

Context sensitivity: A prognostic patient characteristic for digital psychotherapy

Thomas Derrick Hull

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2021

© 2021

Thomas Derrick Hull

All Rights Reserved

Abstract

Context sensitivity: A prognostic patient characteristic for digital psychotherapy

Thomas Derrick Hull

Background: Emotion regulation has been identified as a major contributor to the development of psychopathology and, by extension, to understanding the positive effects of various psychotherapy mechanisms. Little work has been done, however, on the extent to which individual components of emotion regulation operate as prognostic factors in psychological treatment. Context sensitivity and reflective functioning are emerging as important aspects of adaptive emotion regulation capacity and may be related to a portion of patient therapy outcome when investigated as a patient characteristic.

Design and Participants: A sample of 130 adults seeking treatment for depression and anxiety through a digital psychotherapy provider were recruited to participate. Individuals presenting with comorbid severe mental illness or psychosis, significant substance abuse concerns, active suicidal ideation, and active manic states were excluded from participation.

Methods: Participants completed individual difference measures for Five-factor Personality, reflective functioning (i.e., Reflective Functioning Questionnaire; RFQ), and the context sensitivity (i.e., Context Sensitivity Index; CSI), and were followed over three months of psychotherapy. Clinical outcomes were measured with diagnosis-specific symptom measures

such as the PHQ-9 for depression and the GAD-7 for anxiety at baseline and then every 3 weeks for the duration of treatment.

Results: Participants reported significant improvement in depression and anxiety symptoms after three months of treatment ($p < .001$). CSI and RFQ scores were unassociated with baseline symptom severity. Certainty about others' thoughts, an RFQ subscale, was inversely associated with outcome ($p < .05$). CSI scores were unassociated with treatment outcome at 3 months. Lack of insight, an RFQ subscale, significantly improved as a result of treatment when baseline symptoms were high ($p < .05$).

Conclusions: Reflective functioning may be a promising patient characteristic for explaining a modest portion of treatment outcome. Lack of emotional insight improved meaningfully as a result of treatment for individuals with more severe depression and anxiety at baseline. Further research is needed to investigate aspects of emotion regulation as a route towards better understanding outcome in psychotherapy.

Table of Contents

List of Charts, Graphs, Illustrations	ii
Introduction.....	1
Patient Characteristics in Psychotherapy	3
The Potential of Digital Psychotherapy	5
Regulatory Flexibility	7
The Current Study.....	14
Methods.....	18
Results.....	24
Convergence of Context Sensitivity Measures	27
Effect of Treatment on Individual Difference Variables	32
Discussion	41
Limitations and Future Directions	49
References.....	51
Appendix A.....	58

List of Charts, Graphs, Illustrations

Table 1. Demographic Characteristics for Full Sample (p. 28)

Table 2. Relationships between Context Sensitivity Measures (p.31)

Table 3. Linear model of individual difference predictors of baseline symptom scores. (p. 32)

Table 4. Linear model of individual difference predictors of final symptom scores. (p. 34)

Figure 1. Interaction between Pre-treatment and 3 Month Symptoms by level of Baseline Hypermentalization (p. 35)

Table 5. Linear model of symptom improvement predicting 3 Month Cue Presence. (p. 36)

Table 6. Linear model of symptom improvement predicting 3 Month Cue Absence. (p. 37)

Table 7. Linear model of symptom improvement predicting 3 month Hypermentalization. (p. 39)

Table 8. Linear model of symptom improvement predicting 3 month Hypermentalization. (p. 40)

Table 9. Linear model of symptom improvement predicting 3 month Hypomentalization. (p. 40)

Figure 2. Interaction between Pre-treatment Hypomentalization and 3 Month Hypomentalization by all three levels of Symptom change (p. 42)

Figure 3. Interaction between Pre-treatment Hypomentalization and 3 Month Hypomentalization with two levels of Symptom change (p. 43)

Table 10. Linear model of symptom improvement predicting 3 month Hypomentalization (p. 43)

Table 11. Associations between Gender, Symptom Severity, and Individual Differences (p. 63)

Introduction

In response to early criticism (Eysenck, 1952), research on psychotherapy over the last half century has demonstrated consistent and robust effects in alleviating a wide variety of disorders from the so-called common disorders of anxiety (Cuijpers et al., 2014, 2016) and depression (Barth et al., 2016; Mohr et al., 2014), to others such as obsessive-compulsive disorder (OCD; Olatunji et al., 2013), eating disorders (Godfrey et al., 2015), personality disorders (Cristea et al., 2017), and severe conditions like bipolar disorder (Chatterton et al., 2017; Oud et al., 2016) and psychosis (Velthorst et al., 2015). The research demonstrating these effects has largely derived from efficacy or randomized clinical (or controlled) trials (RCTs), but longitudinal, effectiveness study designs that privilege the way therapy is conducted in practice have also demonstrated the salubrious effects of treatment (de Jong & DeRubeis, 2018; Nathan, Stuart, & Dolan, 2000). While helpful for determining the overall impact of psychotherapy, these results cannot identify the specific factors associated with the change experienced by patients between the beginning of therapy and its termination. Investigating this question requires differentiating the effects of the therapy itself from other important factors such as the characteristics of the patient, the characteristics of the therapist, the relationship they establish with each other, and other factors common to all forms of psychotherapy.

The effects of each of these factors have been indirectly estimated, but few have been targeted either experimentally through randomization, or observationally using individual difference measures. Indirect estimates have suggested that patient characteristics in particular account for a large amount of variance in psychotherapy outcomes and could be the most promising avenue for explaining the effects of psychotherapy. Recent advances in emotion

regulation as a patient characteristic offer theoretical and operational resources for understanding the genesis and maintenance of various forms of psychopathology (Kring & Sloan, 2009), especially depression and anxiety (Mennin, Heimberg, Turk, & Fresco, 2005; Rottenberg, Gross, & Gotlib, 2005) and could be an important patient characteristic for understanding therapy outcomes. One particularly promising development in the emotion regulation literature pertains to the concept of regulatory flexibility (Bonanno & Burton, 2013), which highlights the ability to perceive and respond to the demands and opportunities imposed by emotionally triggering or stressful contexts. This individual differences paradigm proposes a number of components thought to impact emotion regulation and the course of psychopathology. These include a person's sensitivity to the demands and opportunities of a situation the person is in (i.e., context sensitivity), followed by their ability to deploy a regulation strategy from among their repertoire suitable to that situation, and finally their ability to register and respond to feedback on how well the strategy worked in that situation. Context sensitivity is the ability to perceive the presence and absence of important cues in a situation, and since it is the first step in the self-regulation process, it will be the primary focus of this investigation. Another potentially impactful individual difference theorized within regulatory flexibility is the ability of the person to understand the psychological states of others within the context of their own emotional states. This is an important backdrop for reflection on and flexibility within situations where emotion regulation is needed (Fonagy et al., 2016), and is here investigated in parallel with and as potentially related to context sensitivity. Finally, five-factor personality traits (McRae & Costa, 1996) offer a general and well-established individual difference framework with several factors related to emotion regulation. While of secondary interest, the five factors will be useful in

differentiating more domain general patient characteristics (the five factors) from more domain specific characteristics like emotion regulation.

Patient Characteristics in Psychotherapy

Research to isolate the various factors of change in psychotherapy has identified five sources of therapy effects that include 1) common factors, 2) expectation and placebo, 3) specific factors, 4) patient characteristics, and 5) therapist characteristics (Wampold, 2001, 2015; Wampold & Imel, 2015). Common factors include the therapeutic alliance and relational bond with the therapist that are important to every type of psychotherapy. Expectation and placebo capture the patient's investment in and hope for positive outcomes from the therapy. The specific factors are the particular interventions theorized to drive clinical change and that distinguish different therapy approaches from each other (e.g., cognitive restructuring in Beckian CBT vs. exploration of fantasy in psychodynamic psychotherapy). Patient and therapist characteristics are often considered to be "extratherapeutic" factors that have a significant impact on outcome but that are not directly controllable by the treatment setting. These five sources each account for different degrees of variance in outcome.

An early estimate (Lambert, 1992) of the relative contribution of each of these factors attributed 30% of outcome to common factors, 15% to specific factors, 15% to expectation effects, and 40% to extratherapy factors and therapist/patient characteristics. While suggestive, these allocations were based on the author's impressions from the literature and lacked any direct empirical evidence. A recent meta-analysis (Cuijpers et al., 2012) sought to more firmly establish these estimates by isolating improvements from 1) within control conditions (to quantify extratherapy and expectation factors), 2) across multiple treatment arms (to quantify common factors), and 3) from superior effects of carefully operationalized treatments over supportive

counseling (to quantify specific factors). Under this approach, extratherapeutic and expectation factors were responsible for 33.3% of outcome, common factors for 49.6% of outcome, and specific factors for 17.1% of outcome. While there continues to be a lot of research on and debate over what counts as a common factor versus what counts as a specific factor, as well as on specific treatment packages and the effect of placebo, comparatively little has been done to investigate extratherapy factors, even though they account for a large proportion of outcome.

There could be a number of reasons for this. One is that the focus on techniques and packages tends to treat other sources of improvement as nuisance variables that are either controlled statistically or through randomization. Another is the practicality of obtaining sufficiently large samples of patients for whom outcome data and individual difference measures exist in one data set. Recruiting subjects for treatment studies while overcoming obstacles of dropout and assessment completion is challenging enough without adding additional measures. One way to work around these challenges is through recent developments in digital mental health. These platforms deliver therapy through modern communication technologies, are gaining in popularity, and have advantages for patient recruitment and data capture that traditional face-to-face settings typically lack. The accessibility and convenience of smartphones and other mobile devices allow therapists to reach people who would otherwise be excluded due to geographic distance or scheduling conflicts. The malleability of the digital interface between therapist and patient also makes it easier for researchers to present assessments as a seamless part of the patient's experience to facilitate assessment completion. The combination of increased accessibility and malleability makes digital health platforms a promising medium for investigating historically intractable research questions.

While data capture and rapid recruitment on digital platforms provides a solution to past research barriers, there are certainly differences in how care is delivered and received on these platforms that may limit the generalizability of the findings to all therapy settings. This issue is discussed first, followed by the overview of a promising individual differences paradigm.

The Potential of Digital Psychotherapy

Despite the large volume of research supporting psychotherapeutic treatment, the majority of adults experiencing psychiatric disorders do not obtain face-to-face treatment due to barriers in accessing mental health care (Young, Klapp, Sherbourne & Wells, 2001; Brody, Khaliq & Thompson, 1997). Significant barriers to care can occur at several levels (Scheppers, Dongen, Dekker, Geertzen & Dekker, 2006) and include issues such as geographic remoteness, economic or insurance constraints, shortage of practitioners, stigmatization, and physical impairment (Alleman, 2002; Fjeldsoe, Marshall, & Miller, 2009; Hollon et al., 2002; Nutting et al., 2002; Young, 2005). These barriers lead to insufficient treatment access, particularly in underserved populations, highlighting the need of innovative mechanisms to enhance treatment delivery (Comer & Barlow, 2014; Mohr et al., 2006).

Technology-based delivery of treatment has become increasingly popular in recent years. Telemedicine interventions may be uniquely positioned to mitigate inequalities of access to care by reducing wait times and increasing accessibility through overcoming geographic and mobility barriers. A number of studies have identified contexts in which therapy delivered via technology platforms can be effective in reducing symptoms across a range of psychiatric diagnoses (Bee et al., 2008; Hull & Mahan, 2017; Kessler, et al., 2009; Nelson, Barnard, & Cain, 2003; Reynolds, Stiles & Grohl, 2006). For example, a meta-analytic review of thirteen studies examining the effectiveness of psychotherapeutic interventions delivered by telephone, internet and

videoconference, identified a medium pooled effect size of .44 for those with depression, and a large pooled effect size of 1.15 for those with anxiety disorders (Bee et al., 2008).

Research on technology-mediated treatment has largely been conducted with respect to its most common synchronous medium, live video. Newer forms of synchronous and asynchronous delivery, such as Multimedia message services (MMS) or “texting,” have so far received much less attention as a form for delivering direct clinical care, but are promising given the wide availability of texting platforms. MMS has been effectively used in the past as an adjunct to clinical care, occupying the role of a reminder system or symptom tracker focusing on promoting healthy lifestyle behavior and medication adherence (Fjeldsoe, Marshall, & Miller, 2009; Militello, Kelly, & Melnyk, 2012). Synchronous MMS, or “live chat,” has been used to directly deliver psychotherapy and was shown to be effective in improving symptom remission when combined with a primary care provider (Kessler et al., 2009). The next generation of MMS has piloted asynchronous modes for delivering care in an attempt to simplify scheduling barriers and expedite the beginning of treatment. In this approach, patients are free to message their provider whenever they like 24/7, and clinicians respond during pre-identified times each day for at least five days a week. Preliminary evidence suggests that this approach is an acceptable and potentially effective medium for delivering therapeutic interventions (Hull & Mahan, 2017).

Ongoing research on the use of asynchronous MMS messaging therapy in treating specific diagnostic types continues to bear out its utility as a broadly effective and acceptable form of delivering care. Longitudinal effectiveness studies on depression and anxiety (Hull, Malgaroli, Connolly, Feuerstein, & Simon, 2020) and on PTSD (Malgaroli, Hull, Wiltsey-Stirman, & Resick, 2020) have shown that between 48% and 53% of treatment seeking patients experience clinically significant change, with another 15% to 20% showing marked, reliable

improvement. Latent growth mixture models of these patient populations have identified that patient characteristics such as gender, age, and readiness for treatment are important in accounting for symptom reduction and treatment success. Similar to face-to-face treatment, measures of the therapeutic alliance have also been shown to predict outcomes for messaging therapy (Hull & Mahan, 2017; Malgaroli, Hull, Wiltsey-Stirman, & Resick, 2020). These findings so far suggest that differences in therapy change mechanisms, if any, between traditional face-to-face settings and digital mental health settings have had little impact on treatment success. Thus there is reason to expect that patient characteristics, especially from the perspective of emotion regulation, are important for psychotherapy regardless of the delivery medium employed. The next section suggests an important role for individual differences in emotion regulation as part of the basis for and maintenance of psychopathology. Regulatory flexibility is highlighted as a particular framework for investigating the role of emotion regulation as an informative patient characteristic in psychotherapy.

Regulatory Flexibility

Individual differences in emotion regulation have been implicated in the development of psychopathology (Buss, Davidson, Kalin, & Goldsmith, 2004; Gerhicke, & Shapiro, 2000; Kring & Sloan, 2009). These differences can be identified across several aspects of emotional experience and control. Recent research has pointed to the importance of the ability to understand and perceive the demands and opportunities imposed by emotionally triggering or stressful contexts, a concept referred to as context sensitivity (Bonanno & Burton, 2013). When emotions are expressed without regard for specific elements of contextual demands, negative outcomes associated with psychopathology, including depression, mania, and grief (Bonanno et

al., 2007; Gruber, Johnson, Oveis, & Keltner, 2008; Rottenberg & Gotlib, 2004; Rottenberg, Gross & Gotlib, 2005) are more likely.

For example, individuals with depression exhibit signs of inflexible emotional responses, primarily blunted expressions regardless of the demands of the environment (Rottenberg, Kasch, Gross, & Gotlib, 2002). Deficits have also been indicated in samples with generalized anxiety where elevations in expressivity of negative emotion in comparison to healthy controls have been identified (Mennin, Heimberg, Turk, & Fresco, 2005). Atypical responding to contextual factors and emotional expressions of others has also been associated with a number of psychological disorders and has been connected to impairments in socialization and communication, increased aggressive behaviors, the inability to modulate behavior according to the social context, or even a failure to avoid actions that cause harm to others (Penton-Voak et al., 2013).

Conversely, the ability to flexibly regulate emotional reactions based on features of the immediate context has been associated with well-being and behavioral adjustment (Bonanno, Papa, Lalande, Westphal, & Coifman, 2004; Bonanno, Pat-Horenczyk, & Noll, 2011; Bonanno et al., 2007; Levy-Gigi et al., 2015; Troy, Shallcross, & Mauss, 2013). Bonanno and Burton (2013) have theorized three general sets of abilities that give rise to regulatory flexibility. The first, sensitivity to context, is an important first step towards flexible regulation because emotions confer an advantage only when deployed in appropriate situations. Second, having a broad repertoire of emotion regulation strategies is critical to addressing the specific and highly variable demands of the situation. Finally, one's perception of the context may not always be accurate, or one's regulatory repertoire may not be sufficient, such that the ability to monitor the effectiveness of emotion regulation efforts and respond to feedback is vital (see Sheppes, Suri, &

Gross, 2015 for a complimentary approach). Research to date has demonstrated that context insensitivity (Coifman & Bonanno, 2010), and a narrow and inflexible repertoire of regulation strategies (Troy et al., 2013), are associated with poor adjustment and well-being.

As a set of cognitive and social skills that kick off any emotion-related event, context sensitivity stands out among the aspects of regulatory flexibility as an individual difference that is particularly promising for predicting baseline severity of psychopathology as well as response to psychotherapy.

Context Sensitivity and the Context Sensitivity Index

It is a generally accepted principle of emotion theory (Lewis, Haviland-Jones, & Barrett, 2010), and of frameworks of action broadly (Carver & Scheier, 2008), that the way we respond to situations is context bound. When emotions are expressed in inappropriate contexts, or when expressions and actions go beyond the bounds of the context, this constitutes context-insensitive emotional responding, and has been tied to psychopathology including depression, anxiety, mania, and prolonged grief (Bonanno et al., 2007; Gruber, Johnson, Oveis, & Keltner, 2008; Rottenberg & Gotlib, 2004; Rottenberg, Gross & Gotlib, 2005).

Context-sensitive responding, on the other hand, involves the ability to shift emotion responses, such as facial expressions, speech, and other bodily actions, in accordance with the demands and opportunities of the context and has been associated with positive adjustment (Aldao, 2013; Bonanno & Burton, 2013; Rottenberg, Gross, & Gotlib, 2005). Various forms of psychopathology can be considered disorders involving emotion context-insensitivity, or a persistent mismatch between the emotion expressed and the context. For example, fear is generally normative and useful, however when expressed repeatedly by an individual with PTSD in a safe setting or at an inappropriate time, it can become pathogenic (Davidson, Jackson, &

Kalin, 2000). For those experiencing grief, feelings of sadness or guilt are natural and expected, but the prolonged and continuous expression of sadness or guilt regardless of what the context demands tends to lead to difficulties adjusting to the loss.

It has been theorized, and data suggests, that context insensitivity is derived from individual differences in the ability to read contextual cues and to identify when important information is misleading or missing (Bonanno, Maccallum, Malgaroli, & Hou, 2020). This work further specifies that context sensitivity be broken down into subsets of abilities such as being able to tell how much control one has over a situation, how much control others have over a situation, how threatening a situation is, and the urgency to react to a situation. Importantly, research suggests that there are significant individual differences in these abilities (Bonanno, Maccallum, Malgaroli, & Hou, 2020; Cheng, Chiu, Hong, & Cheung, 2001; Coifman & Bonanno, 2010), which highlighted the need for a standardized measure to capture these differences that can be broadly administered and that is distinct to context sensitivity.

The Context Sensitivity Index (CSI; Bonanno, Maccallum, Malgaroli, & Hou, 2020) was developed to address several gaps in the literature on context and self-regulation. First, it sought to overcome the limitations of emotion-evoking films and interview tasks by separating the perception of contextual cues from the response to these cues. Second, it was designed to be completed by individuals without having to come into a lab to ease administrative burden. Third, it relied on standardized hypothetical scenarios to minimize recall bias and to minimize reliance on an individual's self-insight regarding context sensitivity, which are typical problems for self-report questionnaires (Schwartz, Neale, Marco, Shiffman, & Stone, 1999). Finally, items sought to measure a person's sensitivity to the contextual cues that are present, as well as their ability to determine when important cues are absent. Five studies have helped to establish the measure's

validity and reliability to assess context sensitivity (Bonanno, Maccallum, Malgaroli, & Hou, 2020) across six different scenarios. Results supported the separation of Cue Presence and Cue Absence subscales, each of which exhibited associations with different classes of variables. For example, whereas the Cue Presence scale was associated with measures of flexibility and context perception, the Cue Absence scale was associated with measures of stress, depression and anxiety symptoms. The CSI is thus an important addition to research on regulatory flexibility and highlights its usefulness as a tool for assessing individual differences and patient characteristics for psychotherapy outcome research and psychopathology.

Reflective Functioning and the Five-Factor Model

As discussed, the CSI presents individuals with scenarios eliciting their judgments to establish the extent to which they are responding to cues in a given situation. An important corollary to their performance on this task is the extent to which individuals self-report sensitivity to the actions of others and their own feelings across contexts in their daily lives. The Reflective Functioning Questionnaire (RFQ) was designed to assess the degree to which individuals make assumptions about the goals and desires of other people (a subscale termed hypermentalization), as well as the extent to which individuals lack insight into their feelings (a subscale termed hypomentalization). Previous research has shown that RFQ scores predict symptom severity and that the measure demonstrates good convergent and discriminant validity with associations to measures of empathy and perspective-taking (Fonagy et al., 2016). This work has shown that the two subscales, hypomentalization and hypermentalization, are associated differently with the clinical presentation of those studied. Hypomentalization reflects deficits in a person's ability to be aware of their emotional experiences and have insight into their feelings. Hypermentalization is a state in which a person projects beliefs and desires onto

others that the other people do not have and to maintain certainty about their assumptions even in the face of contradictory behavior or other evidence. Hypomentalization was particularly associated with measures of depression and anxiety symptoms, whereas hypermentalization was most associated with relational volatility and conflict with others. However, the measure has only been used diagnostically to date and has not been studied as an aspect of treatment outcome, which will be a novel application of the measure in this study. The RFQ will also be investigated as a convergent, but separable aspect of context sensitivity that captures a person's understanding of how they identify cues from others (i.e. hypermentalization) and how they identify cues from themselves (i.e., hypomentalization).

A final approach to individual differences is the five-factor model of personality (McRae & Costa, 1996). This model has been studied in an extremely large number of contexts, including psychotherapy (see Bucher, Suzuki, & Samuel, 2019 for a review), and has been shown to generally be associated with many areas of human functioning. For this reason, including it as a more domain-general measure of personality alongside measures more specific to emotion regulation expands the types of patient characteristics investigated in the context of psychotherapy. While its association with variables of interest in psychotherapy research is modest, the five-factor model has been found to be a rate limiting factor on improvement (Zinbarg, Uliaszek, & Adler, 2008), as a predictor of whether someone will benefit more from pharmacotherapy versus psychotherapy (Bagby, Quilty, & Ryder, 2008), and as a moderator of the therapeutic alliance (Bucher, Suzuki, & Samuel, 2019). Extraversion, conscientiousness and openness tend to predict better therapy outcomes, whereas neuroticism runs counter to successful treatment (Bucher, Suzuki, & Samuel, 2019). The relationship between any of the five factors and context sensitivity is unclear and will be treated as exploratory in the analysis.

The Current Study

The current study seeks to address the lack of research on patient characteristics as an important aspect of psychotherapy outcomes by measuring the context sensitivity of treatment seeking adults using individual difference assessments on a digital mental health platform that include the CSI, RFQ, and five-factor model. The relationship between individual difference scores and baseline symptom severity for depression and anxiety, as well as the relationship between individual difference scores and improvement in symptom scores over the duration of treatment were investigated. Another important question is the extent to which treatment leads to changes in individual differences. This was investigated by administering the CSI and RFQ at both the beginning and end of treatment.

Although neither the RFQ nor CSI have been investigated within the context of therapy outcomes, we offer some hypotheses regarding the extent to which individual difference variables will change as a result of therapy. The CSI reports on two subscales, Cue Absence and Cue Presence, which reflect the individual's ability to sense whether important cues are present or absent. Previous work suggests that Cue Absence, especially, is lower in individuals with depression and anxiety (Bonanno, Maccallum, Malgaroli, & Hou, 2020), and it is possible that Cue Absence will evidence the most improvement in successful therapies. The RFQ is similarly composed of two subscales, hypomentalization and hypermentalization, which correspond to lack of emotional insight and lack of social insight respectively. Here it is anticipated that hypomentalization will see the most improvement as a result of therapy given that low scores on this scale are associated with a wide variety of disorders (Fonagy et al., 2016), whereas the hypermentalization scale most accurately distinguishes individuals with severe pathology, such as Borderline Personality Disorder. Since this sample is composed of individuals with depression

and anxiety disorders, we expect little variation and little improvement in hypermentalization scores. The relationship between both Cue Absence and hypomentalization with depression and anxiety symptoms suggests they may bear a relationship to each other as well, though theoretically they are dissociable concepts. Research on the Cue Presence scale of the CSI demonstrated a relationship between Cue Presence and other measures of context sensitivity and flexibility (Bonanno, Maccallum, Malgaroli, & Hou, 2020). Insofar as hypermentalization is a weakness in reading social cues, Cue Presence may be inversely related to hypermentalization scores.

Regarding the five-factor model, although five-factor traits typically do not change much over time, some research has found slight increases in certain factors as a result of therapy (Bleidorn et al, 2019). Since several studies have already investigated five-factor change through therapy, and since the measure is by far the longest of those given, a second five-factor administration will not be offered to the participants. As a result, change in any of the five factors was not investigated here.

In summary, we hypothesize the following:

1. Higher scores on the CSI and specific personality dimensions (i.e., extraversion, conscientiousness, and openness) will be associated with lower baseline symptom severity. Lower scores on the RFQ will be associated with lower baseline symptom severity.
 - a. We expect higher CSI scores and lower RFQ scores to predict lower baseline symptoms
 - b. The Cue Absence subscale of the CSI is predicted to be more strongly associated with baseline symptoms than the Cue Presence subscale

- c. The Hypomentalization subscale of the RFQ is predicted to be more strongly associated with baseline symptoms than the Hypermentalization subscale
2. Higher CSI, lower RFQ, and higher personality scores are expected to be associated with greater symptom improvement overall.
 - a. We expect higher CSI and lower RFQ scores to predict greater symptom improvement
 - b. Both the Cue Absence and Cue Presence subscales of the CSI are predicted to be associated with symptom improvement since both reflect skills that would likely enable a patient to derive benefit from therapy and transfer these benefits to other contexts
 - c. The Hypomentalization subscale of the RFQ is predicted to be more strongly associated with improvement than the Hypermentalization subscale, primarily because we do not expect a lot of variance in hypermentalization scores for a sample with depression and anxiety
3. Greater symptom improvement is expected to be associated with increases in CSI scores and decreases in RFQ scores. Improvements in personality scores from therapy have already been documented (Bleidorn et al, 2019) and are not investigated here.
 - a. Symptom improvement will be more strongly associated with increases in the Cue Absence subscale than the Cue Presence subscale, given the association between Cue Absence and symptom ratings in previous work
 - b. Symptom improvement will be more strongly associated with decreases in the Hypomentalization subscale of the RFQ than the Hypermentalization subscale

4. We expect a complementary association between the CSI and RFQ indices, supporting the notion that both are related to abilities to identify important cues from self and others. Associations are expected to weak however, given that the CSI is a performative measure, whereas the RFQ is a self report measure.
 - a. Cue Absence and Hypomentalization, while theoretically distinct, are predicted to be modestly associated given their relationship to anxiety and depression symptom scores
5. We made no specific hypotheses about the relationship between the five-factor model and either the CSI or RFQ, or their subscales.

Methods

Setting

The study was conducted with a digital mental health platform (Talkspace) used by independently practicing, licensed therapists in the United States. The platform is accessible through internet search, through Employee Assistance Programs, and as a behavioral health benefit through several individual insurance plans. Patients first meet with an intake clinician through a live messaging system to conduct a brief, standardized intake to identify the presenting complaint, patient treatment history, and the patient's provider preferences. This information informs a matching algorithm that prioritizes and presents three providers with the desired characteristics for the patient to choose among. Once a clinician is chosen, the provider is alerted, and the patient is immediately introduced to the messaging "room" where treatment takes place. Patients complete a self-report baseline assessment and the provider walks them through the informed consent and emergency contact process after which treatment can begin. Observations in this study include data collected as part of organizational quality assurance and program management processes between March 11, 2020 and June 17, 2020. All patients and clinicians gave written consent to the use of their data in a de-identified, aggregate format as part of the user agreement before they began using the platform. Study procedures were approved as exempt by the institutional review board at Teachers College, Columbia University (20-412).

Participants

Patients

Participants were individuals who presented with a chief complaint of anxiety or depression, were seeking treatment through the service, and who completed at least one PHQ-9 and/or GAD-7. Inclusion criteria consisted of: (1) being English speakers in the United States,

(2) between the ages of 18 and 65, (3) having regular internet or cellphone access, (4) receiving a depression or anxiety diagnosis from their assigned licensed mental health provider based on a clinical intake and live messaging or video-based interview, as recorded in the electronic medical record with ICD-10 codes, (5) scoring 10 or higher on the PHQ-9 and/or GAD-7. Exclusion criteria consisted of current or past diagnoses of: (1) bipolar disorder, (2) any schizophrenia spectrum and psychotic disorder, or psychotic features, (3) any medical or neurological condition that would better account for the symptoms, (4) substance or alcohol use disorder (5) any condition requiring hospitalization; or (6) suicidal thoughts and/or behavior sufficient to be marked a “Yes” on any of questions three through six (at least thoughts about a potential suicide method), on the Columbia Suicide Severity Rating Scale Lifetime-Recent Screen (Oquendo, Halberstam, & Mann, 2003), requiring a more intensive level of care that interrupted treatment on the platform. Patients meeting the above criteria and who had scores for the CSI, RFQ and five-factor measures were included.

Clinicians

Clinicians in the provider network were currently licensed in one or more states, were required to have at least a Masters degree, and had at least three years of post-licensure experience delivering mental health care. Clinicians were matched only to patients where licensure included the patient’s residence. There were a total of 17 clinicians – 45% of whom reported five to nine years of post-licensure experience, and 37.5% reporting ten or more years of experience. Ninety-one percent (91%) were female. Providers had a mean age of 40 (SD = 9.4) years, and they reported a mix of orientations as part of their provider profile: 66% cognitive-behavioral treatment, 40% third-wave cognitive behavioral, and 25.5% primarily psychodynamic.

Intervention

Clinicians and patients asynchronously exchanged text-, audio-, and video-based messages using a secure, HIPAA-compliant platform accessible on mobile devices and on desktop computers. Patients could freely send messages at any time without limit, and all messages were stored for the clinician when they returned to review the message history. Therapists responded to messages from their patients at least once a day, five days a week. Clinicians were expected to adhere to all reporting, professional, and ethical standards for their respective fields, and appropriate referrals were provided for patients judged to need a higher level of care.

The number of words exchanged between therapists and patients is automatically counted as meta-data by the platform, and was used as a proxy to quantify the extent of therapeutic interaction through the asynchronous messaging medium. Raw counts of words sent by clinicians and patients were used in analyses estimating contributing factors to treatment outcome.

Assessments

Patients were assessed for depression and anxiety symptoms at baseline and then every three weeks for the duration of treatment, or until the patient opted to stop receiving assessments. Assessments are introduced to patients as an important aspect of their care that facilitates goal setting and to track progress. In this study, five assessments from baseline to week 12 were analyzed, including: Baseline, Week 3, Week 6, Week 9, and Week 12.

The 9-item Patient Health Questionnaire (Kroenke & Spitzer, 2002) was used to identify the clinical severity of depression. Responses on all items were given on a 4-point Likert scale (0

=*Not at all* to 3 = *Nearly every day*) with a total maximum score of 27. Scores greater or equal than 10 have been shown to have high sensitivity and specificity as a threshold for clinical depression, or at least moderate depression (Kroenke, Spitzer, & Williams, 2001; Kroenke et al., 2001).

Anxiety symptoms were assessed with the 7-item Generalized Anxiety Disorder questionnaire (Spitzer et al., 2006). Responses on all items were given on a 4-point Likert scale (0 = *Not at all* to 3 = *Nearly every day*) with a total maximum score of 21. Scores of 10 or above have been shown having high sensitivity and specificity as a clinically significant threshold for at least moderate anxiety (Kroenke et al., 2007).

The CSI presents six scenarios of plausible everyday events and asks respondents to complete three to four questions for each scenario on a 7-point Likert scale (1 = *None/Not at all* to 7 = *Very Much/Extremely*) that gauge the respondents' sense of the controllability, urgency, threatening feeling, and cooperation potential in the situation (Bonanno, Maccallum, Malgaroli, & Hou, 2018). The measure totals 20 items. Ten items contribute to the subscale for Cue Presence with a minimum score of 10 and maximum score of 70. The other ten items map onto the Cue Absence subscale, are reverse coded, and have a minimum score of 10 and maximum of 70. An example scenario is "A friend calls and asks you to do a favor for their partner, whom you don't like." A Cue Presence prompt for this scenario is "How much control do you have over what happens next?" and a Cue Absence prompt is "How threatening is this situation?"

The RFQ is an 8-item survey that measures reflective functioning or mentalization with two subscales, hypermentalization and hypomentalization (Fonagy et al, 2016). Each item is rated on a 7-point Likert scale (1 = *Strongly Disagree* to 7 = *Strongly Agree*). The measure looks for extreme responses and has a straightforward recoding scheme that converts low responses (1

through 3) and high responses (5 through 7) into values between 1 and 3, all other values are recoded as 0. Items 1 through 6 map onto the hypermentalization subscale with a minimum possible score of 0 and a maximum score of 18. Items 2 and 4 through 8 map onto the hypomentalization subscale with a minimum possible score of 0 and a maximum score of 18. A sample hypermentalization item is “People’s thoughts are a mystery to me,” and a sample hypomentalization item is “I always know what I feel.” As a measure of emotional insight and sensitivity to the goals and desires of other people, it has been found to be an important aspect of differences in clinical presentation (Fonagy et al., 2016).

The five-factor instrument used here was collated by DeYoung, Quilty, and Peterson (2007). It has a total of 100 items and 40 were selected to equally represent each of the five factors and to reduce respondent fatigue. It uses a 5-point Likert scale (1 = *Strongly Disagree* to 5 = *Strongly Agree*) and allocates 8 items to each of the five factors. Examples of each of the factors are “I carry out my plans,” (Conscientiousness), “I love to reflect on things” (Openness), “I am not interested in other people’s problems” (Agreeableness), “I make friends easily” (Extraversion), and “I get angry easily” (Neuroticism). The items for each factor are averaged together to get a total score for that factor.

Data analysis plan

Clinical Outcomes

Clinical outcomes will be computed for effect size (Cohen’s d), for clinically significant change and reliable improvement metrics (following Jacobson & Truax, 1992), for remission criteria (scores of 4 or less), and for reduction of cardinal symptoms for depression (depressed mood and loss of interest; Kennedy, 2008; Kroenke, Spitzer, & Williams, 2003) and anxiety (nervousness

or anxious mood and uncontrollable worry; Kroenke et al., 2007). Number of patients experiencing a 50% reduction in symptoms is also reported.

Relation of Individual Difference Variables to Baseline Severity

Hierarchical regression predicting baseline severity scores for the GAD-7 and PHQ-9 as the dependent variables from scores for the CSI, RFQ, and FFM. Regression and all model testing will be conducted in JASP, version 0.12.1.

Relation of Individual Difference Variables to Outcome

Hierarchical regression predicting symptoms scores at 3 months for the GAD-7 and PHQ-9 combined as the dependent variable, controlling for platform usage (specifically, words written by the patient and therapist), baseline GAD-7 and PHQ-9 scores, and patient education level in step one, then incorporating scores from the CSI, RFQ, and FFM into step 2 to isolate the association of individual difference variables on change in symptom measures. Patient education and words written are the only demographic and usage variables associated with outcomes on this platform to date. Regression and all model testing will be run using JASP, version 0.12.1.

Effect of Treatment on Individual Difference Variables

The RFQ and CSI were administered at the beginning of treatment and at 3 months of completed treatment. A paired samples t-test will be run to check for general change between pre- and at 3 months of treatment. Hierarchical regression predicting scores from the RFQ and CSI combined at 3 months of treatment as the dependent variable, controlling for platform usage (specifically, words written by the patient and therapist), baseline RFQ and CSI scores, and patient education in step one, then incorporating change scores from the GAD-7 and PHQ-9 from baseline to 3 months into step 2 to isolate the association of symptom improvement on the individual difference variables. Regression and all model testing will be run using JASP, version 0.12.1.

Results

Sample Characteristics

Of the 237 patients who agree to participate, 130 patients met inclusion criteria and had complete scores for GAD-7, PHQ-9, CSI, RFQ and five-factor measures. Patients were between the ages of 18 and 65, with the majority (55%) falling between 26-35 years of age. Women were 63.1% of the patient sample, and 65.4% of patients had Bachelor's degrees or higher education level. There were no difference in baseline measures or demographics between the 130 who completed the study and the 107 who started, but did not complete. Table 1 provides the full distribution of demographic and clinical characteristics.

Table 1.
Demographic Characteristics for Full Sample (N=130).

Variable	Percentage (N)
<i>Age:</i>	
18-25	24% (31)
26-35	55.4% (72)
36-49	14.9% (19)
50+	5.7% (8)
<i>Education:</i>	
Bachelor Degree or Higher	65.4% (85)

High School Diploma or Less	34.6% (45)
<i>Race/Ethnicity:</i>	
European American	62.3% (81)
African American	14.8% (19)
Southeast/Asian American	12.0% (16)
Native American	0.5% (1)
Other	10.4% (13)
Hispanic/Latinx	12.7% (17)
<i>Gender:</i>	
Female	63.1% (82)
Male	30.7% (40)
[other identity]	6.2% (8)
<i>Patient's State:</i>	
California	16.2% (21)
New York	13.9% (18)
Washington	4.8% (6)
New Jersey	4.6% (6)

[other U.S. State]	60.5% (79)
--------------------	------------

Treatment duration was on average 90.4 days ($SD=15.2$). At three months of treatment, participants reported significantly less depression ($M = 7.03$, $SD = 5.2$, $t(129) = 5.92$, $p < .001$, $d = 0.52$, 95% CI_d [0.335, 0.701]), than prior to treatment ($M = 9.4$, $SD = 4.78$), and less anxiety ($M = 7.37$, $SD = 4.87$, $t(129) = 6.44$, $p < .001$, $d = 0.56$, 95% CI_d [0.378, 0.749]), than prior to treatment ($M = 10.4$, $SD = 4.85$). Viewed through the framework of reliable and clinically significant change (Jacobson & Truax, 1992), 43% of the sample reported PHQ-9 score reductions of 5 or more points and fell below the established threshold for probable depression, and 51% of the sample reported GAD-7 score decreases of 5 or more points and fell below the established threshold for probable anxiety by their last observation. Full remission (defined as sum scores of 4 points or less) at 3 months was reported by 30.9% of patients on the GAD-7, and 33% for the PHQ-9. Average scores for the cardinal symptoms of depression were 4.2 ($SD = .82$) prior to treatment and 2.1 ($SD = .51$) at three months, with 38% reporting no cardinal depression symptoms at three months. Average scores for the cardinal symptoms of anxiety were 4.7 ($SD = .72$) prior to treatment and 2.6 ($SD = .43$) at three months, with 40% reporting no cardinal anxiety symptoms at three months.

Treatment engagement, as measured by word count, was an average of 475.2 words ($SD = 784.9$) per week of treatment generated by patients and 323.0 ($SD = 650.7$) by therapists, about 50% less than the average previously reported for a very large sample of patients on this platform (cf. Malgaroli et al, 2020).

All of the following analyses are based on the full sample and look at change over three months of treatment. However, thirty-nine ($n = 39$, 30% of the full sample) participants

continued treatment past the three month mark, averaging a 267 day ($SD = 90.4$) treatment course and reported significantly less depression ($M = 7.33, SD = 5.6, t(38) = 2.77, p < .009, d = 0.44, 95\% CI_d [0.112, 0.771]$), than prior to treatment ($M = 9.74, SD = 4.34$), and less anxiety ($M = 7.08, SD = 4.70, t(38) = 2.94, p < .006, d = 0.47, 95\% CI_d [0.137, 0.800]$), than prior to treatment ($M = 9.90, SD = 4.84$).

Convergence of Context Sensitivity Measures

See Table 2 for relationships between the subscales of the RFQ and CSI. The negative association previously reported (Fonagy et al., 2016) between the Hypermentalizing and Hypomentalizing subscales was supported, pre-treatment, $r(130) = -.685, p < .001$, and after 3 months of treatment, $r(130) = -.653, p < .001$. Contrary to the positive association between the Cue Presence and Cue Absence subscales previously reported (Bonanno et al., 2018), a negative correlation was found, pre-treatment, $r(130) = -.395, p < .001$, and after 3 months of treatment, $r(130) = -.427, p < .001$. The hypothesized inverse relationship between Cue Absence and Hypomentalizing, and inverse relationship between Cue Presence and Hypermentalizing were not supported (see Table 2). Unexpectedly, there was a small positive association between the Cue Absence subscale and Hypermentalizing subscale at 3 months of treatment, $r(130) = .271, p < .002$.

Table 2. Relationships between Context Sensitivity Measures

Pre-Treatment	Cue Presence	Cue Absence
Hypermentalizing	.062	.014
Hypomentalizing	-.057	-.060
3 Months	Cue Presence	Cue Absence

Hypermentalization	-.165	.271**
Hypomentalization	.113	-.103

Note. *p<.05, **p<.01, ***p<.001

Individual Difference Measures and Baseline Symptoms

To examine the relationship between context sensitivity and baseline symptoms, we used two step hierarchical multiple regression to control for associations with the five-factor personality dimensions in step one, while RFQ and CSI subscales were analyzed in step. Depression and anxiety symptoms were combined in accordance with the finding that difficulties with context sensitivity were not specific to any single diagnostic category (Bonanno et al, 2018). Regression statistics are reported in Table 3.

Table 3. Linear model of individual difference predictors of baseline symptom scores.

	β	SE	r	sr ²	R	R ²	
Step 1						.424	.180
Constant		11.8					
Extraversion	-.162	.138	-.152	.026			
Conscientiousness	-.077	.192	-.132	.007			
Openness	.287**	.213	.129	.075			
Agreeableness	.084	.251	.127	.008			
Neuroticism	.322***	.155	.301***	.096			
Step 2						.454	.206
Constant		11.6					
Extraversion	-.151	.140	-.152	.026			
Conscientiousness	-.082	.197	-.132	.007			

Openness	.289**	.221	.129	.078
Agreeableness	.100	.262	.127	.010
Neuroticism	.241*	.175	.301***	.049
Cue Presence	.035	.082	.085	.002
Cue Absence	-.080	.133	-.083	.003
Hypermentalization	-.079	.236	-.203*	.003
Hypomentalization	.098	.258	.223*	.006

* $p < .05$, ** $p < .01$, *** $p < .001$

The regression model at step one reflects previously documented associations between the five-factor traits and symptom scores, $F(5,124) = 5.44$, $p < .001$, with Extraversion negatively associated and Neuroticism positively associated. Greater Openness scores were also associated with greater baseline symptoms. When the four context sensitivity measures were included in step two, none were significant predictors of baseline symptom scores.

Individual Difference Measures and Symptom Improvement

To examine the relationship between context sensitivity and symptom improvement resulting from treatment, we used three step hierarchical multiple regression to control for associations between words written by the client and baseline symptoms in step one, with RFQ and CSI scores in step two, and possible interactions in step three. All variables were centered prior to analysis to ensure low multicollinearity. The five-factor dimensions were not included as they were not a focus of this study. Initial analyses also suggested that obtained education, which previous work associated with symptom improvement, was not associated with symptom improvement in this sample ($r(130) = .013$, $p = .860$) and was thus not included in the model. The forward method was used in step three to identify significant interactions. Regression statistics are reported in Table 4.

Table 4. Linear model of individual difference predictors of final symptom scores.

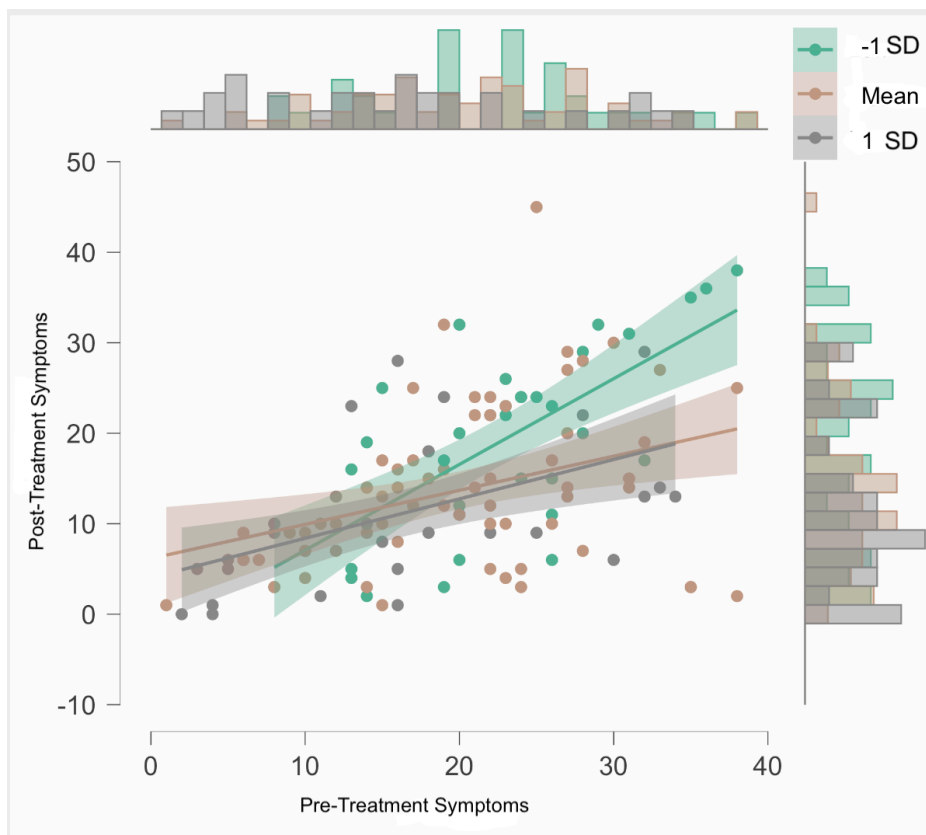
	β	SE	r	sr^2	R	R^2
Step 1					.530	.281
Constant		1.814				
Client Words Written	-.141	<.001	-.158	.027		
Baseline Symptom Scores	.506***	.045	.511***	.262		
Step 2					.567	.322
Constant		1.836				
Client Words Written	-.147*	<.001	-.158	.030		
Baseline Symptom Scores	.457***	.084	.511***	.224		
Cue Presence	-.006	.080	.020	<.001		
Cue Absence	-.050	.118	-.079	.003		
Hypermentalization	-.157	.223	-.288***	.019		
Hypomentalization	.057	.250	.272**	.003		
Step 3					.593	.352
Constant		1.841				
Client Words Written	-.155*	<.001	-.158	.035		
Baseline Symptom Scores	.482***	.083	.511***	.247		
Cue Presence	-.002	.079	.020	<.001		
Cue Absence	-.058	.116	-.079	.004		
Hypermentalization	-.199*	.223	-.288***	.030		
Hypomentalization	.019	.248	.272**	<.001		
Hyperment X Baseline Sxs	-.179*	.018	-.079	.045		

* $p < .05$, ** $p < .01$, *** $p < .001$

The regression model at step one reflects previously documented associations between client activity measured by number of words written and lower symptom scores after 3 Months of

treatment, $F(5,124) = 9.72, p < .001$. When the four context sensitivity measures were included in step two, none were significant predictors of 3 month symptom scores. An interaction detected in step three suggested that for individuals ($n=34$) with low hypermentalization scores ($z\text{-score} > -1$), having higher baseline symptoms led to higher 3 month symptoms, $b = .110, t = 2.35, p = .021$ (See Figure 1). Simple slope effects for average and high values of hypermentalization were not significant ($p > .05$).

Figure 1. Interaction between Pre-treatment and 3 Month Symptoms by level of Baseline Hypermentalization



Note: Plotted lines reflect three levels, one for scores a standard deviation or more below the mean (green), another for scores between -1 and 1 standard deviation around the mean (tan), and a third for scores a standard deviation above the mean (grey).

Effect of Treatment on Individual Difference Variables

Predicting Final Cue Presence Score

To examine the relationship of treatment on changes in Cue Presence scores, we used three step hierarchical multiple regression to control for associations between pre-treatment Cue Presence scores in step one, with pre-treatment symptom scores and change in symptoms scores after 3 months of treatment in step two, and the interaction between pre-treatment symptom severity and change in symptom scores in step three. All variables were centered prior to analysis to ensure low multicollinearity. Regression statistics are reported in Table 5.

Table 5. Linear model of symptom improvement predicting 3 Month Cue Presence.

	β	SE	r	sr^2	R	R^2	
Step 1						.545	.297
Constant		.627					
Cue Presence Baseline	.545***	.065	.545***	.297			
Step 2						.547	.299
Constant		.631					
Cue Presence Baseline	.547***	.065	.545***	.296			
Baseline Symptoms	.020	.081	.045	<.001			
Symptom Change	-.051	.078	-.009	.002			
Step 3						.554	.307
Constant		.633					
Cue Presence Baseline	.539***	.066	.545***	.286			
Baseline Symptoms	.009	.081	.045	<.001			

Symptom Change	.017	.094	-.009	<.001
Baseline Sxs X Sxs	-.110	.009	-.120	.008

Change

*p < .05, ** p < .01, *** p < .001

The regression model at step one reflects the expected association between pre- and 3 month scores for any repeated measure, $F(1,128) = 17.95, p < .001$. When baseline symptom severity and symptom improvement between pre- and 3 month scores were included in step two, none were significant predictors of 3 month Cue Presence scores. The interaction between baseline symptom severity and degree of symptom improvement pre- to 3 month scores was not a significant predictor of 3 month Cue Presence scores.

Predicting Final Cue Absence Score

To examine the relationship of treatment on changes in Cue Absence scores, we used three step hierarchical multiple regression to control for associations between pre-treatment Cue Absence scores in step one, with pre-treatment symptom scores and change in symptoms scores at 3 months in step two, and the interaction between pre-treatment symptom severity and change in symptom scores in step three. All variables were centered prior to analysis to ensure low multicollinearity. Regression statistics are reported in Table 6.

Table 6. Linear model of symptom improvement predicting 3 Month Cue Absence.

	β	SE	r	sr^2	R	R^2
Step 1					.409	.167
Constant		.564				
Cue Absence Baseline	.409***	.086	.409***	.167		

Step 2				.434	.189
Constant		4.389			
Cue Absence Baseline	.409***	.086	.409***	.166	
Baseline Symptoms	.008	.072	.035	<.001	
Symptom Change	.143	.069	.148	.017	
Step 3				.436	.190
Cue Absence Baseline	.409***	.087	.409***	.166	
Baseline Symptoms	.013	.073	.035	<.001	
Symptom Change	.115	.084	.148	.007	
Baseline Sxs X Sxs	.046	.008	.116	.001	

Change

* $p < .05$, ** $p < .01$, *** $p < .001$

The regression model at step one reflects the expected association between pre- and 3 month scores for any repeated measure, $F(1,128) = 9.76, p < .001$. When baseline symptom severity and symptom improvement between pre- and 3 months were included in step two, none were significant predictors of 3 month Cue Absence scores. The interaction between baseline symptom severity and degree of symptom improvement pre- to 3 months was not a significant predictor of 3 month Cue Absence scores.

Predicting Final Hypermentalization Score

To examine the relationship of treatment on changes in Hypermentalization scores, we used three step hierarchical multiple regression to control for associations between pre-treatment Hypermentalization scores in step one, with pre-treatment symptom scores and change in symptoms scores at 3 months in step two, and the interaction between pre-treatment symptom

severity and change in symptom scores in step three. All variables were centered prior to analysis to ensure low multicollinearity. Regression statistics are reported in Table 7.

Table 7. Linear model of symptom improvement predicting 3 month Hypermentalization.

	β	SE	r	sr^2	R	R^2
Step 1						.493
Constant		.265			.702	
Hyperment. Baseline	.702***	.061	.702***	.492		
Step 2						.515
Constant		.462			.717	
Hyperment. Baseline	.670***	.064	.702***	.408		
Baseline Symptoms	-.071	.035	-.135	.004		
Symptom Change	.168*	.033	.210*	.022		
Step 3						.515
Hyperment. Baseline	.670***	.064	.702***	.399		
Baseline Symptoms	-.071	.035	-.135	.004		
Symptom Change	.170*	.039	.210*	.016		
Baseline Sxs X Sxs	-.003	.004	.220*	>.001		

* $p < .05$, ** $p < .01$, *** $p < .001$

The regression model at step one reflects the expected association between pre- and 3 month scores for any repeated measure, $F(1,128) = 124.25, p < .001$. When baseline symptom severity and symptom improvement between pre- and 3 months were included in step two, Symptom Change was a significant predictor of 3 month Hypermentalization scores. The interaction between baseline symptom severity and degree of symptom improvement pre- to 3 months was not a significant predictor of 3 month Hypermentalization scores.

An alternative approach was also used that avoided the use of change scores when predicting change in Hypermentalization over time (see Table 8). This alternative used the residuals from Step 1 of the regression equation for predicting symptom outcome at 3 months (See Table 4, Step 1).

Table 8. Linear model of symptom improvement predicting 3 month Hypermentalization.

	β	<i>SE B</i>	<i>r</i>	<i>sr</i> ²	<i>R</i>	<i>R</i> ²
Step 1					.702	.493
Constant		.265				
Hyperment. Baseline	.702***	.061	.702***	.493		
Step 2					.719	.516
Constant		.260				
Hyperment. Baseline	.667***	.061	.702***	.423		
Residuals	-.083**	.033	-.306***	.024		

Predicting Final Hypomentalization Score

To examine the relationship of treatment on changes in Hypomentalization scores, we used three step hierarchical multiple regression to control for associations between pre-treatment Hypomentalization scores in step one, with pre-treatment symptom scores and change in symptoms scores at 3 months in step two, and the interaction between pre-treatment symptom severity and change in symptom scores in step three. All variables were centered prior to analysis to ensure low multicollinearity. Regression statistics are reported in Table 9.

Table 9. Linear model of symptom improvement predicting 3 month Hypomentalization.

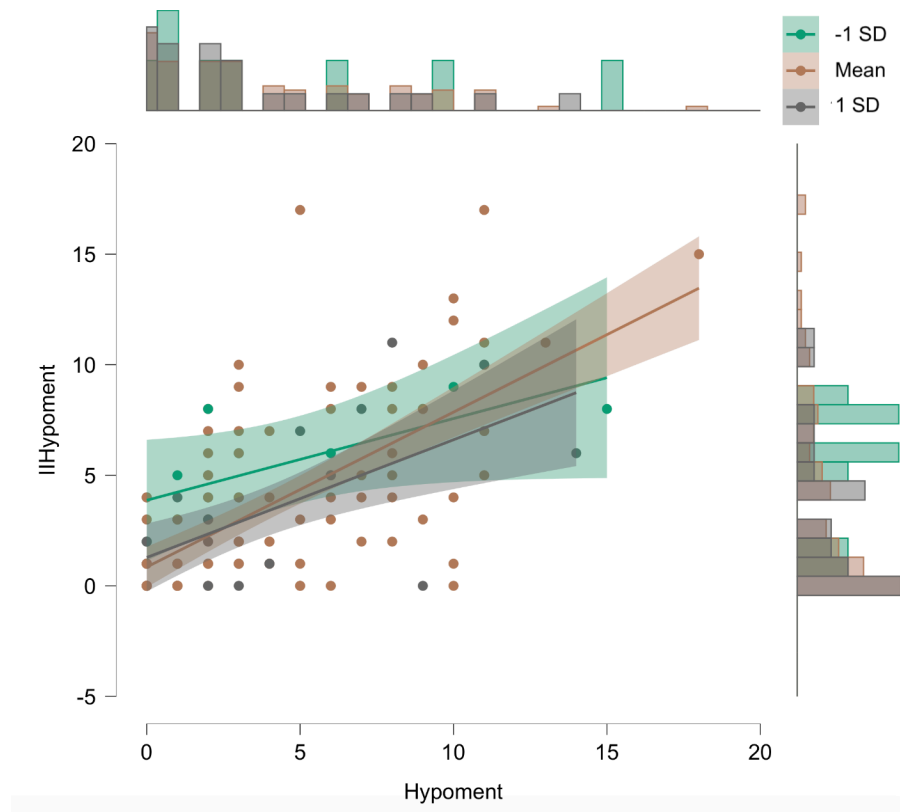
	β	<i>SE B</i>	<i>r</i>	<i>sr</i> ²	<i>R</i>	<i>R</i> ²	
Step 1						.645	.416
Constant		.263					
Hypoment. Baseline	.645***	.067	.645***	.416			
Step 2						.677	.458
Constant		.255					
Hypoment. Baseline	.588***	.068	.645***	.317			
Baseline Symptoms	.192**	.034	.236**	.028			
Symptom Change	-.203**	.032	-.164*	.033			
Step 3						.687	.472
Hypoment. Baseline	.567***	.069	.645***	.287			
Baseline Symptoms	.183**	.034	.236**	.025			
Symptom Change	-1.116	.038	-.164*	.007			
Baseline Sxs X Sxs	-.149*	.004	-.298***	.014			
Change							

* $p < .05$, ** $p < .01$, *** $p < .001$

The regression model at step one reflects the expected association between pre- and 3 month scores for any repeated measure, $F(1,128) = 91.22, p < .001$. When baseline symptom severity and symptom improvement between pre- and 3 months were included in step two, both were significant predictors of 3 month Hypomentalization scores in the expected direction. The interaction between baseline symptom severity and degree of symptom improvement pre- to 3 months was trending to predict final Hypomentalization scores and explained a statistically significant, though small proportion of variance in 3 month Hypomentalization scores, $\Delta R^2 = .014, F(4, 125) = 27.96, p < .001$. The interaction suggested that individuals ($n=24$) reporting the most symptom improvement ($z\text{-score} > 1$) pre-treatment to 3 months had lower 3 month

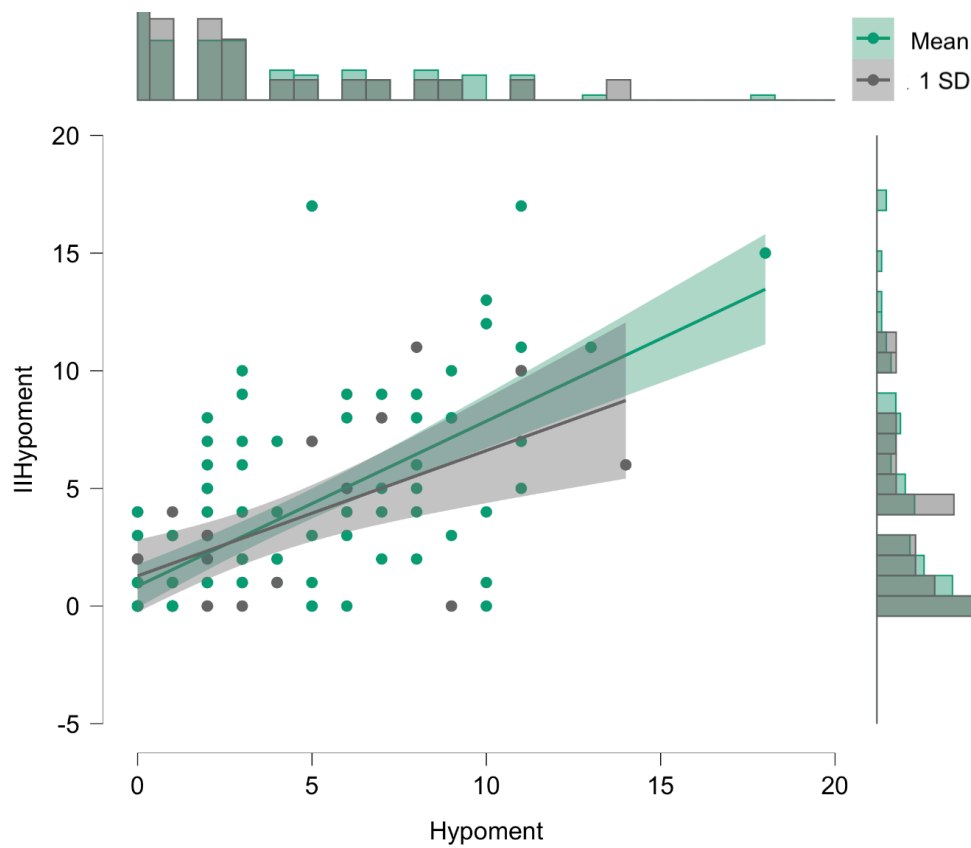
Hypomentalization scores, $b = -.278$, $t = -1.35$, $p = .012$ (See Figure 2). Simple slope effects for average and low values of symptom change were not significant ($p > .10$). To make the interaction easier to see, an additional graph was generated including only the two largest classes, mean scores and scores greater than one standard deviations above the mean (See Figure 3).

Figure 2. Interaction between Pre-treatment Hypomentalization and 3 Month Hypomentalization by all three levels of Symptom change



Note: Plotted lines reflect three levels, one for scores a standard deviation or more below the mean (green), another for scores between -1 and 1 standard deviation around the mean (tan), and a third for scores a standard deviation above the mean (grey).

Figure 3. Interaction between Pre-treatment Hypomentalization and 3 Month Hypomentalization with two levels of Symptom change



Note: Plotted lines reflect two levels, one for scores between -1 and 1 standard deviation around the mean (tan), and a second for scores a standard deviation above the mean (grey).

An alternative approach was also used that avoided the use of change scores when predicting change in individual difference variables (see Table 10). This alternative used the residuals from Step 1 of the regression equation for predicting symptom outcome at 3 months (See Table 4, Step 1).

Table 10. Linear model of symptom improvement predicting 3 month Hypomentalization.

	β	<i>SE B</i>	<i>r</i>	<i>sr</i> ²	<i>R</i>	<i>R</i> ²
Step 1					.645	.416
Constant		.263				
Hypoment. Baseline	.645***	.067	.645***	.416		
Step 2					.669	.447
Constant		.255				
Hypoment. Baseline	.611***	.067	.645***	.361		
Residuals	.087**	.033	.293***	.031		

Discussion

Individual differences in emotion regulation may be an important route for explaining the finding that much of outcome in psychotherapy is accounted for by characteristics that patients bring with them to the treatment setting. In beginning to explore the impact of emotion regulation abilities on treatment outcome, we focused on context sensitivity as an important first step in the emotion regulation (Bonanno & Burton, 2013) process that may have a significant relationship with pre-treatment severity and 3-month improvement. We selected two instruments, the Context Sensitivity Index (CSI; Bonanno, Maccallum, Malgaroli, & Hou, 2020) and the Reflective Functioning Questionnaire (RFQ; Fonagy et al., 2016) to capture multiple aspects of context sensitivity that could be easily distributed to individuals receiving care on a digital therapy platform. Finally, emotion regulation is a common target of psychotherapy and may influence therapy effects while also being influenced by successful therapy. We designed instrument deployment to match the timing of context sensitivity measures with symptom scales to look for changes in clinical and emotion regulation constructs pre- and at 3 months of treatment.

In this study, we found half of the sample reported clinically significant change and about 40% reported no cardinal symptoms and full remission at 3 months with symptom scores less than five. The expected convergent relationship between the RFQ and CSI was not detected. Higher Openness and Neuroticism scores were associated with greater symptom severity at baseline, however we did not find the hypothesized association between higher baseline severity and either poor reflective functioning or lower context sensitivity.

In turning to reflective functioning and context sensitivity as indicators of outcome in digital psychotherapy, we found an unexpected association between baseline lower

Hypermentalization and worse outcomes at 3 months. No other baseline scales for context sensitivity or reflective functioning were associated with outcomes at 3 months. We also expected greater symptom improvement to be associated with improvements in context sensitivity and reflective functioning. While we did not find an association between context sensitivity reported at 3 months and symptom improvement, associations were found for both reflective functioning constructs. Specifically, greater symptom improvement was surprisingly associated with worse hypermentalization at 3 months, whereas greater symptom improvements were associated with improved hypomentalization, especially when baseline symptoms were high.

Results for symptom measures support previous findings on the effectiveness of the digital platform for reducing symptoms among individuals presenting with depression and anxiety diagnoses (DellaCrosse, Mahan, & Hull, 2018; Hull & Mahan, 2017; Malgaroli, Hull, Connolly, Feuerstein, & Simon, 2020). The association between the number of words written by the patient as part of the treatment and improvement was also found in this sample, congruent with previous work, though overall engagement, as measured by word count, was about 40% lower than has been previously reported (Malgaroli, Hull, Connolly, Feuerstein, & Simon, 2020). One possible explanation is that many of the participants began treatment right before or around the beginning of the novel coronavirus disease 2019 (COVID-19) pandemic as it was being addressed by physical distancing and shutdown measures in the United States. The significant adjustment required during this time may have made it difficult to engage with treatment as seen previously.

Previous work has found associations between both context sensitivity and depression and anxiety (Bonanno, Maccallum, Malgaroli, & Hou, 2020), and reflective functioning and

depression and anxiety (Fonagy et al., 2016). Context sensitivity and reflective functioning also share a theoretical interest in how individuals attend to cues in the environment that are relevant for self-regulation. We therefore expected some modest overlap between the subscales of the CSI and RFQ, which measure these constructs. In looking specifically at the subscales of each measure, we predicted that the Cue Presence aspect of the CSI would be negatively correlated with Hypermentalization, given that Cue Presence reflects accurate identification of cues in a context, whereas Hypermentalization is the imposition of assumptions on a context by the individual. Similarly, we expected that Cue Absence would be inversely related to Hypomentalization since Cue Absence is the recognition that needed interpretive information is missing, whereas Hypomentalization reflects uncertainty regarding one's own feelings and the feelings of others. Cue Absence and Hypomentalization were also previously associated with symptoms of depression and anxiety. Contrary to expectations, these postulated aspects of context sensitivity bore no relationship to each other. Instead, a small and unexpected relationship between Cue Absence and Hypermentalization was found only for the second administration at 3 months of treatment, though when making the appropriate adjustments for multiple comparisons, this association appears to be spurious.

A critical question is why these measures did not converge as expected. One possibility is that the CSI measures the outcomes of an implicit judgment process by asking respondents to give ratings of their sense of control, urgency, and cooperation in the given context. It does not ask about the judgment process itself. For example, whether they missed a social cue, assumed an erroneous social cue, or failed to account for aspects of their emotional reaction to a feature of the context. This is one reason why performance-based measures like the CSI must be evaluated against norms for how people tend to respond. While this offers the advantages of performative

measures, such that items are causal indicators instead of effect indicators (cf. Bonanno, Maccallum, Malgaroli, & Hou, 2020), it leaves the concrete details of the performance itself unspecified. In other words, the CSI can tell you how well someone performed relative to empirically derived norms, but it cannot tell you how or why they performed the way they did. Since the types of cues relevant for successful CSI performance are unknown, it is difficult to determine *a priori* when an instrument measuring particular cues will be associated with CSI performance. It could be that the RFQ identifies cues that are simply unrelated to the ways in which people make judgments of the CSI scenarios, even if they could be thought of as potentially relevant contextual cues on their face. Identifying convergent measures that specify particular cue types offers an intriguing possibility for future research into how judgments of CSI scenarios are made.

Another possible reason the CSI and RFQ were unrelated is that the CSI is designed to measure reactions to stressors, whereas the RFQ is a more global measure of how well individuals are able to evaluate their own feelings and the thoughts of others. The domain of application may be separate enough that even though both have been associated with depression and anxiety, they are associated with these symptoms in very different ways. The RFQ reflects one's understanding of and the ability to report on general dispositions of how they relate to themselves and to others. Whereas CSI scores reflect one's implicit formation of judgments related to control, urgency, and cooperation for given stressful scenarios. Perhaps the dispositions of the RFQ have little bearing on one's sense of control, urgency, or cooperation when facing a stressor.

We also found a relatively strong inverse relationship between Cue Presence and Cue Absence, contrary to the positive relationship previously reported. This difference may be

another reason for the lack of relationship between CSI and RFQ measures. There was no evidence of monotonic responding in the CSI responses, nor any for unusually extreme sum scores. This is the first time that the CSI has been deployed among a clinical, treatment seeking population and may pull for different types of responses than for the population it was normed on. As alluded to above, stressors may be a more common contributor to reported depression and anxiety symptoms for a nonclinical population than for a clinical one and this may pull the intuitions of CSI respondents in different ways. Another possibility is that this sample reported higher than usual Hypermentalization scores, suggesting that this sample has a greater tendency to attribute unambiguous mental states to others than is usual. Indeed, Hypermentalization scores in this sample are double that reported in Fonagy et al (2016) and this may have led participants to make more unusual judgments on the CSI than seen previously.

When looking at associations between individual differences and baseline symptom severity, we found the expected relationships between Extraversion and lower baseline symptoms, and Neuroticism and higher baseline symptoms. The positive association between Openness and greater symptom severity is the opposite of that typically reported (Bleidorn et al., 2019). Several Openness items ask about interest in activities involving travel, going to public places, or seeking stimulation outside of the home. It is possible that COVID-19 restrictions may impact individuals with these kinds of interests more than others, which could be driving this association. It could also be that individuals with higher Openness are attracted to the platform, given the novel nature of this type of treatment, and so are more represented among this sample. CSI and RFQ scores were not associated with baseline symptom severity, contrary to expectations. This remained true even when removing the five-factor dimensions from the model. This lack of replication may reflect a difference in measuring participant dispositions at

the beginning of treatment rather than in the course of daily living. The decision to enter treatment is a combination of one's history with depression and anxiety, the timing of factors in one's living situation and recent, acute events (Pescosolido, Gardner, & Lubell, 1998; Jameson & Robert, 2010). These additional factors may contribute to baseline symptom scores far beyond factors of emotion regulation and context sensitivity. Controlling for an individual's circumstance upon entering treatment may help to isolate the contribution of context sensitivity and reflective functioning relative to baseline symptoms in future research.

In analyzing the association between RFQ scores and treatment outcome, lower Hypermentalization was related to worse outcomes, with an interaction suggesting that when Hypermentalization was low, 3 month symptoms were more likely to be high when baseline symptoms were high. In addition, the residual model suggested that symptom improvement was associated with worsening Hypermentalization. This is not easy to account for given that higher Hypermentalization has been previously associated with greater psychopathology (Fonagy et al., 2016). Since those with low Hypermentalization in this sample were the least likely to improve, it may be that the therapeutic relationship was especially important for individuals reporting high levels of Hypermentalization. However, individuals with high Hypermentalization typically struggle to relate to others or utilize relationships effectively. Perhaps the messaging platform provides additional social distance from the therapist and this enables these patients to make more progress or stick with therapy compared to face-to-face settings. Another possibility is that by reducing depression and anxiety symptoms, troubling hypermentalization tendencies come more to the fore for patients. Further research is needed to fully understand this association.

Results for the relationship between symptom reduction in treatment and changes in RFQ scores suggest that Hypomentalization is the most malleable of the constructs, especially when

baseline symptoms are more severe and the patient experiences more change in treatment. This supports the previously reported association between Hypomentalization and depression and anxiety symptoms, but goes further in suggesting that therapy may serve to increase one's ability to be aware of one's own feelings and thoughts, as well as the feelings and thoughts of others as measured by the Hypomentalization subscale. Virtually every form of psychotherapy seeks to improve these capacities (Cuijpers, Reijnders, & Huibers, 2019). Nevertheless, the relationship is not a straightforward one, only being present in the extremes. This raises two possibilities. One, that focusing on cognitive and emotion awareness is especially important for individuals with severe depression and anxiety, and two, that the Hypomentalization measure simply lacks sensitivity when cases are less severe. Future research utilizing therapy process measures like hand coded rubrics or natural language processing might be able to better determine the true utility of awareness practices on outcome and changes in Hypomentalization.

Lastly, a critical question is why the CSI did not bear any of the expected associations. We have offered several specific possibilities above, including clinical sample, unusually high presence of Hypermentalization, the complex factors that bring individuals to treatment over those reporting depression and anxiety symptoms in a community sample, and the possibility that context sensitivity may change slowly, if at all, in response to treatment. An additional general consideration is that context sensitivity on its own may be insufficient for determining the role of emotion regulation in psychotherapy. The Regulatory Flexibility approach (Bonanno & Burton, 2013) identifies two additional processes, repertoire and feedback, as equally important in understanding how emotion regulation events unfold. It is possible that strengths in one's repertoire or ability to adjust to feedback may make up for weaknesses in context sensitivity or

that individuals have other ways of compensating that make isolating context sensitivity as an emotion regulation factor inadvisable when investigating clinical populations.

Limitations and Future Directions

While our findings support previously reported clinical outcomes, engagement metrics, and some relationship between RFQ scores and outcome, there are limitations to this study. First, investigating individual differences with a sample of this size severely limits the power of the analyses and may leave certain associations unidentified while exaggerating the influence of others. In addition, the exploratory nature of the analyses will tend to capitalize on multiple comparisons. For example, the weak, but statistically significant association between Cue Absence and Hypermentalization at 3 months was neither predicted from theory nor from previous empirical work. The association between Openness and baseline symptoms is similar. If *p* values were corrected for multiple comparisons in that case, it would no longer be significant. This limitation may also apply to the regression models conducted, especially where effects were the reverse of those expected, such as lower Hypermentalization scores predicting greater symptoms at 3 months. Finally, the unique historical event of COVID-19 emerged halfway through recruitment of participants into this study. It is difficult to fully account for the impact of this event on the findings, except to note that engagement was lower than typical, and that the study had to be prematurely concluded to enable the digital platform to adjust to increased demand and greater severity of cases presenting for care (Hull et al., 2021). As a result of these limitations, it may be premature to make determinations regarding whether the measures of context sensitivity employed here have utility in accounting for important patient characteristics in psychotherapy or not. Future research utilizing the advantages of digital platforms for recruiting larger samples will be important for furthering investigating the impact of individual differences in emotion regulation on therapy outcome.

Disclosures: Hull is an employee of the platform that provided the data for this study.

References

- Aldao, A. (2013). The future of emotion regulation research: Capturing context. *Perspectives on Psychological Science*, 8(2), 155-172.
- Alleman, J. R. (2002). Online counseling: The Internet and mental health treatment. *Psychotherapy: Theory/Research/Practice/Training*, 39, 2, 199–209. doi:10.1037/0033-3204.39.2.199
- Bagby, R.M., Quilty, L.C., & Ryder, A.C. (2008). Personality and depression. *Canadian Psychiatric Review*, 53(1), 14-25.
- Barth, J., Munder, T., Gerger, H., Nüesch, E., Trelle, S., Znoj, H., ... & Cuijpers, P. (2016). Comparative efficacy of seven psychotherapeutic interventions for patients with depression: a network meta-analysis. *Focus*, 14(2), 229-243.
- Bee, P. E., Bower, P., Lovell, K., Gilbody, S., Richards, D., Gask, L., & Roach, P. (2008). Psychotherapy mediated by remote communication technologies: a meta-analytic review. *BMC psychiatry*, 8, 1, 60. doi: 10.1186/1471-244X-8-60
- Bleidorn, W., Hill, P. L., Back, M. D., Denissen, J. J., Hennecke, M., Hopwood, C. J., ... & Orth, U. (2019). The policy relevance of personality traits. *American psychologist*, 74(9), 1056-1067.
- Bonanno, G.A., Colak, D. M., Keltner, D., Shiota, M. N., Papa, A., Noll, J. G., ... Trickett, P. K. (2007). Context matters: the benefits and costs of expressing positive emotion among survivors of childhood sexual abuse. *Emotion (Washington, D.C.)*, 7(4), 824–837. doi:10.1037/1528-3542.7.4.824
- Bonanno, G.A., Papa, A., Lalande, K., Westphal, M., & Coifman, K. (2004). The importance of being flexible: The ability to both enhance and suppress emotional expression predicts long-term adjustment. *Psychological Science*, 15(7), 482–487. doi:10.1111/j.0956-7976.2004.00705.x
- Bonanno, G. a., Pat-Horenczyk, R., & Noll, J. (2011). Coping flexibility and trauma: The Perceived Ability to Cope With Trauma (PACT) scale. *Psychological Trauma: Theory, Research, Practice, and Policy*, 3(2), 117–129. doi:10.1037/a0020921
- Bonanno, G.A., & Burton, C.L. (2013). Regulatory flexibility: An individual differences perspective on coping and emotion regulation. *Perspectives on Psychological Science*, 8(6), 591-612.
- Bonanno, G. A., Maccallum, F., Malgaroli, M., & Hou, W. K. (2020). The Context Sensitivity

- Index (CSI): Measuring the Ability to Identify the Presence and Absence of Stressor Context Cues. *Assessment*, 27(2), 261-273.
- Brody, D.S., Khaliq, A.A., & Thompson, T.L., (1997). Patients' perspectives on the management of emotional distress in primary care settings. *J Gen Intern Med*, 12, 403-406. doi: 10.1046/j.1525-1497.1997.00070.x
- Bucher, M.A., Suzuki, T., & Samuel, D.B. (2019). A meta-analytic review of personality traits and their associations with mental health treatment outcomes. *Clinical Psychology Review*, 70, 51-63.
- Buss, K.A., Davidson, R.J., Kalin, N.H., & Goldsmith, H.H. (2004). Context-specific freezing and associated physiological reactivity as a dysregulated fear response. *Developmental Psychology*, 40(4), 583.
- Carver, C. S., & Scheier, M. F. (2008). Feedback processes in the simultaneous regulation of action and affect. *Handbook of motivation science*, 308-324.
- Chatterton, M. L., Stockings, E., Berk, M., Barendregt, J. J., Carter, R., & Mihalopoulos, C. (2017). Psychosocial therapies for the adjunctive treatment of bipolar disorder in adults: network meta-analysis. *The British Journal of Psychiatry*, 210(5), 333-341.
- Cheng, C., Chiu, C. Y., Hong, Y. Y., & Cheung, J. S. (2001). Discriminative facility and its role in the perceived quality of interactional experiences. *Journal of personality*, 69(5), 765-785.
- Coifman, K.G., & Bonanno, G.A. (2010). When distress does not become depression: emotion context sensitivity and adjustment to bereavement. *Journal of Abnormal Psychology*, 119(3), 479.
- Comer, J. S., & Barlow, D. H. (2014). The occasional case against broad dissemination and implementation: Retaining a role for specialty care in the delivery of psychological treatments. *American Psychologist*, 69(1), 1.
- Cristea, I. A., Gentili, C., Cotet, C. D., Palomba, D., Barbui, C., & Cuijpers, P. (2017). Efficacy of psychotherapies for borderline personality disorder: a systematic review and meta-analysis. *JAMA Psychiatry*, 74(4), 319-328.
- Cuijpers, P., Driessen, E., Hollon, S. D., van Oppen, P., Barth, J., & Andersson, G. (2012). The efficacy of non-directive supportive therapy for adult depression: a meta-analysis. *Clinical psychology review*, 32(4), 280-291.
- Cuijpers, P., Sijbrandij, M., Koole, S., Huibers, M., Berking, M., & Andersson, G. (2014). Psychological treatment of generalized anxiety disorder: a meta-analysis. *Clinical psychology review*, 34(2), 130-140.
- Cuijpers, P., Donker, T., Weissman, M. M., Ravitz, P., & Cristea, I. A. (2016). Interpersonal

- psychotherapy for mental health problems: a comprehensive meta-analysis. *American Journal of Psychiatry*, 173(7), 680-687.
- Cuijpers, P., Reijnders, M., & Huibers, M.J.H. (2019). The role of common factors in psychotherapy outcomes. *Annu. Rev. Clin. Psychol.*, 15, 207-231.
- Davidson, R.J., Jackson, D.C., & Kalin, N.H. (2000). Emotion, plasticity, context, and regulation: perspectives from affective neuroscience. *Psychological Bulletin*, 126(6), 890.
- de Jong, K., & DeRubeis, R. J. (2018). Effectiveness of psychotherapy. In J. N. Butcher & J. M. Hooley (Eds.), *APA handbooks in psychology®. APA handbook of psychopathology: Psychopathology: Understanding, assessing, and treating adult mental disorders* (p. 705–725). American Psychological Association.
- DellaCrosse, M., Mahan, K., & Hull, T.D. (2018). The effect of messaging therapy for depression and anxiety on employee productivity. *Journal of Technology in Behavioral Science*, 4(1), 1-5. DOI 10.1007/s41347-018-0064-4
- DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 Aspects of the Big Five. *Journal of Personality and Social Psychology*, 93, 880-896.
- Eysenck, H.J. (1952). The effects of psychotherapy: an evaluation. *Journal of Consulting Psychology*, 16(5), 319-324.
- Fjeldsoe, B. S., Marshall, A. L., & Miller, Y. D. (2009). Behavior change interventions delivered by mobile telephone short-message service. *American journal of preventive medicine*, 36, 2, 165-173. doi:10.1016/j.amepre.2008.09.040
- Fonagy, P., Luyten, P., Moulton-Perkins, A., Lee, Y., Warren, F., Howard, S., Ghinai, R., Fearon, P., & Lowyck, B. (2016). Development and validation of a self-report measure of mentalizing: The reflective functioning questionnaire. *PLoS ONE*, 11(7), 1-28.
- Gehricke, J.G., & Shapiro, D. (2000). Reduced facial expression and social context in major depression: discrepancies between facial muscle activity and self-reported emotion. *Psychiatry Research*, 95(2), 157-167.
- Godfrey, K. M., Gallo, L. C., & Afari, N. (2015). Mindfulness-based interventions for binge eating: A systematic review and meta-analysis. *Journal of behavioral medicine*, 38(2), 348-362.
- Gruber, J., Johnson, S. L., Oveis, C., & Keltner, D. (2008). Risk for mania and positive emotional responding: too much of a good thing? *Emotion*, 8(1), 23–33.
- Hollon, S., Munoz, R.F., Barlow, D.H., Beardslee, W.R., Bell, C.C., Bernal, G., Clarke, G.N., Franciosi, L.P., Kazdin, A.E., Kohn, L., Linehan, M.M., Markowitz, J.C., Persons, J.B., Niederehe, G., & Sommers, D. (2002). Psychosocial intervention development for the

- prevention and treatment of depression: promoting innovation and increasing access. *Biological psychiatry*, 52, 610-630. doi: 10.1016/S0006-3223(02)01384-7.
- Hull, T.D., Levine, J., Bantilan, N., Desai, A., & Majumder, M. (2021). Analyzing digital evidence from a telemental health platform to assess complex psychological responses to the COVID-19 pandemic: Content analysis of text messages. *JMIR: Formative Research*, 5(2), 1-8.
- Hull, T.D., & Mahan, K., (2017). A Study of Asynchronous Mobile-Enabled SMS Text Psychotherapy. *Telemedicine and e-Health* 23, 3, 240-247. doi:10.1089/tmj.2016.0114.
- Hull, T.D., Malgaroli, M., Connolly, P., Feuerstein, S., & Simon, N.M. (2020). The treatment effectiveness of asynchronous text therapy for depression and anxiety: A naturalistic observation study. *BMC Psychiatry*, 20(1), 1-12.
- Malgaroli, M., Hull, T.D., Wiltsey-Stirman, S., & Resick, P. (2020). Message delivery of treatment for PTSD: A longitudinal observational study of symptoms trajectories. *Journal of Medical Internet Research*. Doi: 10.2196/15587
- Jacobson, N. S., & Truax, P. (1992). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59(1), 12-16.
- Kennedy, S. H. (2008). Core symptoms of major depressive disorder: relevance to diagnosis and treatment. *Dialogues in clinical neuroscience*, 10(3), 271.
- Kessler, D., Lewis, G., Kaur, S., Wiles, N., King, M., Weich, S., Sharp, D., Araya, R., Hollinghurst, S., & Peters, T. J. (2009). Therapist-delivered internet psychotherapy for depression in primary care: A randomised controlled trial. *The Lancet*, 374, 9690, 628-634. doi:10.1016/s0140-6736(09)61257-5
- Kring, A. M., & Sloan, D. M. (Eds.). (2009). *Emotion regulation and psychopathology: A transdiagnostic approach to etiology and treatment*. Guilford Press.
- Kroenke K, Spitzer RL, Williams JB. (2003). The Patient Health Questionnaire-2: Validity of a Two-Item Depression Screener. *Medical Care*, 41, 1284-92.
- Kroenke, K., & Spitzer, R. L. (2002). The PHQ-9: A new depression diagnostic and severity measure. *Psychiatric annals*, 32(9), 509-515.
- Kroenke, K., West, S. L., Swindle, R., Gilsonan, A., Eckert, G. J., Dolor, R., ... & Weinberger, M. (2001). Similar effectiveness of paroxetine, fluoxetine, and sertraline in primary care: A randomized trial. *JAMA*, 286(23), 2947-2955.
- Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16(9), 606-613.

- Kroenke, K., Spitzer, R. L., Williams, J. B., Monahan, P. O., & Löwe, B. (2007). Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. *Annals of Internal Medicine*, *146*(5), 317-325.
- Lambert, M.J. (1992). Psychotherapy outcome research: Implications for integrative and eclectic therapists. In *Handbook of Psychotherapy Integration* (Norcross, J.C., Goldfield, M.R., Eds.), 94-129. New York: Basic Books.
- Lewis, M., Haviland-Jones, J. M., & Barrett, L. F. (Eds.). (2010). *Handbook of emotions*. Guilford Press.
- Levy-Gigi, E., Bonanno, G. A., Shapiro, A. R., Richter-Levin, G., Keri, S., & Sheppes, G. (2015). Emotion Regulatory Flexibility Sheds Light on the Elusive Relationship Between Repeated Traumatic Exposure and Posttraumatic Stress Disorder Symptoms. *Clinical Psychological Science*, 1–12. doi:10.1177/2167702615577783
- Mennin, D.S., Heimberg, R.G., Turk, C.L., & Fresco, D.M. (2005). Preliminary evidence for an emotion dysregulation model of generalized anxiety disorder. *Behaviour Research and Therapy*, *43*(10), 1281-1310.
- McCrae, R.R., & Costa, P.T. (1996). Toward a new generation of personality theories: Theoretical contexts for the five-factor model. Wiggins, Jerry S. (Ed.). *The five-factor model of personality: Theoretical perspectives* (pp. 51-87). New York: Guilford Press.
- Militello, L. K., Kelly, S. A., & Melnyk, B. M. (2012), Systematic Review of Text-Messaging Interventions to Promote Healthy Behaviors in Pediatric and Adolescent Populations: Implications for Clinical Practice and Research. *Worldviews on Evidence-Based Nursing*, *9*, 66–77. doi:10.1111/j.1741-6787.2011.00239.x
- Mohr, D. C., Hart, S. L., Howard, I., Julian, L., Vella, L., Catledge, C., & Feldman, M. D. (2006). Barriers to psychotherapy among depressed and nondepressed primary care patients. *Annals of Behavioral Medicine*, *32*, 3, 254-258. doi:10.1207/s15324796abm3203_12
- Mohr, D. C., Ho, J., Hart, T. L., Baron, K. G., Berendsen, M., Beckner, V., ... & Schroder, K. E. (2014). Control condition design and implementation features in controlled trials: a meta-analysis of trials evaluating psychotherapy for depression. *Translational behavioral medicine*, *4*(4), 407-423.
- Nathan, P.E., Stuart, S.P., & Dolan, S.L. (2000). Research on psychotherapy efficacy and effectiveness: Between Scylla and Charybdis?. *Psychological bulletin*, *126*(6), 964.
- Nelson, E. L., Barnard, M., & Cain, S. (2003). Treating Childhood Depression Over Videoconferencing. *Telemedicine Journal and e-Health*, *9*, 49-55. doi:10.1089/153056203763317648

- Nutting, P.A., Rost, K., Dickinson, M., Werner, J.J., Dickinson, P., Smith, J.L., & Gallovic, B. (2002). Barriers to initiating depression treatment in primary care practice. *J Gen Intern Med, 17*, 103-111. doi: 10.1046/j.1525-1497.2002.10128.x
- Olatunji, B. O., Davis, M. L., Powers, M. B., & Smits, J. A. (2013). Cognitive-behavioral therapy for obsessive-compulsive disorder: A meta-analysis of treatment outcome and moderators. *Journal of Psychiatric Research, 47*(1), 33-41.
- Oquendo, M. A., Halberstam, B., & Mann, J. J. (2003). Risk factors for suicidal behavior. Standardized evaluation in clinical practice, *22*, 103-129.
- Oud, M., Mayo-Wilson, E., Braidwood, R., Schulte, P., Jones, S. H., Morriss, R., ... & Kendall, T. (2016). Psychological interventions for adults with bipolar disorder: systematic review and meta-analysis. *The British Journal of Psychiatry, 208*(3), 213-222.
- Penton-Voak, I. S., Thomas, J., Gage, S. H., McMurrin, M., McDonald, S., & Munafò, M. R. (2013). Increasing recognition of happiness in ambiguous facial expressions reduces anger and aggressive behavior. *Psychological Science, 24*(5), 688-697.
- Pescosolido, B. A., Gardner, C. B., & Lubell, K. M. (1998). How people get into mental health services: Stories of choice, coercion and “muddling through” from “first-timers”. *Social science & medicine, 46*(2), 275-286.
- Reynolds, D. A. J., Stiles, W. B., & Grohol, J. M. (2006). An investigation of session impact and alliance in internet based psychotherapy: Preliminary results. *Counselling and Psychotherapy Research, 6*, 3, 164-168. doi: 10.1080/14733140600853617
- Rottenberg J, & Gotlib I.H. (2004) Socioemotional functioning in depression. In M. Power (Ed.) *Mood disorders: A handbook of science and practice* (pp. 61-77). New York: Wiley.
- Rottenberg, J., Gross, J.J., & Gotlib, I.H. (2005). Emotion context insensitivity in major depressive disorder. *Journal of Abnormal Psychology, 114*(4), 627.
- Rottenberg, J., Kasch, K. L., Gross, J. J., & Gotlib, I. H. (2002). Sadness and amusement reactivity differentially predict concurrent and prospective functioning in major depressive disorder. *Emotion, 2*(2), 135.
- Scheppers, E., Van Dongen, E., Dekker, J., Geertzen, J., & Dekker, J. (2006). Potential barriers to the use of health services among ethnic minorities: a review. *Family practice, 23*, 3, 325-348. doi:10.1093/fampra/cmi113
- Sheppes, G., Suri, G., & Gross, J. J. (2015). Emotion Regulation and Psychopathology. *Annual Review of Clinical Psychology, 11*, 379-405.
- Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing

- generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine*, 166(10), 1092-1097.
- Troy, A. S., Shallcross, A. J., & Mauss, I. B. (2013). A person-by-situation approach to emotion regulation: cognitive reappraisal can either help or hurt, depending on the context. *Psychological Science*, 24(12), 2505–14. doi:10.1177/0956797613496434
- Velthorst, E., Koeter, M., Van Der Gaag, M., Nieman, D. H., Fett, A. K., Smit, F., ... & De Haan, L. (2015). Adapted cognitive-behavioural therapy required for targeting negative symptoms in schizophrenia: meta-analysis and meta-regression. *Psychological medicine*, 45(3), 453-465.
- Wampold, B. E. (2001). *The great psychotherapy debate: Models, methods, and findings*. Mahwah, NJ: Lawrence Erlbaum.
- Wampold, B. E. (2015). How important are the common factors in psychotherapy? An update. *World Psychiatry*, 14(3), 270-277.
- Wampold, B. E. & Imel, Z.E. (2015). *The great psychotherapy debate: The evidence for what makes psychotherapy work (2nd Ed.)*. New York: Routledge.
- Young, A., Klapp, R., Sherbourne, C., & Wells, K. (2001). The quality of care for depressive and anxiety disorders in the United States. *Arch Gen Psychiatry*, 58, 55-61. doi:10.1001/archpsyc.58.1.55
- Young, K. S. (2005). An empirical examination of client attitudes towards online counseling. *CyberPsychology and Behavior*, 8, 172–177. doi: 10.1089/cpb.2005.8.172
- Zinbarg, R.E., Uliaszek, A.A., & Adler, J.M. (2008). The role of personality in psychotherapy for anxiety and depression. *Journal of Personality*, 76(6), 1649-1688.

Appendix A

Table 11. Associations between Gender, Symptom Severity, and Individual Differences

Score	Female (n = 82)	Male (n = 13)	Other (n = 8)
Symptoms			
Baseline Symptoms	19.0 (9.1)	22.4 (9.9)	22.1 (4.9)
3 Month Symptoms	13.5 (9.5)	11.5 (6.6)	19.0 (10.8)
CSI			
Baseline Cue Pres	51.6 (9.8)	51.5 (10.1)	53.0 (9.7)
3 Month Cue Pres	52.9 (8.4)	56.7 (9.9)	56.6 (8.1)
Baseline Cue Abs	50.1 (6.8)	53.2 (5.3)	46.9 (5.1)
3 Month Cue Abs	50.3 (7.1)	50.5 (8.7)	47.4 (7.1)
RFQ			
Baseline Hyperm.	6.1 (4.7)	5.4 (4.1)	3.8 (2.1)
3 Month Hyperm.	6.1 (4.4)	5.8 (4.1)	3.3 (3.2)
Baseline Hypom.	4.1 (4.2)	3.23 (4.3)	4.8 (3.4)
3 Month Hypom.	3.8 (3.9)	2.8 (3.3)	6.0 (3.9)