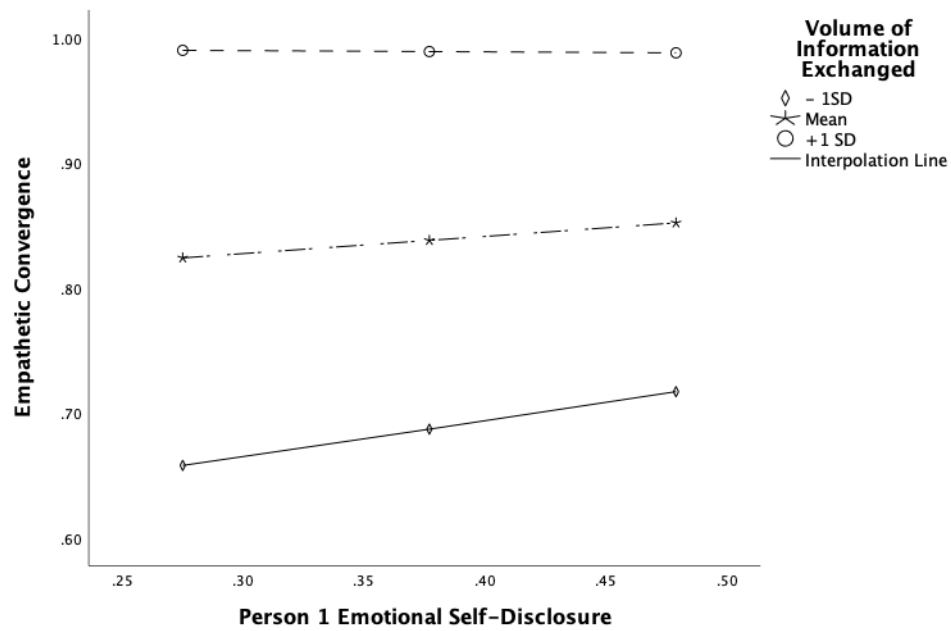


Figure 8. Moderation of volume of information exchanged on the relationship between initial self-disclosure and empathetic convergence.

Table 4. Johnson-Neyman output of the moderating effect of the relationship between initial self-disclosure and empathetic convergence.

| Volume of Information | b | SE | t | p | LLCI | ULCI |
|---|---|---|---|---|---|---|
| **1.39** | **0.51** | **0.05** | **10.09** | **0.00** | **0.41** | **0.61** |
| **1.81** | **0.47** | **0.04** | **10.85** | **0.00** | **0.39** | **0.56** |
| **2.23** | **0.43** | **0.04** | **11.72** | **0.00** | **0.36** | **0.50** |
| **2.65** | **0.39** | **0.03** | **12.58** | **0.00** | **0.33** | **0.45** |
| **3.07** | **0.35** | **0.03** | **13.06** | **0.00** | **0.30** | **0.41** |
| **3.49** | **0.31** | **0.03** | **12.47** | **0.00** | **0.26** | **0.36** |
| **3.91** | **0.27** | **0.03** | **10.52** | **0.00** | **0.22** | **0.32** |
| **4.33** | **0.23** | **0.03** | **7.94** | **0.00** | **0.18** | **0.29** |
| **4.75** | **0.19** | **0.04** | **5.60** | **0.00** | **0.13** | **0.26** |
| **5.17** | **0.15** | **0.04** | **3.76** | **0.00** | **0.07** | **0.23** |
| **5.59** | **0.11** | **0.05** | **2.39** | **0.02** | **0.02** | **0.21** |
| **5.75** | **0.10** | **0.05** | **1.96** | **0.05** | **0.00** | **0.20** |
| 6.01 | 0.08 | 0.06 | 1.35 | 0.18 | -0.03 | 0.18 |
| 6.43 | 0.04 | 0.06 | 0.55 | 0.58 | -0.09 | 0.16 |
| 6.85 | -0.01 | 0.07 | -0.07 | 0.95 | -0.14 | 0.13 |
| 7.27 | -0.04 | 0.08 | -0.56 | 0.57 | -0.20 | 0.11 |
| 7.69 | -0.08 | 0.09 | -0.97 | 0.33 | -0.26 | 0.09 |
| 8.11 | -0.12 | 0.10 | -1.301 | 0.19 | -0.31 | 0.06 |
| 8.53 | -0.16 | 0.10 | -1.582 | 0.11 | -0.37 | 0.04 |
| 8.95 | -0.20 | 0.11 | -1.821 | 0.07 | -0.42 | 0.02 |
| 9.24 | -0.23 | 0.12 | -1.961 | 0.05 | -0.46 | 0.00 |
| **9.37** | **-0.24** | **0.12** | **-2.026** | **0.04** | **-0.48** | **-0.01** |

*Note: The region of significant moderation is shown in bold. Moderator value defining the region of significance 5.57 >volume > 9.37 with 62.10% significant.*

Hypothesis 4 predicted that the volume of information exchanged would moderate the relationship between empathetic convergence and reciprocal self-disclosure. In examining the direct paths, the relationship between empathetic convergence and reciprocal self-disclosure is significant and positive ($b = .17$, $SE = .04$, $t = 4.17$, $p < .001$, CI [.09, .25]). However, the interaction of volume of information exchanged on the relationship between empathetic convergence reciprocal self-disclosure is not significant ($b = -.01$, $SE = .01$, $t = -.56$, $p = .58$, CI [-.03, .02]). Therefore, Hypothesis 4Hypothesis 4 is not supported. As Figure 9 illustrates, the practical differences in these relationships are negligible.
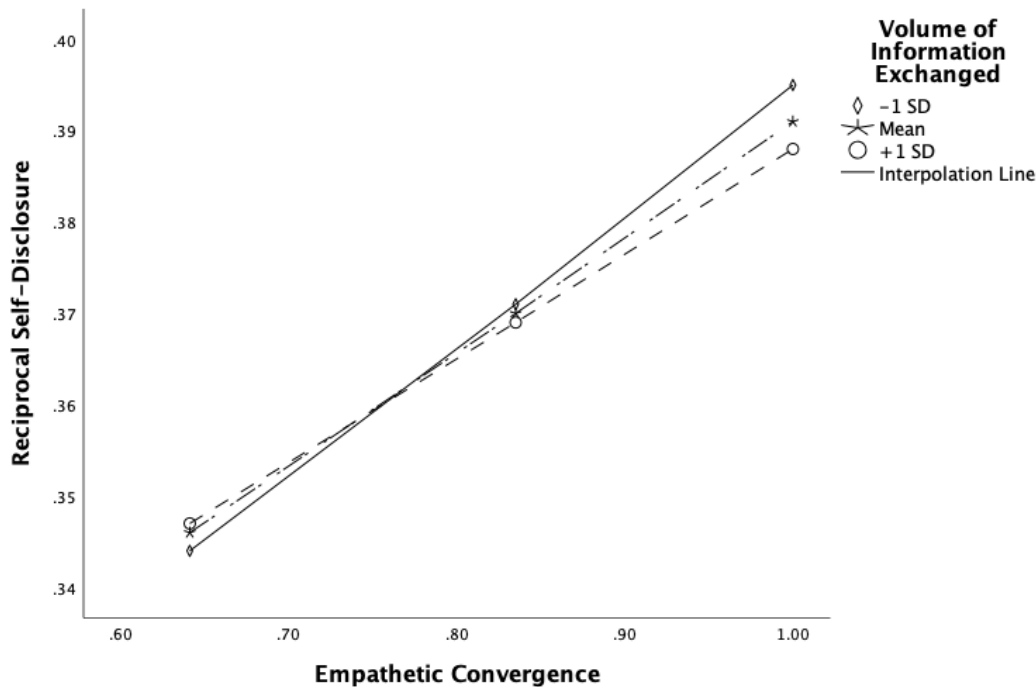


Figure 9. Moderation of volume of information exchanged on the relationship between empathetic convergence and reciprocal self-disclosure.

### Hypothesis 5: Testing for Moderated Mediation Effects

In Hypothesis 5, this study tested that the volume of information exchanged would enhance the indirect relationship between initial emotional self-disclosure and reciprocal self-disclosure through empathetic convergence by moderating the three paths between initial self-disclosure and reciprocal self-disclosure, initial self-disclosure and empathetic convergence, and empathetic convergence and reciprocal self-disclosure. To examine this, I used Model 59 of the PROCESS macro with 5,000 bootstrapped samples that provided estimates of the indirect

effects; if the confidence intervals within these estimates do not include 0, they are considered statistically significant.

Table 5 demonstrates a more nuanced view of the moderated mediation. Specifically, the effect size is small and only present at lower levels of information exchanged; significant mediation only emerged within pairs that had lower amounts of information exchanged (-1SD: $b$ = .04, $SE$ = .04, 95% bootstrapped CI [.01, .07]). Only in pairs with lower amounts of information exchanged was there greater reciprocal self-disclosure via empathetic convergence. The indirect effect was not significant in pairs that had an average or above average amount of conversation. Therefore, hypothesis 5 was partially supported.

Table 5. Results of the indirect path in the moderated-mediation model.

| | $b$ | $S.E.$ | Bootstrapped 95% CI | |
| --- | --- | --- | --- | --- |
| | | | LL | UL |
| Moderated-mediation: ESD ➔ Empathetic Convergence ➔ Reciprocal ESD (Bootstrapped estimates) | | | | |
| Lower amounts of information exchanged (–1 SD) | 0.04 | 0.01 | 0.01* | 0.07* |
| Average amounts of information exchanged | 0.02 | 0.01 | -0.002 | 0.04 |
| Above average amounts of information exchanged (+1 SD) | -0.001 | 0.02 | -0.04 | 0.03 |

*Note: * $p$ < .05. Confidence interval derived from 5,000 bootstrapped models. LL and UL = Lower and upper limits, respectively. This moderated mediation model was tested using model 59 of PROCESS v. 4.0 for SPSS.*

This chapter described the analyses used to test the four hypotheses and summarized the results of those analyses. Hypothesis 1 predicted that the relationship between self-disclosure and reciprocal self-disclosure would be mediated by the degree of empathetic convergence. This hypothesis was supported, and the results indicated that empathetic convergence mediates the relationship at a small, but statistically significant level. Hypotheses 2-4 predicted that the volume of information exchanged would moderate the relationships between self-disclosure and reciprocal self-disclosure, self-disclosure and empathetic convergence, and empathetic convergence and reciprocal self-disclosure, respectively. Hypotheses 2 and 3 were supported, whereas hypothesis 4 was not. Finally, hypothesis 5, which predicted that the volume of information exchanged would moderate the indirect effect of initial self-disclosure on reciprocal

self-disclosure through empathetic convergence was partially supported, as the indirect effects were significant only at lower levels of volume exchanged (-1 SD). The next chapter will discuss these findings and the implications of the results.

# CHAPTER 5: DISCUSSION

The purpose of this paper was to examine the variables that influence reciprocal self-disclosure in a computer-mediated environment and may lead to hyperpersonal communication due to an increasingly emotionally connected feedback spiral. Specifically, the study tested a model that proposed that the initial self-disclosure would result in increased reciprocal self-disclosure, when mediated by empathetic convergence, and that all aspects of this relationship were moderated by the volume of information exchanged between the two individuals. This chapter examines the results of the five hypotheses, discusses the implications of those results, and suggests modifications to the hyperpersonal model in light of those results. The limitations of this project are also reviewed, as well as suggested future research areas to address these limitations.

## Summary of Findings

The goal of this research was to examine the mechanisms by which self-disclosure increases in computer-mediated relationships. To meet this goal, I defined emotional self-disclosure as a percentage of the conversation that contained specific words (i.e., words related to death, negative emotions, discrepancies, sex, family, third person pronouns, prepositions, and verbs), as these words correlate to the presence of emotional self-disclosure (Houghton & Joinson, 2012). Based on research linking self-disclosure and greater intimacy and the hyperpersonal model that predicts a spiral of increasingly intimate feelings created through selective self-presentation of oneself in an online setting, I predicted that when an initial conversant engaged in emotional self-disclosure, the spiral would appear as an increased amount of reciprocal self-disclosure, and that this relationship would be moderated based on the volume of information exchanged over time (Jiang et al., 2011; Walther, 1996). In other words, as a dyad exchanges more information, when one person discloses something personal, the other is more likely to reciprocate. Furthermore, since research shows that self-disclosure creates empathy, I also predicted that empathetic convergence, as measured by linguistic alignment, would mediate the relationship between initial and reciprocal self-disclosure (Malloch & Taylor, 2019). As demonstrated in the previous chapter, the primary findings from these analyses show that

56

empathy and the volume of information exchanged play mediating and moderating roles, respectively, in the relationship between initial and reciprocal self-disclosure. Specifically, empathy is a partial mediator in the relationship between initial and reciprocal self-disclosure, meaning that although self-disclosure does cause some amount of reciprocal self-disclosure, initial self-disclosure also works through empathy to cause reciprocal self-disclosure; a person engaging in self-disclosure makes use of the empathetic alignment co-created within the relationship, and this results in a fertile field for reciprocal self-disclosure.

The volume of information exchanged plays an important role in emotional self-disclosure. First, the volume of information moderates the relationship between initial and reciprocal self-disclosure, as the volume of information exchanged in relationships increases the amount of reciprocal self-disclosure transmitted. Importantly, this relationship only appears once a certain volume is reached; relatively low amounts of self-disclosure do not inspire reciprocal self-disclosure. This is tantamount to an unrequited affection; no matter how much emotional self-disclosure a person attempts to woo a person with, if their relationship has not passed a certain base level, the object of affection is unlikely to respond in kind. After that point, higher volumes of information exchanged do result in higher levels of reciprocal self-disclosure. It is possible there is an upper limit to this effect, but it was not found within this dataset. This supports the spiraling effect of these factors within the hyperpersonal model since the model predicts increasing intimacy and willingness to engage in reciprocal emotional self-disclosure as the total of volume exchanged within a relationship increases.

### *Volume of information exchanged and its effect on empathetic convergence*

The volume of information exchanged does significantly affect the relationship between initial self-disclosure and empathetic convergence. At lower and mid-levels (-1SD and mean), it positively affects empathetic convergence; pairs who have exchanged less information see increases in empathetic alignment when engaging in self-disclosure. However, at higher volumes of exchanged information (+1SD), there is not a significant relationship; increased self-disclosure has no effect on empathetic alignment. An additional interesting item of post-hoc analysis is that at the *very* highest levels (LnVLXC = 9.37), the relationship becomes significant again, but shifts from a positive relationship to a negative relationship. Thus, when a pair communicates *huge* amounts via CMC, increased self-disclosure leads to a pair falling out of

empathetic sync. Although this data is based on a very small subset of the overall dataset examined (specifically five pairs of participants), this result suggests several interesting possibilities that may point to a worthy future study of "super-communicators."

First, when viewed through the lens of Social Penetration Theory, this decoupling could be an indicator of a failing relationship in line with the "dissolution" stage predicted by Taylor and Altman (1987). Hall and Baym (2012) note that within close friendships, using text messages for relational maintenance can produce a plethora of negative consequences and feelings resulting from the increased contact within an omnipresent medium. The pressure of being "always on call" may result in weakening relationships and the resultant decreases within empathetic alignment. The confluence of being a "super-communicator" in a CMC medium and using that as a primary platform to conduct relational maintenance may have a deleterious effect on the relationship and push it towards dissolution. In the case of romantic relationships, the language that dyads use has been consistently shown to be able to predict future dissolution (Gottman & Krokoff, 1989; Kanter et al., 2022). Within CMC-specific contexts, linguistic patterns that emerge from individuals writing about their relationships on social media can predict relationship dissolution up to three months in advance of the actual breakup (Seraj et al., 2021). Although that research did not specifically address whether these patterns exist within interpersonal CMC, such as that which happens via text message, it is reasonable to think that if FtF interpersonal speech patterns can predict dissolution, and CMC text patterns can predict dissolution, that CMC interpersonal text patterns would also be predictive.

Conversely, it is also possible, given that this divergence happens in most voluminous of relationships, it could correlate to the most successful of partnerships in the "stable exchange" stage of a relationship (Carpenter & Greene, 2016). Although this stage is predicted to have both a large breadth and depth of self-disclosure, those elements may be occurring in other communication channels, such as FtF. Therefore, as the volume increases, the proportion of self-disclosure and empathetically aligned conversation that happened earlier in the relationship may be outweighed by the amount of perfunctory, task-based communication that happens later in the relationship. This idea is further supported by research that shows that there is an upper limit of empathetic alignment which plateaus as relationships mature (Brinberg & Ram, 2021). The specific communication patterns of super-communicators should be analyzed in more detail in future research.

An important element of that is the idea that, as relationships mature, pairs communicate differently in their online interactions than they did earlier in their relationship. This would have dramatic effects on the amount and frequency of self-disclosure that a pair is willing to engage in via CMC. Research has shown that the way that a person presents themselves in an online-only context differs from how they present themselves offline, in line with the hyperpersonal model (Gibbs et al., 2006). Thus, if a pair begins their communication online before transitioning to an offline environment, they may be *less* inclined to engage in empathetic alignment since they no longer factor in selective self-presentation. This may be at odds with the idea of a plateau of empathetic alignment as that research did not indicate that increased time resulted in a reduction of alignment, but since that study did not specifically account for the volume of information exchanged, merely the time in which it was exchanged, further investigation is needed to probe these specific effects.

The idea that a CMC-based relationship would become multimodal is not new and is centered in media multiplexity theory (MMT, Haythornthwaite, 2005). Although SIP and the hyperpersonal model see time and volume as the most important explanatory elements of the depth of an online relationship, MMT draws on the relationship between Granovetter's (1973) conceptualization of strong/weak interpersonal ties and how the number of channels (FtF, CMC, video call, etc.) affects the relationship. It proposes that as the strength of ties strengthens, pairs will diversify the number of individual channels they use, and as intimacy (and subsequent self-disclosure) grows, the way that pairs use a particular communication channel changes and transforms as well. MMT neither integrates a specific communication level at which a pair moves to another channel nor does it address the ESD feedback spiral as the hyperpersonal model does. It is possible that the empathy drop-off that the results show is due to relationships reaching a certain point and the pair subsequently diversifying their channel usage; more investigation would be necessary to draw these conclusions.

Although MMT may explain the negative moderation of volume of information exchanged upon the relationship between initial self-disclosure and empathetic alignment, it does less to explain the *positive* moderation of the volume of information exchanged on the relationship between self-disclosure and reciprocal self-disclosure. This more closely is represented by the hyperpersonal model, since it predicts the development of an ever-deepening intimacy. Future research should attempt to account for the nature of the relationship as well. It is

possible that, in line with MMT, the nature of the relationship is influential as well. Although MMT predicts a divide between active relationships being weak (acquaintances, coworkers in another department, etc.) and strong (friends, romantic partners, coworkers), the strong ties may behave differently depending on the specific nature of the relationship. For example, if the tie is platonic, it may be more likely to stay purely on the single channel, so while the degree of overall empathetic convergence remains at a platonic level, the self-disclosure spiral still can surpass its FtF-only counterpart. If the strong tie is romantic in nature, then the pair moves to additional channels, resulting in the disruption of the hyperpersonal spiral. It is also possible that the different types of empathy are a factor, or that different types of strong tie relationships rely more heavily on a particular empathetic style, thus influencing the results (Bachelor, 1988). In order to account for this possibility, additional research should be conducted that measures empathetic convergence by accounting for the specific relationships by tie strength and relationship label.

It is likely that there is a middle ground which incorporates understanding of the multiple communication channels from MMT with the influence of the volume of information exchanged over time and positive feedback loops derived from the hyperpersonal model. This theory could be used to explain and predict various phenomenon within CMC environments. One potential application for this merger could be uncovering the specific points at which people using online dating platforms decide to leave the platform, meet in person, etc. It could also have implications for the prevention of and intervention in technology-facilitated violence, such as doxing, child grooming, and cyberstalking, since those crimes all require three elements: an initial online meeting, the potential criminal believing they have connection (or deliberately creating connection) to the potential victim, then making use of other channels to inflict harm (Bailey et al., 2021; Lorenzo-Dus et al., 2016). Future investigations should attempt to merge MMT and the hyperpersonal model into a unified hyperpersonal multiplexity theory and probe the boundaries and situations in which the resultant theory adequately explains the observed phenomena.

## Limitations

There are three primary limitations in this study that should be addressed in future research.

The first and most significant limitation is that although the hyperpersonal model was built to explain relationships that begin and are experienced predominantly via a text-only CMC setting, the dataset used in this model was not exclusively that. Instead, the dataset contained conversations from people who had both CMC and FtF interactions. As such, intimacy – much less "hyperpersonal" intimacy – cannot be assumed to form in the same ways as Walther predicted within a solely/majority CMC setting. The reason for this discrepancy is due to the lack of availability of a dataset that contains solely CMC between two individual parties. Many CMC datasets exist that are similar to the one used in this project, in that the communicants have both CMC and FtF interactions, or they happen in the context of small groups, such as within chat rooms, Discord servers, etc. Although those datasets likely have fascinating interactions that can and should be examined, they necessarily have many confounding variables that would make the analysis of dyads challenging. For example, within a chat room, it can be difficult to discern a dialogue between two individuals. Some attempts have been made to use machine learning to parse these small group conversations into a turn-based dyadic format (see Lowe et al., 2015, for one approach), but these one-on-one conversations are largely not available for public analysis.

There are several potential solutions to this. First, researchers could look towards donation and simply recruit individuals to upload private conversations as the developers of the BOLT dataset did. A second – and more ethically dubious – approach would be partnering with the hosts of these logs (e.g., Discord, Signal, Slack, Telegram, etc.) to download and use their corpora. Although this is often allowed and permitted through the terms of service of those services, it cannot be claimed to have informed consent (Verma, 2014). Researchers could also partner with an organization that uses one of these communication tools and work an informed consent strategy into their onboarding. All these approaches would require de-identification of messages, which is a time-intensive task, or the use of an NIH Certificate of Confidentiality to meet IRB ethics requirements.

The second potential limitation in this study is that the dataset is homogenous, consisting of native English-speakers who voluntarily donated their data. This presents several possible issues. First, conversations did not necessarily begin at the "beginning" of a relationship, thus creating a possible truncation of crucial relationship-building interactions; future studies should capture the entirety of a pair's relationship, or explicitly compare the patterns found in full relationships versus those that have been abbreviated. Next, the willingness to engage in self-

disclosure and the formation of hyperpersonal relationships that surpass their FtF counterparts in terms of feelings of closeness and intimacy is intimately tied to one's culture. Specifically, being a member of a collectivist versus individualist culture affects the degree to which and circumstances under which a person will disclose personal information (Liu & Wang, 2018). When self-disclosure does take place, the cultural differences seem to largely disappear; both members of collectivist and individualist cultures report more liking, commitment, and willingness to tailor their communication when they experience increased self-disclosure (Yum & Hara, 2005). Furthermore, gender, disclosure valence (positive/negative), and specific platform affect the degree to which certain sentiments and disclosure are considered socially appropriate (Waterloo et al., 2018). For example, on WhatsApp (a popular instant message service that supports both group and dyadic chat) males are less likely to disclose any emotions than females, but there was no difference in negatively-valenced emotions (e.g., sadness, anger, etc.) on Twitter or Instagram. Still other research has found that gender does not play a role in online self-disclosure at all (Kim & Dindia, 2011). Since there is debate on this issue, future studies should consider including gender as a covariate and should be cautious when applying the present study's conclusions to a heterogenous population.

Finally, the dataset used in this study was collected from mid-2013 to mid-2015. The users in this dataset were operating in a post-smartphone, post-social media, pre-pandemic CMC milieu. Research conducted post-pandemic suggests that forced isolation and quarantine affected the ways that individuals use CMC to relate to others (Maheux et al., 2021). It remains to be seen if this change is permanent or temporary, but future research should take note of the timeframe (i.e., pre- or post-pandemic) in which the dataset was gathered to determine the long-term effect (if any) more clearly on communication competence. Emerging adults, whose communication and relational patterns were still developing during the pandemic, may show particular differences in terms of whether increased CMC affects their interpersonal relationships.

Finally, it is important to note that although the results in this study may be statistically significant, the effect sizes are relatively small. Although the volume of information exchanged does moderate the mediation relationship of initial self-disclosure working through empathy to cause reciprocal self-disclosure, the effects are subtle. Big data's ability to detect small effect sizes within large samples is well-known, but still presents ethical dilemmas related to both publishing this data and in suggesting implications and areas for future research (Khalilzadeh &

Tasci, 2017). To address these concerns and heed Khalilzadeh and Tasci's (2017) advice to "report all available measures to allow readers to obtain a full range of information so that they can form their own interpretations of scientific studies conducted with large samples or big data" (p. 96), the full results from SPSS used within this paper and the variables extracted from the dataset are available via the Purdue University Research Repository.

## The Future of the Hyperpersonal Model

Although this research was designed to test the key tenet of the hyperpersonal model that the volume of information exchanged between two people over time affects the emotional self-disclosure they engage in, some contextualization of that model should be given. When it was first developed, less than 37% of American households had a computer at home and a mere 18% had Internet access; smartphones would not be invented for another decade (US Census, 2010). As such, the average person's relationship with computer-mediated communication mediums were fundamentally different in that they *did not exist*. Instead, a relatively small portion of people used a primarily text-based medium for communicating with a similar demographic. In short, geeks chatted with geeks, and the hyperpersonal model was built on the idea of CMC as the exclusive communication medium.

In the decades since, the userbase and usage patterns have transformed dramatically; nearly all Americans have access to a computer, smartphone, tablet, or other Internet-connected device, and even in the Global South, access to the Internet and mobile communication technologies is far from uncommon (Chen, 2021; Martin, 2021). Furthermore, higher processor speeds and increased bandwidth have changed the nature of our online communication from purely text-based to a multidimensional relationship based on a combination of text, video, symbolic imagery (i.e., emojis and emoticons), and audio. Even if a pair initially meets via a text-only communication channel, they are often driven to diversify their relationship through other channels, even culminating in FtF interaction (Pettersen, 2016). Our online social lives are now comprehensive, robust, and richly textured. This research *only* examined CMC dialogues in a text-message based medium. Therefore, future research should probe additional CMC-based communication channels where people dialogue, such as online forums (e.g., Reddit, 4Chan, Telegram, etc.) and social media sites (e.g., Facebook, Twitter, SnapChat, etc.), as well as FtF settings in a variety of contexts. Existing research has shown that people use their

communication and information channels in different ways that are particulate to the channel; they use Twitter in a fundamentally different way than they use Facebook. Therefore, it is reasonable to expect that the interpersonal communication they conduct on these platforms may also vary.

Regardless of the disparate channels used in CMC-based relationships, qualitative evidence still suggests that text-based online communication adds a layer of depth that does not exist offline. Within this study, Hypothesis 2 represents quantitative evidence that the volume of information influences the amount of emotional self-disclosure they are willing to engage in. Future studies should examine these patterns in more detail to learn the specific reasons that modern communicators choose a text-only medium, and the situations in which they choose to use different mediums. Potential studies could explore if there is a volume of communication at which dyads add additional channels to their relationship, or if they prefer to engage in emotionally intense conversations (i.e., those with more ESD present) in different mediums. If so, there may be a triggering volume of disclosure, after which there are diminishing returns using the CMC medium. This could have implications for content moderation and creating intersessions in online hostility, such as the idea of a certain level of incivility triggering a time-out so that parties can calm down. It could also imply a switch with regards to crime crossing from CMC to FtF; if a person no longer feels like arguing online is sufficient, they may be more likely to engage in offline violence. It is important to note that this research did not have any comparison to FtF interactions, so these arenas are highly speculative and should be explored in future research.

Future research should also explore the time element. Although this study used each individual dyad as the unit of analysis, the temporal dimension is another element to explore. The research of Brinberg et al. (2021) used the individual dyad as its unit of analysis, and points to a rapid rise of empathetic convergence as a dyad continues to communicate, followed by a leveling-out as the pair grows more comfortable within their relationship. However, that research does not touch specifically on the amount of self-disclosure that those dyads engage in, nor does it examine any potential causal relationship between the self-disclosure acts of each specific party. If the same trends uncovered in this study hold true in a per-day analysis, the hyperpersonal model does not incorporate an emotional disclosure equilibrium and instead implies that a pair will be caught in an endless feedback loop of increasingly intense intimacy.

This equilibrium should be factored into the hyperpersonal model for it to remain reflective of current research trends.

Next, as natural language processing and machine learning become more robust, researchers should consider implementing them in the search for the answers to the above questions. Self-reported survey data often has issues with reliability and validity due to participants incorrectly reporting time spent on tasks. Therefore, the ability to view how people use platforms when unobserved (or when they forget that observation is possible) is critical to being able to uncover truths.

Finally, in the interest of equity, consideration must be given to dataset access. Although there are a great many datasets available for researchers to use, high-quality datasets, such as the one used in this paper, are often trapped behind paywalls or require expensive memberships to access. These financial access barriers prevent access to all but the most well-funded researchers at the most robustly financed institutions. Although datasets often are not functional by default, requiring much post-processing to be useful, once that post-processing is complete, additional development is generally unnecessary (Weller & Kinder-Kurlanda, 2015). Researchers compiling these datasets should commit to releasing the datasets, once de-identified, under an open source or Creative Commons license. Another potential solution is approaching it as temporarily embargoed data; once a set period has elapsed, the data is released to the general public. In short, requiring payment to access that was collected via publicly funded projects should be broadly reconsidered and individually eschewed.

## Conclusion

As the Internet develops and grows in its ubiquity, understanding how relationships develop and are maintained remains a critical element of understanding a core aspect of the average person's life. Although the hyperpersonal model has existed in literature for over a quarter century, there has been a dearth of quantitative evidence of the increasingly intimate feedback loops it predicts, and scholars have called for a consistently defined and easily replicable model that can be used to this end (Kim & Dindia, 2011). This study addresses that concern and provides a model that can be used with other datasets to examine hyperpersonal relationships more closely. I also interrogated the hyperpersonal model and demonstrated how the indirect relationship between emotional self-disclosure and reciprocal self-disclosure through

empathetic convergence is moderated by the volume of information exchanged over time. Additionally, I advanced an idea of how future research can merge media multiplexity theory with the hyperpersonal model to form a hyperpersonal multiplexity theory that can be used to understand the points at which people move to additional channels based on the level of emotional self-disclosure they find themselves engaging in. Through understanding this relationship more thoroughly, we can better conceptualize the different stages of friendship and relationships when they have an online element and build processes and products that will encourage those relationships to form connections. Furthermore, although the hyperpersonal model fundamentally examines friendship and feedback loops of positive interaction, perhaps one day, by providing a model that will allow us to know at what point intercession can deter a relationship that would otherwise be harmful or destructive, it can also give us hope for the future.

# REFERENCES

Agbatogun, A. O. (2014). Developing learners' second language communicative competence through active learning: Clickers or communicative approach? *Educational Technology and Society*, *17*(2), 257–269.

Altman, I., & Taylor, D. A. (1973). *Social penetration: The development of interpersonal relationships*. Holt, Rinehart & Winston.

Altman, I., Vinsel, A., & Brown, B. B. (1981). Dialectic conceptions in social psychology: An application to social penetration and privacy regulation. *Advances in Experimental Social Psychology*, *14*, 107–160. https://doi.org/10.1016/S0065-2601(08)60371-8

Anderson, T. L., & Emmers-Sommer, T. M. (2006). Predictors of relationship satisfaction in online romantic relationships. *Communication Studies*, *57*(2), 153–172. https://doi.org/10.1080/10510970600666834

Antheunis, M. L., Valkenburg, P. M., & Peter, J. (2007). Computer-mediated communication and interpersonal attraction: An experimental test of two explanatory hypotheses. *CyberPsychology & Behavior*, *10*(6), 831–836. https://doi.org/10.1089/cpb.2007.9945

Attrill, A., & Jalil, R. (2011). Revealing only the superficial me: Exploring categorical self-disclosure online. *Computers in Human Behavior*, *27*(5). https://doi.org/10.1016/j.chb.2011.02.001

Babcock, M. J., Ta, V. P., & Ickes, W. (2014). Latent semantic similarity and language style matching in initial dyadic interactions. *Journal of Language and Social Psychology*, *33*(1), 78–88. https://doi.org/10.1177/0261927X13499331

Bachelor, A. (1988). How clients perceive therapist empathy: A content analysis of "received" empathy. *Psychotherapy: Theory, Research, Practice, Training*, *25*(2), 227–240. https://doi.org/10.1037/h0085337

Bailey, J., Flynn, A., & Henry, N. (2021). *The Emerald International handbook of technology-facilitated violence and abuse*. Emerald Publishing Limited. https://doi.org/10.1108/9781839828485

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6). https://doi.org/10.1037//0022-3514.51.6.1173

Baroni, M., Dinu, G., & Kruszewski, G. (2014). Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 238–247. https://doi.org/10.3115/v1/P14-1023

Bay, M. (2009). *Homo conexus: The connected humans*. Gyldendal Business.

Brinberg, M., & Ram, N. (2021). Do new romantic couples use more similar language over time? Evidence from intensive longitudinal text messages. *Journal of Communication*, *71*(3), 454–477. https://doi.org/10.1093/joc/jqab012

Brinberg, M., Vanderbilt, R. R., Solomon, D. H., Brinberg, D., & Ram, N. (2021). Using technology to unobtrusively observe relationship development. *Journal of Social and Personal Relationships*, 1–22. https://doi.org/10.1177/02654075211028654

Burgoon, J. K., & Hale, J. L. (1987). Validation and measurement of the fundamental themes of relational communication. *Communication Monographs*, *54*(1), 19–41. https://doi.org/10.1080/03637758709390214

Byron, K. (2008). Carrying too heavy a load? The communication and miscommunication of emotion by email. *Academy of Management Review*, *33*(2), 309–327. https://doi.org/10.5465/AMR.2008.31193163

Carlson, J. R., & Zmud, R. W. (1999). Channel expansion theory and the experiential nature of media richness perceptions. *Academy of Management Journal*, *42*(2). https://doi.org/10.2307/257090

Carpenter, A., & Greene, K. (2016). Social Penetration Theory. In C. R. Berger & M. E. Roloff (Eds.), *The International Encyclopedia of Interpersonal Communication* (pp. 1–5). Wiley. https://doi.org/10.1002/9781118540190.wbeic0160

Chan, D. K.-S., & Cheng, G. H.-L. (2004). A comparison of offline and online friendship qualities at different stages of relationship development. *Journal of Social and Personal Relationships*, *21*(3), 305–320. https://doi.org/10.1177/0265407504042834

Chen, R. (2021). *A demand-side view of mobile Internet adoption in the Global South*. https://openknowledge.worldbank.org/bitstream/handle/10986/35302/A-Demand-Side-View-of-Mobile-Internet-Adoption-in-the-Global-South.pdf

Chen, S., Fore, D., Strassel, S., Lee, H., & Wright, J. (2018). *BOLT English SMS/chat*. Linguistic Data Consortium. https://doi.org/10.35111/hkfc-7865

Coleman, L. H., Paternite, C. E., & Sherman, R. C. (1999). A reexamination of deindividuation in synchronous computer-mediated communication. *Computers in Human Behavior*, *15*(1), 51–65. https://doi.org/10.1016/S0747-5632(98)00032-6

Collins, N. L., & Miller, L. C. (1994). Self-disclosure and liking: A meta-analytic review. *Psychological Bulletin*, *116*(3), 457–475. https://doi.org/10.1037/0033-2909.116.3.457

Daft, R. L., & Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. *Management Science*, *32*(5), 513–644. https://doi.org/10.1287/mnsc.32.5.554

Dennis, A. R., & Kinney, S. T. (1998). Testing media richness theory in the new media: The effects of cues, feedback, and task equivocality. *Information Systems Research*, *9*(3). https://doi.org/10.1287/isre.9.3.256

Dingel, J. I., & Neiman, B. (2020). How many jobs can be done at home? *NBEC Working Paper Series 26948*. https://www.nber.org/system/files/working_papers/w26948/w26948.pdf

Downs, E., & Shyam Sundar, S. (2011). "We won" vs. "They lost": Exploring ego-enhancement and self-preservation tendencies in the context of video game play. *Entertainment Computing*. https://doi.org/10.1016/j.entcom.2011.03.012

Dresner, E., & Herring, S. C. (2010). Functions of the nonverbal in CMC: Emoticons and illocutionary force. *Communication Theory*, *20*(3), 249–268. https://doi.org/10.1111/j.1468-2885.2010.01362.x

Felson, R. B., & Steadman, H. J. (1983). Situational factors in disputes leading to criminal violence. *Criminology*, *21*(1). https://doi.org/10.1111/j.1745-9125.1983.tb00251.x

Freeman, G. Z., Bardzell, J., & Bardzell, S. (2016). Intimate experiences in virtual worlds: The interplay among hyperpersonal communication, avatar-based systems, and experiential drives. *IConference 2016 Proceedings*, 1–10. https://doi.org/10.9776/16239

Furr, J. B., Carreiro, A., & McArthur, J. A. (2016). Strategic approaches to disability disclosure on social media. *Disability and Society*. https://doi.org/10.1080/09687599.2016.1256272

Gibbs, J. L., Ellison, N. B., & Heino, R. D. (2006). Self-presentation in online personals: The role of anticipated future interaction, self-disclosure, and perceived success in internet dating. *Communication Research*, *33*(2). https://doi.org/10.1177/0093650205285368

Gottman, J. M., & Krokoff, L. J. (1989). Marital Interaction and Satisfaction: A Longitudinal View. *Journal of Consulting and Clinical Psychology*, *57*(1). https://doi.org/10.1037//0022-006x.57.1.47

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, *78*(6), 1360–1380.

Günther, F., Dudschig, C., & Kaup, B. (2014). LSAfun - An R package for computations based on latent semantic analysis. *Behavior Research Methods*, *47*(4), 930–944. https://doi.org/10.3758/s13428-014-0529-0

Gupta, B., Iyer, L., & Weisskirch, R. (2010). Facilitating global e-commerce: A comparison of consumers' willingness to disclose personal information online in the U.S. and in India. *Journal of Electronic Commerce Research*, *11*(1), 41–52.

Hall, J. A., & Baym, N. K. (2012). Calling and texting (too much): Mobile maintenance expectations, (over)dependence, entrapment, and friendship satisfaction. *New Media and Society*, *14*(2), 316–331. https://doi.org/10.1177/1461444811415047

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford Press.

Haythornthwaite, C. (2005). Social networks and Internet connectivity effects. *Information, Communication & Society*, *8*(2), 125–147. https://doi.org/10.1080/13691180500146185

Henderson, S., & Gilding, M. (2004). "I've never clicked this much with anyone in my life": Trust and hyperpersonal communication in online friendships. *New Media and Society*, *6*(4), 487–506. https://doi.org/10.1177/146144804044331

Hertel, G., Schroer, J., Batinic, B., & Naumann, S. (2008). Do shy people prefer to send e-mail? Personality effects on communication media preferences in threatening and nonthreatening situations. *Social Psychology*, *39*(4). https://doi.org/10.1027/1864-9335.39.4.231

Houghton, D. J., & Joinson, A. N. (2012). Linguistic markers of secrets and sensitive self-disclosure in Twitter. *45th Hawai'i International Conference on System Sciences*, 3480–3489. https://doi.org/10.1109/HICSS.2012.415

Ireland, M. E., Slatcher, R. B., Eastwick, P. W., Scissors, L. E., Finkel, E. J., & Pennebaker, J. W. (2011). Language style matching predicts relationship initiation and stability. *Psychological Science*, *22*(1), 39–44. https://doi.org/10.1177/0956797610392928

Jackson, D. (2019). *Disclosing the undisclosed: Social, emotional, and attitudinal information as modeled predictors of #MeToo posts* [Master's Thesis, Purdue University]. https://doi.org/10.25394/PGS.8026184.v1

Jacobson, D. (1999). Impression formation in cyberspace: Online expectations and offline experiences in text-based virtual communities. In *Journal of Computer-Mediated Communication* (Vol. 5, Issue 1). https://doi.org/10.1111/j.1083-6101.1999.tb00333.x

Jiang, L. C., Bazarova, N. N., & Hancock, J. T. (2011). The disclosure-intimacy link in computer-mediated communication: An attributional extension of the hyperpersonal model. *Human Communication Research*, *37*(1), 58–77. https://doi.org/10.1111/j.1468-2958.2010.01393.x

Kanter, J. B., Lavner, J. A., Lannin, D. G., Hilgard, J., & Monk, J. K. (2022). Does couple communication predict later relationship quality and dissolution? A meta-analysis. *Journal of Marriage and Family*, *84*(2). https://doi.org/10.1111/jomf.12804

Kashian, N., Jang, J. woo, Shin, S. Y., Dai, Y., & Walther, J. B. (2017). Self-disclosure and liking in computer-mediated communication. *Computers in Human Behavior*, *71*, 275–283. https://doi.org/10.1016/j.chb.2017.01.041

Khalilzadeh, J., & Tasci, A. D. A. (2017). Large sample size, significance level, and the effect size: Solutions to perils of using big data for academic research. *Tourism Management*, *62*, 89–96. https://doi.org/10.1016/j.tourman.2017.03.026

Kim, J., & Dindia, K. (2011). Online self-disclosure : A review of research. In K. B. Wright & L. M. Webb (Eds.), *Computer-mediated communication in personal relationships* (pp. 156–180). Peter Lang.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240. https://doi.org/10.1037/0033-295X.104.2.211

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*(2–3), 259–284. https://doi.org/10.1080/01638539809545028

Lea, M., & Spears, R. (1991). Computer-mediated communication, de-individuation and group

    decision-making. *International Journal of Man-Machine Studies*, *34*(2), 283–301.

    https://doi.org/10.1016/0020-7373(91)90045-9

Lev-On, A. (2015). Equalization of What? Status, Expertise, and Tenure in Virtual Communities

    of Practice. *Human Service Organizations Management, Leadership and Governance*,

    *39*(2). https://doi.org/10.1080/23303131.2015.1011763

Liu, Z., & Wang, X. (2018). How to regulate individuals' privacy boundaries on social network

    sites: A cross-cultural comparison. *Information and Management*, *55*(8).

    https://doi.org/10.1016/j.im.2018.05.006

Locke, S. D., & Gilbert, B. O. (1995). Method of psychological assessment, self-disclosure, and

    experiential differences: A study of computer, questionnaire, and interview assessment

    formats. *Journal of Social Behavior and Personality*, *10*(1), 255–263.

Lorenzo-Dus, N., Izura, C., & Pérez-Tattam, R. (2016). Understanding grooming discourse in

    computer-mediated environments. *Discourse, Context and Media*, *12*, 40–50.

    https://doi.org/10.1016/j.dcm.2016.02.004

Lowe, R., Pow, N., Serban, I. v., & Pineau, J. (2015). The Ubuntu dialogue corpus: A large

    dataset for research in unstructured multi-turn dialogue systems. *SIGDIAL 2015 - 16th*

    *Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 285–294.

    https://doi.org/10.18653/v1/w15-4640

Maheux, A. J., Nesi, J., Galla, B. M., Roberts, S. R., & Choukas-Bradley, S. (2021). #Grateful:

    Longitudinal Associations Between Adolescents' Social Media Use and Gratitude During

    the COVID-19 Pandemic. *Journal of Research on Adolescence*, *31*(3).

    https://doi.org/10.1111/jora.12650

Malloch, Y. Z., & Taylor, L. D. (2019). Emotional self-disclosure in online breast cancer support

    groups: Examining theme, reciprocity, and linguistic style matching. *Health*

    *Communication*, *34*(7), 764–773. https://doi.org/10.1080/10410236.2018.1434737

Martin, M. (2021). *Computer and Internet Use in the United States: 2018*.

    https://www.census.gov/content/dam/Census/library/publications/2021/acs/acs-49.pdf

Nguyen, M., Bin, Y. S., & Campbell, A. (2012). Comparing online and offline self-disclosure: A

    systematic review. *Cyberpsychology, Behavior, and Social Networking*, *15*(2), 103–111.

    https://doi.org/10.1089/cyber.2011.0277

O'Sullivan, P. B. (2000). What you don't know won't hurt me: Impression management functions of communication channels in relationships. *Human Communication Research*, *26*(3). https://doi.org/10.1093/hcr/26.3.403

Park, N., Jin, B., & Annie Jin, S.-A. (2011). Effects of self-disclosure on relational intimacy in Facebook. *Computers in Human Behavior*, *27*(5), 1974–1983. https://doi.org/10.1016/j.chb.2011.05.004

Parks, M. R., & Floyd, K. (1996). Making friends in cyberspace. *Journal of Communication*, *46*(1), 80–97. https://doi.org/10.1111/j.1460-2466.1996.tb01462.x

Parks, M. R., & Roberts, L. D. (1998). "Making MOOsic": The development of personal relationships on line and a comparison to their off-line counterparts. *Journal of Social and Personal Relationships*, *15*(4), 517–537. https://doi.org/10.1177/0265407598154005

Pennebaker, J. W., Booth, R. J., Boyd, R. L., & E., F. M. (2015). *Linguistic inquiry and word count: LIWC2015* (No. 2015). Pennebaker Conglomerate. http://www.liwc.net

Pennebaker, J. W., Booth, R. J., Boyd, R. L., & E., F. M. (2022). *Linguistic inquiry and word count: LIWC2022*. Pennebaker Conglomerate.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. University of Texas at Austin. https://doi.org/10.1115/MSEC2014-3991

Pettersen, L. (2016). The role of offline places for communication and social interaction in online and virtual spaces in the multinational workplace. *Nordicom Review*, *37*(Special Issue), 131–146. https://doi.org/10.1515/nor-2016-0028

R Core Team. (2018). *R: A language and environment for statistical computing*.

Robertson, A., Liza, F. F., Nguyen, D., McGillivray, B., & Hale, S. A. (2021). Semantic journeys: Quantifying change in emoji meaning from 2012-2018. *4th International Workshop on Emoji Understanding and Applications in Social Media*. http://arxiv.org/abs/2105.00846

Rosenfeld, M. J., Thomas, R. J., & Hausen, S. (2019). Disintermediating your friends: How online dating in the United States displaces other ways of meeting. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(36). https://doi.org/10.1073/pnas.1908630116

Seraj, S., Blackburn, K. G., & Pennebaker, J. W. (2021). Language left behind on social media exposes the emotional and cognitive costs of a romantic breakup. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(7). https://doi.org/10.1073/pnas.2017154118

Sheer, V. C., & Chen, L. (2004). Improving media richness theory: A study of interaction goals, message valence, and task complexity in manager-subordinate communication. *Management Communication Quarterly*, *18*(1). https://doi.org/10.1177/0893318904265803

Short, J., Williams, E., & Christie, B. (1976). *The Social psychology of telecommunications*. John Wiley and Sons.

Sillars, A., & Zorn, T. E. (2021). Hypernegative interpretation of negatively perceived email at work. *Management Communication Quarterly*, *35*(2), 171–200. https://doi.org/10.1177/0893318920979828

Snell, W. E., Miller, R. S., & Belk, S. S. (1988). Development of the emotional self-disclosure scale. *Sex Roles*, *18*(1–2), 59–73. https://doi.org/10.1007/BF00288017

Song, Z., Fore, D., Strassel, S., Lee, H., & Wright, J. (2018). *BOLT English SMS/Chat ReadMe*. https://catalog.ldc.upenn.edu/docs/LDC2018T19/README.txt

Staude-Müller, F., Hansen, B., & Voss, M. (2012). How stressful is online victimization? Effects of victim's personality and properties of the incident. *European Journal of Developmental Psychology*, *9*(2), 260–274. https://doi.org/10.1080/17405629.2011.643170

Swarns, R. L. (1996). *Arrest at Columbia shows risks in on-line dating*. New York Times.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*(1), 24–54. https://doi.org/10.1177/0261927X09351676

Taylor, D. A., & Altman, I. (1987). Communication in interpersonal relationships: Social penetration processes. *Interpersonal Processes: New Directions in Communication Research. Sage Annual Reviews of Communication Research*.

Thompson, D., & Filik, R. (2016). Sarcasm in written communication: Emoticons are efficient markers of intention. *Journal of Computer-Mediated Communication*, *21*(2), 105–120. https://doi.org/10.1111/jcc4.12156

Tidwell, L. C., & Walther, J. B. (2002). Computer-mediated communication effects on disclosure, impressions, and interpersonal evaluations: Getting to know one another a bit at a time. *Human Communication Research*, *28*(3), 317–348. https://doi.org/10.1111/j.1468-2958.2002.tb00811.x

Toma, C. L., & Hancock, J. T. (2011). A new twist on love's labor: Self-presentation in online dating profiles. In K. B. W. Wright & L. M. Webb (Eds.), *Computer-mediated communication in personal relationships* (pp. 41–55). Peter Lang. https://doi.org/10.1016/S0262-4079(09)63019-0

Toma, C. L., Hancock, J. T., & Ellison, N. B. (2008). Separating fact from fiction: An examination of deceptive self-presentation in online dating profiles. *Personality and Social Psychology Bulletin*, *34*(8), 1023–1036. https://doi.org/10.1177/0146167208318067

US Census. (2010). *Computer and Internet use in the United States: 1984 to 2009*. https://www.census.gov/data/tables/time-series/demo/computer-internet/computer-use-1984-2009.html

VanLear, C. A., Sheehan, M., Withers, L. A., & Walker, R. A. (2005). AA online: the enactment of supportive computer mediated communication. *Western Journal of Communication*, *69*(1), 5–26. https://doi.org/10.1080/10570310500033941

Verma, I. M. (2014). Editorial expression of concern and correction for "Experimental evidence of massive-scale emotional contagion through social networks." *Proc Natl Acad Sci*, *111*(29), 10779. www.pnas.org/cgi/doi/10.1073/pnas.1412469111

Vijayakumar, N., & Pfeifer, J. H. (2020). Self-disclosure during adolescence: exploring the means, targets, and types of personal exchanges. *Current Opinion in Psychology*, *31*, 135–140. https://doi.org/10.1016/j.copsyc.2019.08.005

Vossen, H. G. M., Koutamanis, M., & Walther, J. B. (2017). An experimental test of the effects of online and face-to-face feedback on self-esteem. *Cyberpsychology*. https://doi.org/10.5817/CP2017-4-1

Walther, J. B. (1992). Interpersonal effects in computer-mediated interaction: A relational perspective. *Communication Research*, *19*(1), 52–90. https://doi.org/10.1177/009365092019001003

Walther, J. B. (1996). Computer-mediated communication: Impersonal, interpersonal and hyperpersonal interaction. *Communication Research*, *23*(3), 3–43.

Walther, J. B. (2007). Selective self-presentation in computer-mediated communication: Hyperpersonal dimensions of technology, language, and cognition. *Computers in Human Behavior*. https://doi.org/10.1016/j.chb.2006.05.002

Walther, J. B., & Burgoon, J. K. (1992). Relational communication in computer-mediated interaction. *Human Communication Research*, *19*(1), 50–88. https://doi.org/10.1111/j.1468-2958.1992.tb00295.x

Walther, J. B., Kashian, N., Jang, J. W., Shin, S. Y., Dai, Y. (Nancy), & Koutamanis, M. (2016). The effect of message persistence and disclosure on liking in computer-mediated communication. *Media Psychology*, *21*(2), 308–327. https://doi.org/10.1080/15213269.2016.1247718

Walther, J. B., & Parks, M. R. (2002). Cues filtered out, cues filtered in: Computer-mediated communication and relationships. In M. L. Knapp & J. A. Daly (Eds.), *Handbook of Interpersonal Communication* (3rd ed., pp. 529–563). Sage.

Walther, J. B., van der Heide, B., Ramirez, A., Burgoon, J. K., Peña, J., Heide, B. van der, Ramirez, A., Burgoon, J. K., & Peña, J. (2015). Interpersonal and hyperpersonal dimensions of computer-mediated communication. In S. S. Sundar (Ed.), *The handbook of the psychology of communication technology* (pp. 3–22). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118426456.ch1

Walther, J. B., van der Heide, B., Tong, S. T., Carr, C. T., & Atkin, C. K. (2010). Effects of interpersonal goals on inadvertent intrapersonal influence in computer-mediated communication. *Human Communication Research*, *36*(3), 323–347. https://doi.org/10.1111/j.1468-2958.2010.01378.x

Waterloo, S. F., Baumgartner, S. E., Peter, J., & Valkenburg, P. M. (2018). Norms of online expressions of emotion: Comparing Facebook, Twitter, Instagram, and WhatsApp. *New Media and Society*, *20*(5). https://doi.org/10.1177/1461444817707349

Welch, J. R. (2020). *Uses and gratifications of scientific subreddits: An examination of online message typologies and their impact* (Issue August) [Dissertation]. Purdue University.

Weller, K., & Kinder-Kurlanda, K. E. (2015). *Uncovering the Challenges in Collection, Sharing and Documentation: The Hidden Data of Social Media Research?* https://www.ldc.upenn.edu/about

Welles, B. F., Rousse, T., Merrill, N., & Contractor, N. (2014). Virtually friends: An exploration of friendship claims and expectations in immersive virtual worlds. *Journal of Virtual Worlds Research*, *7*(2), 1–15. https://doi.org/10.4101/jvwr.v7i2.7024

Wiemann, J. M., & Backlund, P. (1980). Current theory and research in communicative competence. *Review of Educational Research*, *50*(1), 185–198. https://doi.org/10.3102/00346543050001185

Workman, H., & Coleman, C. A. (2014). "The front page of the Internet": Safe spaces and hyperpersonal communication among females in an online community. *Southwestern Mass Communication Journal*, *29*(2).

Yum, Y. O., & Hara, K. (2005). Computer-mediated relationship development: A cross-cultural comparison. *Journal of Computer-Mediated Communication*, *11*(1). https://doi.org/10.1111/j.1083-6101.2006.tb00307.x

# APPENDIX A: WORD CATEGORIES USED FROM LIWC 2022

Note: Due to the specifics of the LIWC licensing agreement, not all words in all categories are presented. This is intended as an example, not an exhaustive list.

Verbs: Am ask began begin creates descends describe described describes didn't died e-mailed exclude explain followed forgetting getting googles grow hate havent he's here's kick loses mastered obey obeying overcome picking reading resting say see sent sighs slay* speak swerves theyve thinking travels wanted warms weve wish wonder wonders works you'd

Prepositions: about abt across after ahead along amid amidst among* as atop before behind beneath beside besides by despite down except excluding for from hereafter in insides like minus near off on outside over plus regarding respecting sans thru til toward* under underneath unless until unto upon versus via vs within

Third person pronouns: he her hers herself him himself his hissel* oneself she their* them themself themselves they

Family: aunt* babies baby cousin* exhusb* exwive* family fiance fiance's godmother gramp* granddad* granddau* grandkid* grandson* husband* inlaw* ma mama marry maternal* mimi'* mimis mom mom's momma* mommy* motherhood mothering mothers mum mum's mummy* nana niece* pa papas pappy parent* paternity sibling* sister* son sons step-dau* step-fath* stepdau* uncles wife* wive*

Sexual: bdsm bi-sexual* boob* buttfuck* call-girl* callgirl* chlamydia condoms dildo* erectile fuck fwb gays gonorrhea* hard-on* homo hump* incest* jism jissom jizz* lover* makeout* mating nipple* nudi* orgy perver' prudish queer* rape* raping rapist* screw* seduc* sex sexier sexily sexing sexploit* sext* swinger* tit tits titties titty twat* vag vibrator* whore*

Discrepancies: abnormal* could could've couldn't couldnt expect* hope hopeful hopefully ideal* if impossible inadequa* lack lacked mistak* must mustnt need needed needn't normally odd odder ought ought'nt oughta oughtn't oughtve outstanding prefer* problem* rather regardless should should've shouldn't shouldnt undesir* undid undone unneccess* wanna wanted wants wished would've wouldn't wouldnt wouldve

Negative emotions: abuse* advers* aggression* annoys aversi* avoid* battl* condemn* confusing contempt* crude crudely cruel crueler cruelty cynic* damag* decay* defend* demot* devastat* devil* difficulties difficulty disaster dislike dislikes dismay* distraught disturb* doom* emotional empty enrag* fail* fake forbade fuckh* fucktwat* gloom gloomier grimac* grimly grr* guilty heartless* horrible incompeten* inferiority irrita*

Death: autops* behead* bereave* buried bury casket* casualt* cemet* coffin* coroner* corpse* dead demise die died dies doa drown* execution* exterminat* ghost* grief griev* hearse*

immortal* kill* lynch* manslaughter* mausoleum* memoria* morgue* mortal* mortician*
mourn* murder* obit* oded overdosed overdosed plague* reaper* slain slaughter* slay* tomb
tombs war warfare* wars zombie*

# APPENDIX B: LSAFUN COMMANDS USED

```
install.packages("readxl")
install.packages("writexl")
library(readxl)
library(writexl)
library(LSAfun)

# Load the Excel file and convert it to a dataframe
dat <- read_excel("path/Data_For_LSA.xlsx")
dat <- data.frame(dat)

load(file="baroni.rda")

# PreProcess the file to remove punctuation and make the dataset lowercase
dat$lower_sender_msg <- tolower(dat$lower_sender_msg)
dat$lower_sender_msg <- gsub(dat$lower_sender_msg,pattern='[[:punct:]]',replacement="")
dat$lower_receiver_msg <- tolower(dat$lower_receiver_msg)
dat$lower_receiver_msg <- gsub(dat$lower_receiver_msg,pattern='[[:punct:]]',replacement="")

dat$cos <- vector(length=nrow(dat)) #will create an empty column to include the similarity
scores

# Analyze each rope separately
for(i in 1:nrow(dat)){
 dat$cos[i] <- costring(dat$lower_sender_msg[i],dat$lower_receiver_msg[i],tvectors=baroni)

}

# Write the resultant dataframe to an excel file for analysis in SPSS.
write_xlsx(dat,"path /lsa.xlsx")
```