Supporting Electronic Mental Health with Artificial Intelligence
Thought Record Analysis and Guidance

Burger, Franziska

DOI
10.4233/uuid:6aed720a-b970-43c2-9c84-f028a8127230

Publication date
2022

Document Version
Final published version

Citation (APA)
Burger, F. (2022). *Supporting Electronic Mental Health with Artificial Intelligence: Thought Record Analysis and Guidance*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:6aed720a-b970-43c2-9c84-f028a8127230

Supporting Electronic Mental Health for Depression with Artificial Intelligence

Thought Record Analysis and Guidance

Franziska Burger

Supporting Electronic Mental Health for Depression with Artificial Intelligence

Thought Record Analysis and Guidance

Supporting Electronic Mental Health for Depression with Artificial Intelligence

Thought Record Analysis and Guidance

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus, prof. dr. ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates
to be defended publicly on
Thursday, 15 December 2022 at 12:30 o'clock

by

Franziska Viktoria Burger

Master of Science in Artificial Intelligence,
Radboud University Nijmegen, The Netherlands,

born in Waiblingen, Germany.

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus,	chairman
Prof. dr. M.A. Neerincx,	Delft University of Technology, promotor
Dr. ir. W.-P. Brinkman,	Delft University of Technology, promotor
Independent members:	
Prof. dr. C.M. Jonker,	Delft University of Technology
Prof. dr. ir. G.J.P.M. Houben,	Delft University of Technology
Prof. dr. T. Belpaeme,	Ghent University
Prof. dr. E.J. Krahmer,	Tilburg University
Prof. dr. M.M. Specht	Delft University of Technology, reserve

Keywords: computerized therapy, conversational agents, natural language processing, cognitive therapy

Printed by: ProefschriftMaken | www.proefschriftmaken.nl

Front & Back: made possible by DALL-E and WordCloud for Python

ISBN 978-94-6469-147-4

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

To Amalia.

CONTENTS

SUMMARY

Despite the attention that depression has received in the past two decades in the media, politics, and research, the percentage of depressed people who receive minimally adequate treatment remains low, ranging from 3% in low-income countries to 23% in high-income countries. The advancements in information and communication technology provide new opportunities to bridge the treatment gap. So called e-mental health for depression holds the promise of addressing all common barriers by providing low-cost, standardized, scalable, evidence-based, privacy-preserving, immediately available treatment to depressed people in the comfort of their own home any time of day. Moreover, many such developed systems have been found to be effective in randomized controlled trials. Effectiveness of treatment, however, depends to a large extent on adherence to treatment, and this appears to be the crux of e-mental health of depression. Particularly as systems move from controlled research contexts into the wild, dropout rates surge. Additionally, it has been observed that as human involvement in the system decreases from therapist support via administrative support to no support, attrition increases and the effectiveness of treatment decreases. Before concluding that a human presents a necessary constituent of effective e-mental health treatment, though, it may be worthwhile to take a closer look at the concrete technological implementation of existing systems.

To this end, we systematically studied the technological state of the art of existing software systems for the treatment and prevention of depression detailed and evaluated in the scientific literature. The 267 identified systems were characterized on 45 attributes concerning their technological features, their empirical evaluation, and the articles that describe them. This information was entered into a relational database of 14 tables. In order to quantify how technologically advanced a system is, we divided systems into their functional components or *functions* and rated each function with regard to its technological sophistication. Functions could be of five types: intervention, planning support, execution support, monitoring support, and social support. Rating the functions required us to develop a set of five scales that defined different degrees of technological sophistication, with each of the scales covering one type of function. The scales were found to have satisfactory reliability and concurrent validity when applied by independent raters to various samples of the functions. The technological sophistication of an entire system was then computed as the average rating of all functions within a system. With this, we could quantitatively describe the technological state of the art of the e-mental health for depression field in a literature review. This yielded the observation that the majority of systems were of a psychoeducational nature, i.e. they delivered information to the user but did not process information coming from the user. Additionally, the lack of human guidance was not compensated for with more functionality or with more technologically sophisticated functionality in autonomous systems. Lastly, no developments in the technological sophistication of systems could be observed over time for the examined time frame from the 2000 to 2017.

The literature review made it evident that most systems developed up until 2017 were lacking in their response capabilities especially concerning the textual input of users. Considering the advances that have been made in natural language processing with data-driven approaches and the advent of deep learning, an opportunity presented itself to address this shortcoming using these artificial intelligence techniques. Thought recording is a particularly suited exercise to this end for three reasons: first, it forms an integral part of the most used psychotherapeutic approaches for depression, Cognitive Therapy and Cognitive Behavioral Therapy; second, it is based on concise, structured descriptions in natural language; and, third, it is often assigned to patients as homework in face-to-face therapy, so that indirectly motivating patients to complete thought records by signaling understanding can benefit not only e-mental health but also face-to-face treatment. Regularly completing thought records allows patients to identify patterns in their thinking. According to cognitive theory, these patterns are caused by underlying core beliefs and are, in turn, causative of negative emotions, such as the low mood that characterizes depression. Learning to pay attention to the thoughts and recognizing patterns is the first step towards changing them. In the absence of human guidance, conversational agents are a promising technology for supporting users in a human-like manner. These software programs that engage in dialog with the users have frequently been found to be not only acceptable to users but also satisfactory and preferred over other screen-based methods or even humans.

In light of these considerations, we designed and studied a conversational agent that made use of data-driven natural language processing methods to support thought recording. This required developing a means of automatically determining the active core belief(s) of a person in a certain situation from the thoughts that person delineates in a thought record about said situation. For this purpose, we collected 1600 thought records in questionnaire format from 320 healthy participants in a crowd-sourcing setting. Each of the more than 5000 thoughts was manually labeled with respect to a set of nine possible underlying core beliefs. Using this labeled dataset, we trained a number of supervised classification and regression algorithms. All algorithms could label six of the nine core beliefs with correlations of at least $\rho = 0.35$ between the manually assigned labels and the algorithmically assigned labels; for the other three, there were not enough labeled training samples. Across core beliefs, the best performance was obtained with a set of nine recurrent neural networks with correlations on those six well-labelable core beliefs ranging between $\rho = 0.58$ and $\rho = 76$. With these networks, we went on to develop the feedback of a conversational agent to support thought recording to study the feasibility of recording thoughts in dialog with a conversational agent and to increase users' motivation. The agent engaged with 308 sub-clinically depressed participants recruited via crowd-sourcing in two thought recording sessions, an initial practice session and a second session in which participants were asked to complete as many personal thought records beyond the first one as they wanted. The conversational agent provided three different kinds of feedback on the thought record: (1) it acknowledged receiving the thought record, (2) it acknowledged receiving the thought record and gave feedback on the number of thoughts the participant wrote down (effort), and (3) it acknowledged receiving the thought record, gave feedback on the effort, and gave feedback on the content by interpreting the thoughts with respect to the underlying core beliefs. We found the conversational agent to be a feasible means of collecting thought record data in dialog because all participants were capable of completing thought records in this way. The distribution of automatically labeled core beliefs in this population of sub-clinically depressed participants mirrored that of the healthy population in our previous study and that of a clinical population in another study. However, we found no differences between the different feedback strategies of the agent concerning both the self-reported and the behavioral measure of motivation. These results open up a number of avenues for future research, such as studying whether the feedback is processed correctly, whether a population with depression that is motivated by a wish to get healthy might behave or experience the system differently from our sample that was recruited online and did not meet diagnostic criteria for depression, or whether extending the feedback with other conversational techniques from common coaching frameworks like Motivational Interviewing might result in a positive effect on motivation.

Samenvatting

Ondanks de aandacht die depressie de afgelopen twee decennia in de media, de politiek en de wetenschap heeft gekregen, blijft het percentage depressieve mensen dat een minimaal adequate behandeling krijgt laag, variërend van 3 % in lage-inkomenslanden tot 23 % in hoge-inkomenslanden. De vooruitgang in de informatie- en communicatietechnologie brengt nieuwe mogelijkheden om de behandelingskloof te overbruggen. De zogenaamde *elektronische geestelijke gezondheidzorg* (e-mental health voor depressie) *voor depressie* biedt de belofte om alle veelvoorkomende barrières aan te pakken door goedkope, gestandaardiseerde, schaalbare, evidence-based, privacy-beschermende, onmiddellijk beschikbare behandeling aan depressieve mensen te leveren in hun thuisomgeving op elk moment van de dag. Bovendien zijn veel van dergelijke ontwikkelde systemen effectief gebleken in gerandomiseerde gecontroleerde trials. De doeltreffendheid van de behandeling hangt echter in grote mate af van de therapietrouw, en dit lijkt de crux van e-mental health voor depressie te zijn. Met name bij de overgang van een gecontroleerde onderzoeksomgeving naar de praktijk, nemen de uitvalpercentages sterk toe. Bovendien is vastgesteld dat naarmate de menselijke betrokkenheid bij het systeem afneemt, van ondersteuning van de therapeut via administratieve ondersteuning tot geen ondersteuning, de uitval toeneemt en de doeltreffendheid van de behandeling afneemt. Voordat we concluderen dat de mens een noodzakelijk bestanddeel is van een effectieve e-mental health behandeling, kan het echter de moeite waard zijn om de concrete technologische implementatie van bestaande systemen nader te bekijken.

Daartoe hebben we systematisch de technologische stand van zaken bestudeerd van bestaande, in de wetenschappelijke literatuur beschreven en geëvalueerde softwaresystemen voor de behandeling en preventie van depressie. De 267 geïdentificeerde systemen werden gekarakteriseerd aan de hand van 45 attributen met betrekking tot hun technologische kenmerken, hun empirische evaluatie, en de artikelen waarin ze beschreven werden. Deze informatie werd ingevoerd in een relationele database met 14 tabellen. Om te kwantificeren hoe technologisch geavanceerd een systeem is, hebben we de systemen in functionele componenten of *functies* onderverdeeld en elke functie een cijfer gegeven voor de technologische geavanceerdheid ervan. Wij onderscheidden vijf typen functies: interventie, ondersteuning bij planning, ondersteuning bij de uitvoering van onderdelen van de interventie, ondersteuning bij zelf-monitoring en sociale ondersteuning. De beoordeling van de functies vereiste de ontwikkeling van een reeks van vijf schalen die verschillende graden van technologische verfijning definiëren, waarbij elk van de schalen één type functie beschrijft. De schalen bleken voldoende betrouwbaarheid en concurrent validiteit te hebben wanneer ze door onafhankelijke beoordelaars werden toegepast op verschillende steekproeven van de functies. De technologische geavanceerdheid van een heel systeem werd vervolgens berekend als de gemiddelde waardering van alle functies binnen een systeem. Hiermee konden we de technologische stand van zaken op het gebied van e-mental health voor depressie kwantitatief beschrijven in een literatuuroverzicht. Op basis hiervan stelden we vast dat de meerderheid van

de systemen van psycho-educatieve aard waren, dat wil zeggen dat ze informatie aan de gebruiker leverden maar geen informatie van de gebruiker verwerkten. Bovendien werd het gebrek aan menselijke begeleiding niet gecompenseerd door meer functionaliteit of door technologisch geavanceerdere functionaliteit in autonome systemen. Ten slotte hebben we geen ontwikkelingen in de technologische geavanceerdheid van systemen waargenomen in het onderzochte tijdvak van 2000 tot 2017.

Uit het literatuuronderzoek is gebleken dat de meeste systemen die tot 2017 zijn ontwikkeld, tekortschieten in hun reactievermogen, met name wat betreft de tekstuele input van gebruikers. Gezien de vooruitgang in de natuurlijke taalverwerking met datagestuurde benaderingen en de komst van deep learning, zagen we kansen om deze tekortkoming aan te pakken met behulp van kunstmatige-intelligentietechnieken. Gedachteregistratie is hiervoor om drie redenen een bijzonder geschikte oefening: ten eerste vormt het een integraal onderdeel van de meest gebruikte psychotherapeutische behandelingen voor depressie, cognitieve therapie en cognitieve gedragstherapie; ten tweede is het gebaseerd op beknopte, gestructureerde beschrijvingen in natuurlijke taal; en ten derde wordt het vaak aan patiënten meegegeven als huiswerk bij face-to-face therapie, zodat het indirect motiveren van patiënten om gedachteregistraties in te vullen door het signaleren van begrip niet alleen de e-mental health, maar ook de face-to-face behandeling ten goede kan komen. Het regelmatig invullen van gedachteregisters stelt patiënten in staat om patronen in hun denken te identificeren. Volgens de cognitieve theorie worden deze patronen veroorzaakt door onderliggende kernovertuigingen en zijn ze op hun beurt weer de oorzaak van negatieve emoties, zoals de neeslachtige stemming die kenmerkend is voor depressie. Het leren letten op gedachten en het herkennen van patronen is de eerste stap om ze te veranderen. Bij gebrek aan menselijke begeleiding zijn computergestuurde conversationele agenten een veelbelovende technologie om gebruikers op een mensachtige manier te ondersteunen. Deze softwareprogramma's, die een dialoog met de gebruikers aangaan, zijn vaak niet alleen aanvaardbaar bevonden door de gebruikers, maar ook bevredigend en verkozen boven andere schermgebaseerde methoden of zelfs mensen.

Gezien deze overwegingen hebben wij een conversationele agent ontworpen en bestudeerd die gebruik maakt van data-gedreven natuurlijke taalverwerkingsmethoden om gedachteregistratie te ondersteunen. Dit vereiste de ontwikkeling van een manier om automatisch de actieve kernovertuiging(en) van een persoon in een bepaalde situatie af te leiden uit de gedachten die die persoon omschrijft in een gedachtenregistratie over die situatie. Hiervoor verzamelden we 1600 gedachtenregistraties in vragenlijstformaat van 320 gezonde deelnemers gerekruteerd via een crowd-sourcing platform. Elk van de meer dan 5000 gedachten werd handmatig gelabeld met betrekking tot negen mogelijke onderliggende kernovertuigingen. Met behulp van deze gelabelde dataverzameling trainden we een aantal gesuperviseerde classificatie- en regressiealgoritmes. Alle algoritmes konden zes van de negen kernovertuigingen labelen met correlaties van minstens $\rho = 0.35$ tussen de handmatig toegekende labels en de algoritmisch toegekende labels; voor de andere drie waren er niet genoeg gelabelde trainingsvoorbeelden. Voor alle kernovertuigingen werd de beste prestatie verkregen met een verzameling van negen recurrente neurale netwerken met correlaties op die zes goed labelbare kernovertuigingen variërend tussen $\rho = 0.58$ en $\rho = 76$. Met deze netwerken ontwikkelden we vervolgens de terugkoppeling van een conversationele agent om gedachteregistratie te ondersteunen, om de haalbaarheid van gedachteregistratie in een dialoog met een agent te bestuderen en om de motivatie van gebruikers te verhogen. De agent ging in gesprek met 308 subklinisch depressieve deelnemers, gerekruteerd via crowd-sourcing, in twee gedachteregistratiesessies, een eerste oefensessie en een tweede sessie waarin de deelnemers gevraagd werd om zoveel persoonlijke gedachteregistraties te maken als ze wilden. De conversationele agent gaf drie verschillende soorten terugkoppeling op de gedachteregistratie: (1) hij bevestigde de ontvangst van de gedachteregistratie, (2) hij bevestigde de ontvangst van de gedachteregistratie en gaf terugkoppeling over het aantal gedachten dat de deelnemer opschreef (inspanning) en (3) hij bevestigde de ontvangst van de gedachteregistratie, gaf terugkoppeling over de inspanning en gaf terugkoppeling over de inhoud door de gedachten te interpreteren met betrekking tot de onderliggende kernovertuigingen. Wij vonden de conversationele agent een haalbare manier om gedachteregistratiegegevens in een dialoog te verzamelen, omdat alle deelnemers in staat waren op deze manier gedachteregisters in te vullen. De verdeling van automatisch gelabelde kernovertuigingen in deze populatie van subklinisch depressieve deelnemers weerspiegelde die van de gezonde populatie in onze vorige studie en die van een klinische populatie in een andere studie. We vonden echter geen verschillen tussen de verschillende terugkoppelingsstrategieën van de agent met betrekking tot zowel de zelf-gerapporteerde als de gedragsmatige maat van motivatie. Deze resultaten openen een aantal wegen voor toekomstig onderzoek, zoals het bestuderen of de terugkoppeling correct wordt verwerkt, of een populatie met depressie die gemotiveerd is door de wens om gezond te worden zich gedraagt of het systeem anders ervaart dan onze online gerekruteerde steekproef die niet voldeed aan diagnostische criteria voor depressie, of dat het uitbreiden van de terugkoppeling met andere gesprekstechnieken uit gangbare coachingskaders zoals Motivational Interviewing resulteert in een positief effect op de motivatie.

1

INTRODUCTION

1.1. MOTIVATION

A World Health Organization (WHO) report from 2011 predicted depression to be responsible for most of the disease burden worldwide by 2030 [1]. This prediction was made on the basis of considerations such as the high lifetime prevalence of depression, the fact that those affected are often entirely disabled by the illness for prolonged periods of time, and that one depressive episode increases the likelihood of further episodes. It did not factor in a pandemic in 2020 that would not only dramatically enlarge the groups that are affected or at risk [2–5] but would also create a sudden, pervasive need for the remote administration of treatment [6]. These factors have resulted in *e-mental health for depression*, the delivery of therapeutic content by means of technology, moving more and more into therapeutic practice [7].

E-mental health for depression can take many forms with varying degrees of human therapist involvement. Several of the interventions have been found to be as effective as face-to-face treatment [8–10], especially when they are also adhered to in terms of module completion [11, 12]. Nevertheless, Richards *et al.* [13] observed that as human guidance decreases from *therapist* to *administrative* to *no support*, effectiveness of treatment decreases and dropout rates increase. This indicates that there are human factors that serve a motivational function, which had not been effectively mimicked in unguided systems for depression treatment at the time of the study (2012) by Richards *et al.* [13].

Since the relationship between client and therapist has long been thought to be an important factor in therapeutic outcomes [14] and in adherence to treatment [15], a substantial body of research exists on the human factors that may play a role in fostering the alliance between therapist and patient. As Mohr *et al.* [16] point out, however, it cannot simply be assumed that these aspects readily translate to eHealth interventions. In their model of *supportive accountability*, they propose that it is *accountability* that leads to greater adherence to eHealth. Two important factors that are theorized to contribute to this accountability are *the perceived legitimacy of the therapist* and *the bond between therapist and client*, with both hinging on good communication. For therapists to be perceived as legitimate, they must carefully attend and respond to the information and experiences that the client shares (reciprocity) and simultaneously convey the feeling that they have the client's best interest at heart (benevolence). This is in line with the notion of *empathy* in the framework of motivational interviewing [17] and with the finding of Rutjes *et al.* [18] that much of the value of human coaches lies in their ability to closely listen to their clients and to tailor their advice and approach accordingly.

The model of supportive accountability posits social presence, that is, the presence of another human being, as a necessary requirement for accountability in treatment and thus also for increasing adherence to treatment. However, including a human for guidance in the e-mental health loop is not always feasible or even desirable. Particularly entirely autonomous systems hold great potential for the prevention and treatment of depression (a) when there is a paucity of mental health resources, such as an imbalanced patient to therapist ratio [6], (b) in the case of homework assignment that are not intended to be guided by therapists, or (c) when barriers like stigma prevent people from seeking out therapists in the first place. In such cases, artificial intelligence (AI) may provide a means to guide patients in another human's stead. In leaning on the definition of AI provided by Russell *et al.* in their seminal textbook on AI [19], we define AI as computer algorithms or systems that use their perceptions to arrive at human-programmed or machine-learnt rules for behavior that is intended to maximize the chances of attaining their goals. While AI emerged as a research discipline in the 1950s, the increase in processing power of computers, and the frontiers that have been pushed back as a consequence, has resulted in a veritable boom in the past decade. With the myriad of advances made in this research discipline, an opportunity presented itself to investigate the following main research question:

> How can artificial intelligence guide people in e-mental health interventions for depression?

This thesis sets out to answer it.

1.2. RESEARCH QUESTIONS AND HYPOTHESES

Computerized programs for depression have often been found to be poorly adhered to [8–10]. However, before reaching the conclusion that high dropout may be a natural characteristic of autonomous eHealth, several alternative explanations must be considered [20]. One of these is that systems may not be sufficiently engaging. Adherence and attrition have been linked to engagement or lack thereof [21–24], with aspects of the users, their context, and the system all theorized to play a role in engagement. In one of the first systematic reviews of system features of e-mental health for depression, Zhao *et al.* [25] concluded that systems were lacking in interactivity. Thus, if most systems that have been developed to prevent and treat depression to date are low-tech implementations with a unidirectional flow of information from system to user (essentially computerized psychoeducation), should we expect them to be engaging or, in any case, more engaging than self-help books? This invited a closer look at the concrete implementation of system functionality and its technological sophistication, resulting in the first research sub-question:

> RQ1: How can we assess the technological sophistication of an e-mental health systems for the prevention and treatment of major depressive disorder?

As a result of the increasing prevalence of depression and the increasing ubiquity of computers and smartphones, numerous systems for depression prevention and treatment have been developed in the past two decades. Yet, as can be seen from the very short system descriptions in many of the articles, the technology is often regarded as a vessel for the therapeutic content rather than as an active ingredient of the intervention. Kelders *et al.* [26], however, found that 55 % of the variance in adherence to eHealth could be explained by the intervention format and technological features that were implemented in systems. Additionally, several authors of literature reviews have called for a systematic analysis of the components of systems to better understand the system landscape and to ultimately determine the ones that are contributing to clinical outcomes [9, 26, 27]. A first step in this direction was undertaken by Wildeboer *et al.* [28] for e-mental health interventions in a meta-analysis linking persuasive system features with effectiveness outcomes. However, the authors used a set of systems identified by Kelders *et al.* [26] four years earlier and did not update this set with more recent works. Furthermore, they only included systems that were evaluated in trials to assess their effectiveness in reducing symptoms, leading to the exclusion of less rigorously evaluated but possibly more advanced technology. Other reviews have focused on one specific technological feature and its effect on clinical outcomes (e.g. communication modality [29]) or have looked at the absence or presence of certain technological features in systems (e.g. [25]). Yet, before it becomes possible to study the link between the technological realization of systems and clinical outcomes, it is first necessary to determine how systems could be disassembled into components and how these could then be described technologically. These considerations led to the second research sub-question:

> RQ2: What is the technological state of the art of software systems for the prevention and treatment of depression?

Most e-mental health for depression adopts Cognitive Behavior Therapy (CBT) as the underlying therapeutic framework [9, 25]. This corresponds with CBT having crystallized into the gold standard of psychotherapy in the second half of the 20th century [30] for a number of reasons. For one, the models underlying the hypothesized mechanisms of change in CBT correspond to those of the predominant psychological research paradigms *behavioral* and *cognitive psychology*, which have accrued a large body of research in their support. The same can be said for CBT itself, which has been found to be effective for a number of conditions in a very limited time frame of less than ten sessions in various randomized controlled trials [31, 32]. Additionally, the approach is often favored over other approaches by therapists and clients alike due to its transparent work mechanisms and its focus on concrete techniques that can be readily applied outside of the sessions [33]. The active role that the client takes in acquiring these skills combined with the large amount of learning and practice material available for CBT also make it particularly suited for self-administration, be it via self-help books or via a computer or mobile app [33]. As a consequence of the clear rationale, highly structured format, prevalence in e-mental health for depression, and large evidence-base of CBT, the focus of the main research question was narrowed down to supporting CBT with artificial intelligence.

Cognitive Behavior Therapy is the combination of Cognitive Therapy with Behavioral Activation. Historically, CBT evolved from Behavioral Activation (BA) therapy and was extended with the ideas from Cognitive Therapy (CT) in a second wave. Assuming the models for learning and behavior from behaviorism, BA for depression has the aim of getting patients to actively lead their lives as members of a larger community to create opportunities for positive reinforcement by preventing them from isolating themselves. Thanks to consumer *behavior* driving the market economy, there are now technological solutions for the tracking, analysis, and visualization of physical and online activity, for controlled social or virtual rewards, for highly personalized product recommendations, and for timely reminders. Many of these computerized *behavior change* mechanisms can be readily applied to the automation and support of the techniques of BA, such as activity planning and monitoring, rewarding, reminding, or encouraging. As a result, there is already an abundance of ways in which AI can be used to support BA for depression and we thus turned our attention to CT.

According to the father of Cognitive Therapy, Aaron Beck, the therapy is particularly well suited for the treatment of depression. Depression is a mental illness commonly classified as a *mood disorder*. This is reflected in the necessary symptoms that inform its diagnosis: over the course of at least two weeks, a low mood nearly all the time or the loss of interest or pleasure in most daily activities. Beck [34], however, argues for the cognitive primacy in depression rather than the emotive one, instead calling it a *thought disorder*. He developed *Cognitive Theory* to explain how mental illness is caused and maintained by maladaptive core beliefs or *schemas*, beliefs that we hold as truths about ourselves, the world, or the future and that are central to our self-concept. An example for such a maladaptive schema may be: “I am unlovable." Schemas often form during childhood either through our own experiences and observations or from what we are told. When held, they inform our immediate appraisals of situations, which are often automatic and can go entirely unnoticed. To make this concrete, if one holds the schema that one is not lovable, even caring gestures will be played down, viewed with suspicion, not perceived at all, or viewed as criticism, such as when one is paid a compliment and one's first thought is that the other person is just being polite. At the heart of cognitive theory is the premise that it is this first (automatic) thought that makes us feel a certain way in a certain situation rather than the situation causing the negative emotional response directly. The thought that the compliment could not possibly be sincere will result in sadness, while another, positive appraisal would result in happiness and a feeling of connection. From this theory, Beck went on to derive *Cognitive Therapy* [35], a treatment method that strives to unveil, challenge, and ultimately change maladaptive thinking.

An important technique of Cognitive Therapy is thought recording [35]. This is commonly done with the help of pen-and-paper forms. The most rudimentary thought record form is intended as a thought monitoring support tool and consists of a table with columns labeled *situation*, *emotion and intensity*, *automatic thought*, and *behavior*. Patients are asked to record situations that resulted in a maladaptive response (in a patient with depression, this will commonly be situations in which the patient felt overwhelmed by sadness, in a patient with anger management difficulties, this will commonly be situations in which the patient became disproportionately angry, etc.). The aim of the exercise is to teach patients to attend to their thoughts, learn to express them, and ultimately bring to light patterns in thinking that sustain their illness. Once these patterns have been detected, the thought record can be extended with the downward arrow technique [36], a technique intended to uncover maladaptive schemas. It achieves this by repeatedly asking the questions “Why would the previously stated thought be so upsetting to me? What does that say about me?" When the consecutive answers to this question begin to repeat themselves or begin to appear ridiculous, one has probably arrived at a schema. A suite of additional techniques can be included in the thought record form, e.g. formulating alternative healthy automatic thoughts or challenging the maladaptive thought by putting it on the dock of an imaginary court and taking on the roles of defense attorney, prosecution, jury, and judge in an imaginary trial. Since it is assumed that thought records can be most accurately completed close in time to when the upsetting situation occurred, patients are typically asked to record their thoughts as homework assignments between face-to-face sessions and to bring the forms to the session to discuss with the therapist [37]. Unfortunately, patients often struggle to adhere to their homework assignments in cognitive therapy [38–42].

Since much of our conscious thinking occurs in natural language, the techniques of CT critically hinge on language understanding. They demand of patient and therapist an in-depth semantic processing of unconstrained natural language to recognize thought patterns, follow the downward arrow, formulate healthier alternative thoughts, and challenge thoughts via Socratic questioning or by putting them on “trial." Demanding the same kind of processing of a computer program still seems a long way off, despite the advances made in natural language understanding in the past decade [43]. Nonetheless, several recent developments in natural language processing, in part due to deep learning [44], afford an opportunity to explore an automated, in-depth analysis of cognitions formulated by patients.

These considerations inspired a vision for answering our main research question of how artificial intelligence could support depression treatment: a software program engages in dialog with the users to help them complete thought records on a regular basis by giving an interpretation of what they write down in their thought records. This could be achieved by analyzing the thoughts and providing feedback on the content, namely the possible underlying schemas, thus signaling understanding. In addition to giving the user feedback on a content level, an increasingly complex model of the user's schema structure could be constructed as the user completes more thought records over time. This model could be related back to the user as well. This thesis puts forward two hypotheses as important steps towards realizing this vision.

The first hypothesis concerns itself with the analysis of thought records to generate an interpretation on the content level. It was formulated as

> H1: Schemas can be automatically extracted from the thoughts people delineate in thought records including the downward arrow technique.

Support for this hypothesis can be seen in the promising results that have been obtained when applying state-of-the-art natural language processing (NLP) to mental health texts. To date, most of this research has focused on psychological assessment. To this end, social media platforms and forums are crawled to obtain the natural language data, which is then searched for linguistic markers of depression, crisis, or suicidal risk (e.g. [45–51]). The mental health app Koko, for example, was used to collect a dataset of short “posts" of a few sentences. With a subset of these, deep learning models were trained to detect crisis. The best model detected crisis with an F1-score accuracy of 0.80 [49]. In another study [50], the NLP data was taken from 16 topic-based mental health forums on the social media website Reddit. The authors trained a deep learning model that could identify the mental-health related themes of posts (e.g. depression, suicide, anxiety, bipolar disorder) achieving an F1-score accuracy of 0.71 in this multi-class classification task. In a similar task, Benton *et al.* [51] collected tweets from the social media platform Twitter to identify several, not mutually exclusive classes: suicidal risk, atypical mental health, and seven mental health conditions. They could show that multi-task models, models that had access to all class labels for the tweets of a user, were able to leverage correlations between the classes to produce better predictions. Although previous research has not yet directly attempted to identify schemas from text using NLP, these examples suggest that state-of-the-art NLP methods are a viable means to label natural language input with regard to mental health concepts. However, to meaningfully identify schemas using NLP, a taxonomy of schemas is necessary. Such a taxonomy is provided by Millings *et al.* [52], who obtained a classification rubric for schemas from thought records by conducting a content analysis of thought records. The identified schemas thus appear to differ in their language and word usage to the extent that human analysts were able to place them in different categories. If this is the case, a good model trained on sufficient data should be able to pick up on these differences as well. Finally, as in the dataset of Benton *et al.* [51], the schemas identified by Millings *et al.* [52] are not mutually exclusive and might therefore inform each other, possibly further improving prediction accuracy. Taken together, these findings indicate that the automatic labeling of schemas is within reach.

As another crucial step towards our vision, it must be possible for users to complete thought records in dialog with a conversational agent. Conversational agents (CAs) are software programs that engage in dialog with their users. They thus present a promising interface to support patients in dealing with their maladaptive cognitions using natural language. As our second hypothesis, we therefore posit that

> H2: An intelligent conversational agent is a feasible tool for completing thought records in dialog format.

Conversational agents have a comparatively long history in mental health, with the first, Weizenbaum's ELIZA, already having been developed in the 1960's. However, it is important to note that Weizenbaum never intended to create a therapeutic program [53], but rather chose this domain to effectively disguise the program's complete lack of understanding. He argued that this context is the only one in which a user would feel *heard* despite the agent clearly demonstrating its deficits in comprehension, because these deficits would be attributed to the therapeutic technique [53]. The probably first *serious* attempt at creating a CA with semantic understanding of the user for the treatment of depression was the Overcoming Depression program of Colby [54]. The dialog part of the program consisted of a large semantic database and a set of production rules that would produce the right-hand side of the rule after mapping the natural language input from the user to a left-hand side semantic pattern. Such rule-based systems require many hours of engineering labor but are nonetheless likely to misunderstand the user [55, Ch. 8]. It is perhaps not surprising then that despite the number of computerized treatment programs growing steadily in the past decades, the authors of two recent literature surveys called the field of conversational agents for mental health *nascent* [56, 57].

Despite few agents having been developed and the evidence concerning their clinical effectiveness being sparse [58–60], conversational agents supporting or delivering mental health interventions perform favorably in terms of feasibility, acceptability, and user satisfaction [61]. In a scoping review on patient opinions, these were generally satisfied with the CAs they had interacted with and saw them as useful. Furthermore, Lucas *et al.* [62] found that some people prefer to disclose to an agent rather than to a human because they are less afraid of being negatively evaluated and less worried about the impression they were making.

We also expect to find evidence in support of H2 because users could complete thought records when not guided by a conversational agent in other e-mental health for depression interventions. For example, Millings *et al.* [52] constructed their list of schemas from the core beliefs of thought records completed by users of the online CBT program Beating the Blues. Of the entire sample of thought records with downward arrow technique, 87 % had a core belief with at least one identifiable schema. While this may lead one to wonder why agents are needed at all then, it is important to note that this thesis does not take the stance that agents are necessary for thought recording. Rather, we aim to mimic the dialog format of face-to-face therapy with a conversational agent, because we expect the guidance of an agent to have benefits for the user, such as favorable effects on user experience and adherence. Indications in support of this can be found outside of the mental health context. For example, in a study comparing a conversational agent based system for remote patient monitoring of chronic obstructive pulmonary disease, the patients preferred the conversational agent over the normal screen-based method [63]. Similarly, Bickmore *et al.* [64] found, when comparing a relational agent and a non-relational agent to a form-based control condition, that participants in both agent conditions voluntarily looked at more educational pages and engaged in more physical activity than in the control condition. Within the mental health context, Maharjan *et al.* [65] found that people with affective disorders who evaluated a speech-enabled conversational agent in the wild appreciated that it gave them the feeling that someone was listening to them without judging them or looking for solutions.

In agents that engage in dialog, two types of dialog are distinguished: task-oriented and open-domain dialog [66]. The former aims to support users in achieving a specific task in dialog with a conversational agent, which typically takes the initiative (e.g., ordering pizza), while the latter strives to let the agent provide consistent responses to topics and ideas brought up by the user (e.g., small-talk). Although thought records can be written about any kind of life-event and might therefore be considered open-domain, their highly structured format allows collecting thought records in a task-oriented dialog as long as the agent does not engage in a discussion on the content of situation description or the thoughts. The task for the agent is then to obtain responses to all questions of the thought record. Since agents developed in the health and mental health domain in the past have successfully engaged with users in task-oriented dialog (e.g., [67–69]), we expect that a task-oriented dialog for thought recording can be conducted successfully by an agent as well.

1.3. RESEARCH APPROACH AND THESIS STRUCTURE

We approached the first two research questions jointly: the first research question was addressed in the process of conducting a systematic literature review to answer the second research question. For the review, all publications describing a scientifically evaluated e-mental health solution for the prevention or treatment of depression that had been published since 2000 were collected. Each description was read to identify the functional components (functions) of the system. Aside from their functions, systems were characterized with another 44 variables, concerning also their evaluation and more general descriptive information such as their *build year* or *name*. This resulted in a database containing these 45 variables in 14 tables. In the process of filtering the literature and characterizing a first set of systems, it became increasingly apparent that many systems were lacking in interactivity and responsiveness, appearing largely as self-help books put on a website. This led to the first formulation of a scale to assess technological sophistication of a system, with the main focus lying on the interactivity and the system capabilities in terms of responding to user input. Lastly, we determined the concurrent validity of the scales. More details concerning the database creation procedure and the e-mental Health Degree of Technological Sophistication (eHDTS) scales can be found in Chapter 2.

The database, in which each function of a system was assigned a score on the eHDTS, formed the basis for an extensive systematic literature review. We divided this into three parts: a descriptive part concerning the systems, their functions, and their evaluations, an analytical part concerning the technological sophistication of the systems, and a developmental part concerning the evolution of the system landscape across nearly two decades of active research. By covering all three parts, the literature review presented in Chapter 3, strives to give a comprehensive overview of the state of the art of systems for the treatment and prevention of depression by taking a technological perspective. It thus aims to answer the second research question.

Testing the hypothesis that it is possible to identify the underlying schemas from the thoughts noted down in a thought record required a dataset of thought records. We collected such a dataset using the crowd-sourcing platform Amazon Mechanical Turk. To this end, healthy participants were asked to each complete five thought records including the downward arrow technique. We then created a list of all thoughts uttered in these thought records, randomized the order such that thoughts of the same thought record and same participant were no longer in sequence, and manually labeled the thoughts using the schema rubric of Millings *et al.* [52]. However, we applied it separately to individual thoughts along the downward arrow as well as to the final schema rather than only to the schema. We then used two subsets of the data to train and validate three types of machine learning algorithms to also assign a degree for each schema to a given thought. The machine learning algorithms k-nearest neighbors, support vector machines, and recurrent neural networks were chosen for their suitability for the multi-labeling task and their ability to produce meaningful predictions for small datasets. Chapter 4 presents the details concerning the dataset, the coding, the algorithms, and the performance.

To determine whether completing thought records with a conversational agent is feasible, we designed such an agent and asked participants to interact with it. The conversational agent asked participants the questions that constitute the thought record and downward arrow technique and was capable of giving reflective feedback. We used the best performing algorithm that we obtained from testing the first hypothesis to design the feedback of the agent. The agent was developed using the open-source text-based conversational agent development platform Rasa. We recruited sub-clinically depressed participants with the help of Prolific, a crowd-sourcing platform for research. Participants were asked to first practice thought recording with the agent in a first interaction session consisting of two thought records. They were subsequently invited to interact with the agent in a second session. In this session, the agent asked them to complete at least one thought record but as many additional thought records as they wanted. We assessed whether participants were able to complete the thought records in this setting and studied the voluntary comments made by the participants concerning their experience with the conversational agent to come to a conclusion concerning the feasibility of supporting thought recording in this manner. Chapter 5 can be consulted for the details of the study and its results.

The final chapter of this thesis, Chapter 6, summarizes the main findings with regard to the research questions and hypotheses presented above. We further discuss the limitations that qualify the conclusions and point out possible interesting directions for further research. We conclude with the contributions that the thesis makes to the fields of artificial intelligence and clinical psychology as well as to society at large.

Bibliography

1. World Health Organization. Global burden of mental disorders and the need for a comprehensive, coordinated response from health and social sectors at the country level (2011).
2. Office for National Statistics. Coronavirus and depression in adults, Great Britain: June 2020 (2020).
3. Butterworth, P., Rodgers, B. & Windsor, T. D. Financial hardship, socio-economic position and depression: results from the PATH Through Life Survey. *Social science & medicine* **69,** 229–237 (2009).
4. Choi, K. R., Heilemann, M. V., Fauer, A. & Mead, M. A second pandemic: mental health spillover from the novel coronavirus (COVID-19). *Journal of the American Psychiatric Nurses Association* **26,** 340–343 (2020).
5. Ammar, A., Trabelsi, K., Brach, M., Chtourou, H., Boukhris, O., Masmoudi, L., Bouaziz, B., Bentlage, E., How, D., Ahmed, M., *et al.* Effects of home confinement on mental health and lifestyle behaviours during the COVID-19 outbreak: insights from the ECLB-COVID19 multicentre study. *Biology of Sport* **38,** 9 (2021).
6. Wind, T. R., Rijkeboer, M., Andersson, G. & Riper, H. The COVID-19 pandemic: The 'black swan' for mental health care and a turning point for e-health. *Internet interventions* **20** (2020).
7. Ellis, L. A., Meulenbroeks, I., Churruca, K., Pomare, C., Hatem, S., Harrison, R., Zurynski, Y. & Braithwaite, J. The Application of e-Mental Health in Response to COVID-19: Scoping Review and Bibliometric Analysis. *JMIR mental health* **8,** e32948 (2021).
8. Andrews, G., Cuijpers, P., Craske, M. G., McEvoy, P. & Titov, N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. *PloS one* **5,** e13196 (2010).
9. Schröder, J., Berger, T., Westermann, S., Klein, J. P. & Moritz, S. Internet interventions for depression: New developments. *Dialogues in Clinical Neuroscience* **18,** 203–212 (2016).
10. Karyotaki, E., Riper, H., Twisk, J., Hoogendoorn, A., Kleiboer, A., Mira, A., MacKinnon, A., Meyer, B., Botella, C., Littlewood, E., Andersson, G., Christensen, H., Klein, J. P., Schröder, J., Bretón-ópez, J., Scheider, J., Griffiths, K., Farrer, L., Huibers, M. J., Phillips, R., Gilbody, S., Moritz, S., Berger, T., Pop, V., Spek, V. & Cuijpers, P. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms a meta-analysis of individual participant data. *JAMA Psychiatry* **74,** 351–359 (2017).

11. Hilvert-Bruce, Z., Rossouw, P. J., Wong, N., Sunderland, M. & Andrews, G. Adherence as a determinant of effectiveness of internet cognitive behavioural therapy for anxiety and depressive disorders. *Behaviour research and therapy* **50,** 463–468 (2012).

12. Donkin, L., Christensen, H., Naismith, S. L., Neal, B., Hickie, I. B. & Glozier, N. A systematic review of the impact of adherence on the effectiveness of e-therapies. *Journal of medical Internet research* **13,** e52 (2011).

13. Richards, D. & Richardson, T. Computer-based psychological treatments for depression: A systematic review and meta-analysis. *Clinical Psychology Review* **32,** 329–342 (2012).

14. Martin, D. J., Garske, J. P. & Davis, M. K. Relation of the therapeutic alliance with outcome and other variables: a meta-analytic review. *Journal of consulting and clinical psychology* **68,** 438 (2000).

15. Mohl, P. C., Martinez, D., Ticknor, C., Huang, M. & Cordell, L. Early dropouts from psychotherapy. *Journal of Nervous and Mental Disease* (1991).

16. Mohr, D., Cuijpers, P. & Lehman, K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. *Journal of medical Internet research* **13,** e30 (2011).

17. Miller, W. R. & Rollnick, S. *Motivational Interviewing: Preparing People to Change Addictive Behaviour.* (Guilford Press, 1991).

18. Rutjes, H., Willemsen, M. C., Kersten-van Dijk, E. T., de Ruyter, B. & IJsselsteijn, W. A. Better together: opportunities for technology in health coaching from the coach's perspective. in *Proceedings of the 2017 Conference on e-Coaching for Health and Wellbeing.* ([Sl: sn], 2017), 1–14.

19. Russell, S. & Norvig, P. Artificial Intelligence: A Modern Approach (2002).

20. Eysenbach, G. The law of attrition. *Journal of medical Internet research* **7,** e402 (2005).

21. Short, C., Rebar, A., Plotnikoff, R. & Vandelanotte, C. Designing engaging online behaviour change interventions: a proposed model of user engagement. *The European Health Psychologist* **17,** 32–38 (2015).

22. Short, C. E., DeSmet, A., Woods, C., Williams, S. L., Maher, C., Middelweerd, A., Müller, A. M., Wark, P. A., Vandelanotte, C., Poppe, L., *et al.* Measuring engagement in eHealth and mHealth behavior change interventions: viewpoint of methodologies. *Journal of medical Internet research* **20,** e9397 (2018).

23. Ritterband, L. M., Thorndike, F. P., Cox, D. J., Kovatchev, B. P. & Gonder-Frederick, L. A. A behavior change model for internet interventions. *Annals of Behavioral Medicine* **38,** 18–27 (2009).

24. Danaher, B. G., Boles, S. M., Akers, L., Gordon, J. S. & Severson, H. H. Defining participant exposure measures in Web-based health behavior change programs. *Journal of Medical Internet Research* **8,** e542 (2006).

25. Zhao, D., Lustria, M. L. A. & Hendrickse, J. Systematic review of the information and communication technology features of web- and mobile-based psychoeducational interventions for depression. *Patient Education and Counseling* **100,** 1049–1072 (2017).

26. Kelders, S. M., Kok, R. N., Ossebaard, H. C. & Gemert-Pijnen, J. E. V. Persuasive system design does matter: A systematic review of adherence to web-based interventions. *Journal of Medical Internet Research* **14** (2012).

27. van Ballegooijen, W., Cuijpers, P., Van Straten, A., Karyotaki, E., Andersson, G., Smit, J. H. & Riper, H. Adherence to Internet-based and face-to-face cognitive behavioural therapy for depression: a meta-analysis. *PloS one* **9,** e100674 (2014).

28. Wildeboer, G., Kelders, S. M. & van Gemert-Pijnen, J. E. The relationship between persuasive technology principles, adherence and effect of web-Based interventions for mental health: A meta-analysis. *International Journal of Medical Informatics* **96,** 71–85 (2016).

29. Barak, A., Hen, L., Boniel-Nissim, M. & Shapira, N. A comprehensive review and a meta-analysis of the effectiveness of internet-based psychotherapeutic interventions. *Journal of Technology in Human Services* **26,** 109–160 (2008).

30. David, D., Cristea, I. & Hofmann, S. G. Why cognitive behavioral therapy is the current gold standard of psychotherapy. *Frontiers in psychiatry* **9,** 4 (2018).

31. Hazlett-Stevens, H. & Craske, M. G. Brief cognitive-behavioral therapy: Definition and scientific foundations. in *Handbook of Brief Cognitive Behaviour Therapy* (eds Bond, F. W. & Dryden, W.) (John Wiley & Sons, Ltd., 2002).

32. Cape, J., Whittington, C., Buszewicz, M., Wallace, P. & Underwood, L. Brief psychological therapies for anxiety and depression in primary care: meta-analysis and meta-regression. *BMC medicine* **8,** 1–13 (2010).

33. Gaudiano, B. A. Cognitive-behavioural therapies: achievements and challenges. *Evidence-based mental health* **11,** 5–7 (2008).

34. Beck, A. T. Thinking and depression: I. Idiosyncratic content and cognitive distortions. *Archives of general psychiatry* **9,** 324–333 (1963).

35. Beck, A. T. *Depression. Clinical, Experimental and Theoretical Aspects.* (Hoeber, 1967).

36. Burns, D. D. *The Feeling Good Handbook* (Plume, 1999).

37. Kazantzis, N., Deane, F. P., Ronan, K. R. & L'Abate, L. *Using homework assignments in cognitive behavior therapy* (Routledge, 2005).

38. Russo, T. Cognitive counseling for health care compliance. *Journal of rational emotive therapy* **5,** 125–134 (1987).

39. Detweiler, J. B. & Whisman, M. A. The role of homework assignments in cognitive therapy for depression: Potential methods for enhancing adherence. *Clinical Psychology: Science and Practice* **6,** 267–282 (1999).

40. Helbig, S. & Fehm, L. Problems with homework in CBT: Rare exception or rather frequent? *Behavioural and cognitive psychotherapy* **32,** 291 (2004).

41. Kazantzis, N. & Shinkfield, G. Conceptualizing patient barriers to nonadherence with homework assignments. *Cognitive and Behavioral Practice* **14,** 317–324 (2007).

42. Kazantzis, N., Brownfield, N. R., Mosely, L., Usatoff, A. S. & Flighty, A. J. Homework in cognitive behavioral therapy: A systematic review of adherence assessment in anxiety and depression (2011–2016). *Psychiatric Clinics* **40,** 625–639 (2017).

43. Miner, A. S., Milstein, A. & Hancock, J. T. Talking to machines about personal mental health problems. *Jama* **318,** 1217–1218 (2017).

44. Torfi, A., Shirvani, R. A., Keneshloo, Y., Tavaf, N. & Fox, E. A. Natural language processing advancements by deep learning: A survey. *arXiv preprint arXiv:2003.01200* (2020).

45. Guntuku, S. C., Yaden, D. B., Kern, M. L., Ungar, L. H. & Eichstaedt, J. C. Detecting depression and mental illness on social media: an integrative review. *Current Opinion in Behavioral Sciences* **18,** 43–49 (2017).

46. Al-Mosaiwi, M. & Johnstone, T. In an absolute state: Elevated use of absolutist words is a marker specific to anxiety, depression, and suicidal ideation. *Clinical Psychological Science* **6,** 529–542 (2018).

47. Holtzman, N. S. *et al.* A meta-analysis of correlations between depression and first person singular pronoun use. *Journal of Research in Personality* **68,** 63–68 (2017).

48. Newell, E. E., McCoy, S. K., Newman, M. L., Wellman, J. D. & Gardner, S. K. You sound so down: Capturing depressed affect through depressed language. *Journal of Language and Social Psychology* **37,** 451–474 (2018).

49. Kshirsagar, R., Morris, R. & Bowman, S. Detecting and Explaining Crisis. in *Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology—From Linguistic Signal to Clinical Reality* (Association for Computational Linguistics (ACL), 2017), 66–73.

50. Gkotsis, G., Oellrich, A., Velupillai, S., Liakata, M., Hubbard, T. J., Dobson, R. J. & Dutta, R. Characterisation of mental health conditions in social media using Informed Deep Learning. *Scientific reports* **7,** 1–11 (2017).

51. Benton, A., Mitchell, M. & Hovy, D. Multi-task learning for mental health using social media text. *arXiv preprint arXiv:1712.03538* (2017).

52. Millings, A. & Carnelley, K. B. Core belief content examined in a large sample of patients using online cognitive behaviour therapy. *Journal of Affective Disorders* **186,** 275–283 (2015).

53. Weizenbaum, J. ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM* **9,** 36–45 (1966).

54. Colby, K. M. Human-computer conversation in a cognitive therapy program. in *Machine conversations* (Springer, 1999), 9–19.

55. Jurafsky, D. & Martin, J. *Speech and Language Processing (3rd ed. draft)* `https://web.stanford.edu/~jurafsky/slp3/`. Accessed in 2019.

56. Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S. & Torous, J. B. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. *The Canadian Journal of Psychiatry* **64,** 456–464 (2019).

57. Abd-Alrazaq, A. A., Alajlani, M., Alalwan, A. A., Bewick, B. M., Gardner, P. & Househ, M. An overview of the features of chatbots in mental health: A scoping review. *International Journal of Medical Informatics* **132,** 103978 (2019).

58. Provoost, S., Lau, H. M., Ruwaard, J., Riper, H., *et al.* Embodied conversational agents in clinical psychology: a scoping review. *Journal of medical Internet research* **19,** e6553 (2017).

59. Gaffney, H., Mansell, W. & Tai, S. Conversational agents in the treatment of mental health problems: mixed-method systematic review. *JMIR mental health* **6,** e14166 (2019).

60. Car, L. T., Dhinagaran, D. A., Kyaw, B. M., Kowatsch, T., Joty, S., Theng, Y.-L., Atun, R., *et al.* Conversational agents in health care: scoping review and conceptual analysis. *Journal of medical Internet research* **22,** e17158 (2020).

61. Bendig, E., Erb, B., Schulze-Thuesing, L. & Baumeister, H. The next generation: chatbots in clinical psychology and psychotherapy to foster mental health–a scoping review. *Verhaltenstherapie,* 1–13 (2019).

62. Lucas, G. M., Gratch, J., King, A. & Morency, L.-P. It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior* **37,** 94–100 (2014).

63. Cleres, D., Rassouli, F., Brutsche, M., Kowatsch, T. & Barata, F. Lena: a Voice-Based Conversational Agent for Remote Patient Monitoring in Chronic Obstructive Pulmonary Disease in *Proceedings of the 26th ACM Conference on Intelligent User Interfaces (ACM IUI 2021)* (Association for Computing Machinery, 2021).

64. Bickmore, T., Gruber, A. & Picard, R. Establishing the computer–patient working alliance in automated health behavior change interventions. *Patient education and counseling* **59,** 21–30 (2005).

65. Maharjan, R., Doherty, K., Rohani, D. A., Bækgaard, P. & Bardram, J. E. Experiences of a Speech-Enabled Conversational Agent for the Self-Report of Wellbeing Among People Living with Affective Disorders: An In-The-Wild Study. *ACM Transactions on Interactive Intelligent Systems* **12,** 1–29 (2022).

66. Shevat, A. *Designing bots: Creating conversational experiences* (O'Reilly Media, 2017).

67. DeVault, D., Artstein, R., Benn, G., Dey, T., Fast, E., Gainer, A., Georgila, K., Gratch, J., Hartholt, A., Lhommet, M., *et al.* SimSensei Kiosk: A virtual human interviewer for healthcare decision support. in *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems* (International Foundation for Autonomous Agents and Multiagent Systems, 2014), 1061–1068.

68. Tielman, M. L., Neerincx, M. A., Bidarra, R., Kybartas, B. & Brinkman, W.-P. A therapy system for post-traumatic stress disorder using a virtual agent and virtual storytelling to reconstruct traumatic memories. *Journal of medical systems* **41,** 1–10 (2017).

69. Kocielnik, R., Xiao, L., Avrahami, D. & Hsieh, G. Reflection companion: a conversational system for engaging users in reflection on physical activity. in *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* **2** (Association for Computing Machinery, 2018), 1–26.

2

EHEALTH4MDD: A DATABASE OF E-HEALTH SYSTEMS FOR THE PREVENTION AND TREATMENT OF DEPRESSIVE DISORDERS

This chapter is based on Burger, F., Neerincx, M. A. & Brinkman, W. P. Ehealth4MDD: A database of e-health systems for the prevention and treatment of depressive disorders. *Annual Review of CyberTherapy and Telemedicine* **16,** 18–24 (2018)

ABSTRACT

To date, meta-analyses of e-mental health systems for major depressive disorder (MDD) have largely overlooked the technological side of interventions. This warranted the creation of an open access database, EHealth4MDD, for the systematic study of the technological implementation in relation to intervention content, study design, and study outcomes. E-health systems were identified by conducting an exhaustive search on PubMed, Scopus, and Web of Science in 2017. The 5379 retrieved records yielded 267 systems. One coder extracted information from the records on 45 variables, organized into 14 tables in EHealth4MDD. A sample of each high-inference variable was double coded by a second coder to assess reliability. Percent agreement was satisfactory given that coders received no training and the number of possible categories was large. Furthermore, scales were developed to rate the degree of technological sophistication of system functions for each of five function types. Four of these scales demonstrated concurrent validity, as evidenced by the substantial to strong correlations observed when comparing the scales with the results of an unlabeled ordering task. For researchers in both computer science and clinical psychology, the database presents a useful tool to systematically study e-mental health interventions for depression.

2.1. INTRODUCTION

Although depression can be treated effectively, more than half of the approximately 300 million people worldwide suffering from the illness are receiving inadequate or no treatment [2]. E-mental health presents a promising direction in overcoming many of the barriers to and shortcomings of face-to-face treatment [3]. To date, both systematic reviews and meta-analyses in the field have largely focused on delivery aspects (e.g., guidance as a factor of influence [4]) and therapy aspects (e.g., limiting the scope to a specific therapeutic approach [5]) in relation to outcomes while neglecting influences of technology. To the best of our knowledge, only Zhao *et al.* [6] systematically reviewed the presence of certain technological features (communication tools, instructional ICT features, and self-monitoring tools) in psychoeducational e-mental health systems for depression, reaching the conclusion that most interventions lack in technological sophistication.

As has been pointed out in recent surveys, the field would benefit from a clearer picture of the features of e-mental health systems for depression that contribute to outcomes and those that are superfluous [3]. This warranted the creation of an open access database of e-mental health systems for the treatment and prevention of major depressive disorder (MDD), enabling the systematic examination of the composition of these systems and how this relates to their evaluation context and dropout rates.

2.2. METHOD

2.2.1. SEARCH STRATEGY

Considered for inclusion was primary research published in English describing e-mental health support systems, i.e., interventions with therapeutic content delivered on information and communication technology platforms for the prevention or treatment of major depressive disorder in adults developed and evaluated between 2000 and 2017. To ensure quality, only support systems having been scientifically evaluated with end users were included. The time frame was lower-bounded to give an accurate overview of the state of the art: in earlier systems the technological sophistication is limited by the availability of technology rather than being a design choice. Additionally, we excluded research on systems (1) serving only as a medium between therapist and user or between users (2) targeting children, targeting women with postpartum, perinatal, or prenatal depression, targeting caregivers or family members of people with depression, targeting a comorbid psychotic condition, (3) aiming to reduce stigma associated with depression, (4) diagnostic tools or decision aids, (5) lacking psychotherapeutic functionality (e.g., only supporting adherence to antidepressant treatment), (6) having a very narrow scope (i.e., system developed for one specific patient with very specific combination of conditions). The full search-query is provided in Appendix A on the database website [7].

An exhaustive search was conducted for articles and papers published up until 2017 describing eHealth interventions for depression on Scopus, PubMed, and Web of Science. A total of 5379 records were retrieved from the three databases (5359) and research syntheses in the field (20). All records were filtered first on title, then on abstract, and finally on full article by the first author, FB, hereinafter referred to as C1 (see Figure 2.1). Due to resource restrictions, a sample of the records was double coded at each stage by a second, independent coder, C2, a computer science student. Table 2.1 presents sample sizes and agreement scores at each stage. Sample sizes were determined by trading off the available resources against the margin of error of the percent agreement using the methods proposed by Gwet [8].

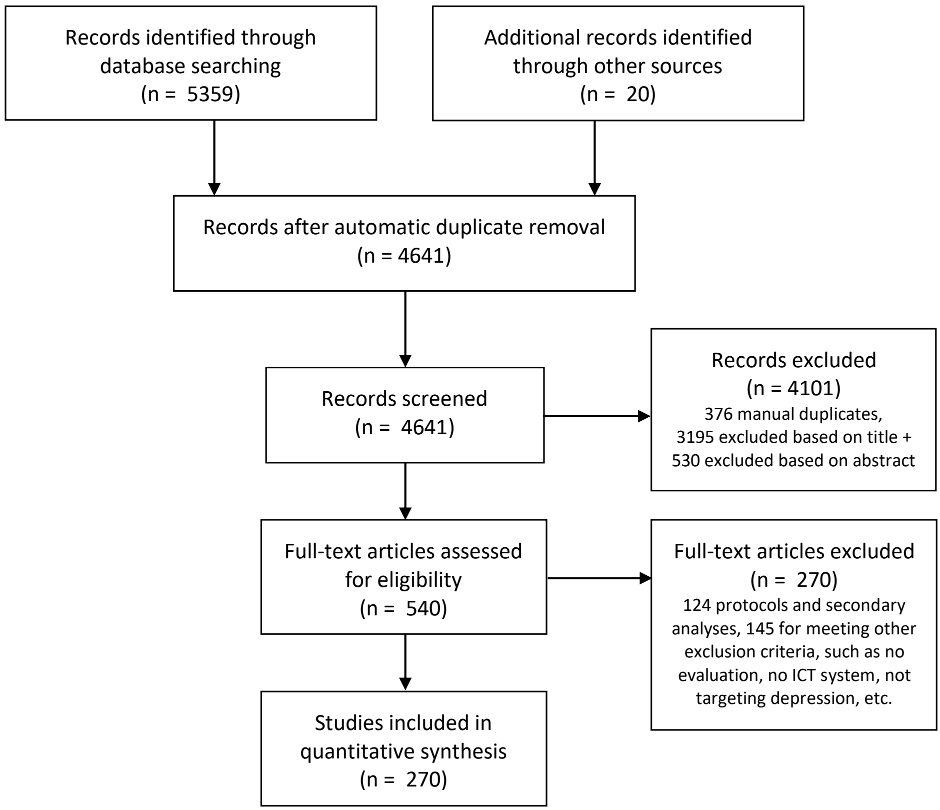

Figure 2.1: Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) diagram of the screening process, as completed by C1.

Abbreviation: ICT – *information and communication technology*

2.2.2. DATA EXTRACTION

Included articles were coded by C1 on a total of 45 variables and entered into a relational database. This database consists of 14 tables grouped into three larger clusters. The systems cluster details the e-mental health systems, their versions, their functionality, and their therapeutic purposes (four tables). In this cluster, systems were characterized on a macro-level (year of completion; whether its purpose is to prevent, treat, or monitor; whether it is guided, unguided, or an adjunct to face-to-face therapy; etc.) and on a micro-level, i.e. their functions. An instantiation of a function is its concrete implementation in a system. Functions were of two types: support functions and intervention functions. Support functions aim at increasing adherence to the intervention. They were again categorized into four subtypes: support functions for treatment planning (e.g., scheduling of sessions), treatment execution (e.g., reminders), monitoring (e.g., monitoring of symptoms), and social support (e.g., therapist support). To determine possible support functions in the domain, adherence strategies as defined by, for example, Oinas-Kukkonen *et al.* [9] were considered. Intervention functions support patient activities aimed at reducing depression symptoms. They are linked to specific classes of therapeutic interventions. The second cluster of tables in the database is the evaluations cluster, detailing the empirical studies of the systems in the systems cluster, their design, the employed measurement instruments, and dropout rates per study arm (four tables). Finally, the publications cluster details articles describing systems and their evaluations and the authors of these articles (three tables). The remaining three tables link systems to evaluations, systems to publications, and publications to evaluations.

Of the 45 variables, 41 were identified as low-inference (e.g., intended duration of the intervention) as they could be extracted directly from the literature, and 4 were identified as high-inference, since extraction required interpretation of the literature. The four high-inference codes were (1) identifying functions in a system description (e.g., SuperBetter implements the intervention function *3 Good Things*), (2) classifying an instantiation as a particular function (e.g., “Participants were instructed to follow one module per week" describes Tunneling), (3) linking an intervention function to a therapeutic intervention class (e.g., *3 Good Things* is a technique in Positive Psychology), and (4) assigning a degree of technological sophistication to an instantiation. For each of the high-inference variables, a random sample was selected and double coded by a second coder to assess reliability.

To *identify functions in a system description*, a graduate student in clinical psychology, C3, was provided with articles describing systems and the complete list of 184 functions. The percentage agreement was calculated as the percentage of overlap in all assigned functions of C1 and C3. To *classify an instantiation of a function*, C3 was again provided with the list of 184 functions as well as with a list of 125 snippets of text from articles describing functions. If C1 and C3 appointed the same function to the description of an instantiation, it was scored as agreement. In *linking a therapeutic intervention class to an intervention function*, C3 was provided with a list of all intervention functions and a set of 25 possible therapeutic intervention classes. While C1 always only selected one therapeutic intervention class, C3 was permitted to select multiple. Coder agreement was calculated by coding as agreement whenever C1’s class was a subset of the class(es) assigned by C3. C3 received no training for the tasks other than a detailed coding manual. Furthermore, C1 *rated all instantiations of functions on their degree of technological sophistication* with the respective e-mental Health Degree of Technological Sophistication scale (eHDTS) developed specifically for this purpose1. Reliability was examined by regarding agreement in a sample that was re-coded by C2.

Concurrent validity of the eHDTS scales was examined by correlating eHDTS scores assigned to instantiations of functions with knowledge of the scale levels (informed scores) with rank scores obtained when coders with computer science or related degrees were asked to sort the same instantiations according to their intuitive understanding of "technological sophistication" (naïve scores). Each of five coders, C4-8, received only instantiation descriptions of a specific category (e.g., only intervention functions or only treatment planning functions). C2, however, received a large sample of component descriptions taken from all the categories to allow for the examination of comparability of the different eHDTS scales. At least one week after the naïve sorting, the coders were asked to rate the same sample component on their respective eHDTS. Spearman correlations were calculated to examine agreement between informed and naïve scores within coders (Intra-Coder Correlation). All raw data and analysis scripts can be accessed online [10].

2.3. RESULTS

2.3.1. RELIABILITY ANALYSES

Both the screening procedure and high-inference codes were subjected to a reliability analysis. Table 2.1 presents the agreement scores obtained between the coders. In the literature screening process, coders agreed in approximately 80 % of cases. Inter-coder reliability was moderate to substantial according to the classification proposed by Landis *et al.* [11]. Since agreement and reliability could only be assessed on a sample, particularly false positives (C1 excluded while C2 included, indicating that other relevant articles of the population may have been missed entirely by C1 and hence may not be in the database) had to be examined closely. Four reasons could be identified: mistakes by C1 (full-paper: 1 record), mistakes by C2 (title: 2 records), C2 misunderstanding an exclusion criterion (title: 4 records, abstract: 2 records, full-paper: 3 records), and precaution on the part of C2 as he was instructed to include records when in doubt (title: 5 records, abstract: 3 records). Precautiously included records were unanimously excluded by re-evaluation at the next filtering stage. False negatives were not analyzed further as they did not pose a threat to database content.

2.3.2. SCALE VALIDATION

To determine concurrent validity of the different eHDTS scales, we correlated the informed scores of the coders with their own naïve scores (intra-coder correlation) as well as with the informed scores of C1 (inter-coder correlation). Table 2.2 shows the Spearman correlations for each scale.

2.3.3. DISTRIBUTION OVER SYSTEM FUNCTIONS

Analyses concerning the content of the database exceed the scope of this work. We therefore only briefly describe the distribution over function instantiations here. As can be seen in the Population column of Table 2, approximately 60 % of function instantiations across systems are of the intervention type. Of the support instantiations, 3 % support the user in establishing adherence strategies initially, while more than half aim at reeling the user back in. In line with research indicating that adherence to the

Table 2.1: Results of the reliability analyses conducted for the literature screening process and selected variables coded for in the database. For assigning a degree of technological sophistication to a function instantiation, a weighted Cohen's κ was calculated since the scale is ordinal.

	Coders	**Population**	**Sample**	**Percent Agreement, 95% CIs**	**Cohen's kappa, 95% CIs**
Literature Screening					
Title Filter	C1, C2	4266	100	.81 [.71,.87]	.50 [.31,.69]
Abstract Filter	C1, C2	1071	44	.80 [.64,.89]	.58 [.34,.82]
Full Article Filter	C1, C2	541	25	.84 [.60,.92]	.69 [.42,.95]
Coding					
Identify functions of system	C1, C3	273	10	.37 [.30,.43]	-
Classify function of instantiation	C1, C3	2224	125	.58 [.49,.66]	.57 [.48,.65]
Map intervention functions to therapeutic intervention class	C1, C3	141	141	.81 [.73,.87]	-
Assign degree of technological sophistication to instantiation	C1, C2	2224	132	.48 [.39,.55]	.60 [.51,.69]

Table 2.2: Results of the scale validation conducted to assess concurrent validity of the degree of technological sophistication scales. Confidence intervals were obtained by bootstrapping.

	Coders	**Population**	**Sample**	**Intra-Coder Correlation, 95% CIs**	**Inter-Coder Correlation, 95% CIs**
Intervention	C1, C4	1344 (60%)	27	.59 [.21,.80]	.81 [.58,.92]
Treatment Planning	C1, C5	29 (1.3%)	20	.82 [.62,.92]	.70 [.21,.92]
Treatment Execution	C1, C6	445 (20%)	27	.52 [.08,.78]	.67 [.31,.86]
Monitoring	C1, C7	140 (6.3%)	29	.27 [-.14,.61]	.52 [.07,.78]
Social support	C1, C8	266 (12%)	29	.52 [.11,.79]	.65 [.24,.87]
All types combined	C1, C2	2224 (100%)	117	.47 [.31,.60]	.58 [.43,.70]

systems is higher when human support is included, one third of support instantiations strive to provide human contact.

2.4. DISCUSSION AND CONCLUSION

The EHealth4MDD database is a systematic inventory of e-mental health systems for the treatment and prevention of depression. It contains 267 such systems with a total of 2224 function instantiations. Approximately 60 % of these functions are of a psychotherapeutic nature while 40 % aim to support the user in adhering to the intervention. In the literature screening process for populating the database, moderate to substantial reliability was obtained. Double coding of high-inference codes yielded satisfactory percent agreement scores in light of the vast number of possible categories. Therefore, the findings show clear consistency between coders and, as common in high-inference coding, some degree of individual subjectivity. Significant correlations of naïve with informed ratings indicate that four of the five different eHDTS scales capture the intuitive understanding of *technological sophistication* as held by those with a computer science or related degree. Furthermore, correlations between pairs of coders using the scales were significant and high on the same four scales, indicating that technological sophistication can be reliably assessed with the scales. However, the database is not without limitations as a single coder coded most data and reliability of this coding could only be assessed on samples. This resulted in large differences between coders in exposure to the data and therefore possibly lower scores of intercoder agreement and reliability than might otherwise be expected.

For researchers in both computer science and clinical psychology, the database presents a useful tool for the systematic study of e-mental health interventions for depression. It allows for a better understanding of system composition and of how functionalities contribute to clinical outcomes. Since the database is open access and implemented as a standard MySQL database, it can be linked with other databases, for example, databases of clinical trials. Furthermore, the accessibility allows the research community to contribute to the maintenance of the database.

Bibliography

1. Burger, F., Neerincx, M. A. & Brinkman, W. P. Ehealth4MDD: A database of e-health systems for the prevention and treatment of depressive disorders. *Annual Review of CyberTherapy and Telemedicine* **16,** 18–24 (2018).
2. World Federation For Mental Health (WFMH). Depression: A Global Crisis. *World Mental Health Day* (2012).
3. Schröder, J., Berger, T., Westermann, S., Klein, J. P. & Moritz, S. Internet interventions for depression: New developments. *Dialogues in Clinical Neuroscience* **18,** 203–212 (2016).
4. Richards, D. & Richardson, T. Computer-based psychological treatments for depression: A systematic review and meta-analysis. *Clinical Psychology Review* **32,** 329–342 (2012).
5. Foroushani, P. S., Schneider, J. & Assareh, N. Meta-review of the effectiveness of computerised CBT in treating depression. *BMC Psychiatry* **11** (2011).
6. Zhao, D., Lustria, M. L. A. & Hendrickse, J. Systematic review of the information and communication technology features of web- and mobile-based psychoeducational interventions for depression. *Patient Education and Counseling* **100,** 1049–1072 (2017).
7. Burger, F. *Website of the eHealth4MDD Database* `http://insyprojects.ewi.tudelft.nl:8888/`. (2018).
8. Gwet, K. L. *Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring the Extent of Agreement Among Raters* (Advanced Analytics, LLC, 2010).
9. Oinas-Kukkonen, H. & Harjumaa, M. Persuasive systems design: Key issues, process model, and system features. *Communications of the Association for Information Systems* **24,** 485–500 (2009).
10. Burger, F., Neerincx, M. A. & Brinkman, W.-P. *Reliability and validity analyses for the coding of information entered into the Ehealth4MDD database* 4TU.ResearchData `https://doi.org/10.4121/uuid:7e7e91ab-7afc-4b48-8915-e2bc80b23c99`. (2018).
11. Landis, J. R. & Koch, G. G. The Measurement of Observer Agreement for Categorical Data. *Biometrics* **33,** 159–174 (1977).

3

Technological State of the Art of e-Mental Health for Depression: Systematic Literature Review

This chapter is based on Burger, F., Neerincx, M. A. & Brinkman, W.-P. Technological State of the Art of Electronic Mental Health Interventions for Major Depressive Disorder: Systematic Literature Review. *Journal of Medical Internet Research* **22,** e12599 (2020)

3

ABSTRACT

Background: *Electronic mental (e-mental) health care for depression aims to overcome barriers to and limitations of face-to-face treatment. Owing to the high and growing demand for mental health care, a large number of such information and communication technology systems have been developed in recent years. Consequently, a diverse system landscape formed.*

Objective: *This literature review aims to give an overview of this landscape of e-mental health systems for the prevention and treatment of major depressive disorder, focusing on three main research questions: (1) What types of systems exist? (2) How technologically advanced are these systems? (3) How has the system landscape evolved between 2000 and 2017?*

Methods: *Publications eligible for inclusion described e-mental health software for the prevention or treatment of major depressive disorder. Additionally, the software had to have been evaluated with end users and developed since 2000. After screening, 270 records remained for inclusion. We constructed a taxonomy concerning software systems, their functions, how technologized these were in their realization, and how systems were evaluated, and then, we extracted this information from the included records. We define here as functions any component of the system that delivers either treatment or adherence support to the user. For this coding process, an elaborate classification hierarchy for functions was developed yielding a total of 133 systems with 2163 functions. The systems and their functions were analyzed quantitatively, with a focus on technological realization.*

Results: *There are various types of systems. However, most are delivered on the World Wide Web (76 %), and most implement cognitive behavioral therapy techniques (85 %). In terms of content, systems contain twice as many treatment functions as adherence support functions on average. Furthermore, autonomous systems, those not including human guidance, are equally as technologized and have one-third less functions than guided ones. Therefore, lack of guidance is neither compensated with additional functions nor compensated by technologizing functions to a greater degree. Although several high-tech solutions could be found, the average system falls between a purely informational system and one that allows for data entry but without automatically processing these data. Moreover, no clear increase in the technological capabilities of systems showed in the field, between 2000 and 2017, despite a marked growth in system quantity. Finally, more sophisticated systems were evaluated less often in comparative trials than less sophisticated ones (*$OR = 0.59$*).*

Conclusions: *The findings indicate that when developers create systems, there is a greater focus on implementing therapeutic treatment than adherence support. Although the field is very active, as evidenced by the growing number of systems developed per year, the technological possibilities explored are limited. In addition to allowing developers to compare their system with others, we anticipate that this review will help researchers identify opportunities in the field.*

3.1. INTRODUCTION

Between 2000 and 2017, researchers have reported more than 100 software interventions for depression in the scientific literature. Although all these systems have the same objective, they vary widely in both content and in the way the content is delivered. Taken together, they thus form a diverse landscape. But what does this landscape actually look like? The purpose of this literature review is to map the terrain by exploring the technological state of the art of electronic mental (e-mental) health interventions for depression.

The systems under study here strive to meet a globally growing need for depression care. The illness affects approximately 300 million people worldwide. Its high lifetime prevalence and high disease burden are further exacerbated by additional episodes often following the first. This renders the pervasive provision of treatment and prevention means imperative. However, the World Health Organization estimates that, currently, half of those suffering from depression are receiving inadequate or no treatment [2].

Information and communication technology (ICT) may present a viable solution to the shortage. The rapid dissemination of ICT over the course of the past two decades has led researchers to explore the provision of therapeutic content on these platforms. Unlike face-to-face treatment, such support systems are scalable, easily accessible, cheap, and standardized, and they can reduce the fear of stigmatization, since they can be used in private and at one's own convenience [3]. In addition to these benefits, numerous meta-analyses attest to the effectiveness of the interventions [4–6].

As a consequence of the high research interest, many systems have been developed to treat or prevent depression. Each system presents a unique solution. In light of this, several recent literature surveys point out that an analysis of the system landscape is in order, as there is little insight into the makeup of systems [3, 7, 8]. Where systems have been reviewed to date, authors have typically adopted one of two core perspectives. Syntheses with a *clinical psychology* perspective have addressed the effectiveness of different types of interventions [4, 9, 10]. Syntheses with a (*persuasive*) *technology* perspective, on the other hand, have addressed the functionality of systems, such as persuasive technology elements [8] or communication modality [11]. This systematic literature review takes the latter perspective. However, rather than studying the implementation or impact of a specific type of function in depth, it compares entire systems on their technological implementation. In doing so, e-mental health systems for depression are regarded as compositions of functions and assessed in terms of their technological realization. The support systems reported in the literature thus form the population under study. The main goal of this review, then, is to provide a comprehensive overview of the system landscape and its technological state. In addition, it identifies some of the challenges and opportunities for the field. However, linking the degree to which systems present high-tech solutions with clinical outcomes is outside of the scope of this review. Nevertheless, with the introduced system characterization and technological sophistication metric, a first step toward such studies is taken. From the extensive, domain-specific analysis presented here, we particularly expect researchers to benefit who are setting out to develop or study support systems for depression. It allows them to compare their system with those already in use and to identify underexplored aspects of these systems. To this end, the following three research questions are addressed:

1. What types of ICT systems for the treatment and prevention of depression have been developed?
2. How technologized are these systems?
3. How has the system landscape evolved between 2000 and 2017?

3.2. METHODS

3.2.1. LITERATURE IDENTIFICATION AND CODING

In this section, we focus on the literature search and filtering as well as coding of data pertaining to the analyses in this study. A detailed account of the construction, the structure, and the information contained in the open-access, relational database that was created for this analysis can be found in Burger *et al.* [12] and on the accompanying website [13].

IDENTIFICATION

The exhaustive search for potentially relevant literature made use of three databases: Scopus, PubMed, and Web of Science. It included English language journal articles, conference papers, and theses published between 2000 and 2017 that are presenting primary research that was conducted with support systems for the prevention or treatment of major depressive disorder or dysthymia in adults. To ensure that systems were actually created and functional at some point, we only considered the literature that reported the results of a system evaluation with end users. Therefore, systems that only had published study protocols available at the time of the search (early 2017) did not qualify. Lists of search terms comprised words around the following concepts that were central to the research interest: ICT, Health Condition, Purpose, Evaluation (Appendix A). They were expanded with controlled vocabulary terms where applicable. Systems were excluded if they were (1) only employing technology for mediated communication, (2) targeting children, postpartum or pregnant women, caregivers of depressed patients, or patients with comorbid psychotic conditions, (3) only aiming to reduce stigma, (4) serving only as diagnostic tools or decision aids, (5) addressing only antidepressant treatment, and (6) having an otherwise too narrow scope, for example, a system developed for a single patient with a specific combination of comorbid conditions.

The three queried databases returned a total of 5359 documents. Forward and backward reference searches on previous literature reviews and meta-analyses yielded an additional 20 records. After the removal of duplicates, 4256 records remained for screening. A lenient inclusion protocol at the title and abstract stages allowed for the inclusion of as many articles as possible concerning a system. Therefore, the exclusion of articles describing study protocols and secondary analyses only occurred at the full paper screening, but they were kept as additional references for clarification purposes. The first author, with a cognitive science background, screened all records at the title, abstract, and full-text stages (see PRISMA [14] diagram in Figure 2.1). A second, independent coder, with a computer science background, double coded a random selection at each stage. Intercoder agreement ranged from 80 % to 84 %, with moderate-to-substantial intercoder reliability (Cohen's κ between 0.50 and 0.69). Appendix B includes a complete list of all 270 articles included in the final synthesis.

CODING

To provide an overview of the aspects of software systems for depression considered here, a simplified taxonomy is presented in Appendix C. The extraction of information resulted in 45 coded attributes. These were either low-inference attributes, that is, the information could be directly copied from the paper, or high-inference attributes, that is, the coder needed to make inferences to arrive at the information. Second coders were neither used to refine the coding procedure nor to obtain a more reliable dataset. However, second coders did double code samples of the high-inference attributes to assess the reliability of the first coder, and the intercoder reliability measures reported here are to be regarded as an indication thereof. All analyses are based on the coding of the first author only.

A key task in the coding process was the division of systems into elementary functional parts, that is, functions. Herein, the focus was limited to functions pertaining to the higher-level layers of software architecture. For example, in the layered software architecture described in the Microsoft Application Architecture Guide [15], the functions would be located in the presentation and application layers. Cross-cutting concerns, such as security, were not considered. Additional criteria by which to evaluate software quality, for example maintainability, integration with other software, or software reliability, are also beyond the scope of this work. The construction of a classification hierarchy (Figure 3.1) preceded the coding process. At the fourth and highest level, two types of functions are possible: intervention functions, which aim to reduce depressive symptomatology in users, and support functions, which aim to increase adherence of the user to the intervention. An example of an intervention function would be the positive psychology exercise to count one's blessings every night, whereas an example of a support function would be sending text message reminders to encourage the user to engage with the system. At the third level, support functions further split into helping the user in (1) planning the intervention, (2) executing the intervention, (3) self-monitoring, or (4) connecting with other supportive people. Two more refined classification levels follow. At the lowest level, 41 classifications make up the support functions (Appendix D) and 145 classifications make up the intervention functions (Appendix E). Inspiration for the lowest-level support functions came largely from persuasive technology design frameworks [16–19], whereas therapy manuals (e.g., [20]) inspired the lowest-level intervention functions. These are often linked to therapeutic intervention frameworks, for example, Activity Planning is a technique of Behavioral Therapy. The intervention frameworks finally cluster into eight therapies at the penultimate level (Appendix E).

A second coder with a background in clinical psychology double coded two parts of the function identification task. The first part required spotting functions in the system description. Taking the functions that were found by the first coder as ground truth, interrater reliability was moderate on this part ($\phi = 0.54$, with a specificity [21] of $d' = 2.31$). The second part required labeling snippets of text that the first coder had identified as functions. For this part, interrater reliability on the 4 different function classification levels (Figure 3.1) was good on average ($\bar{\kappa} = 0.63$), ranging from moderate ($\kappa = 0.55$) to good ($\kappa = 0.72$).

Another key coding task concerned rating the degree to which each function was technologized. A set of scales, the e-mental Health Degree of Technological Sophistication

3

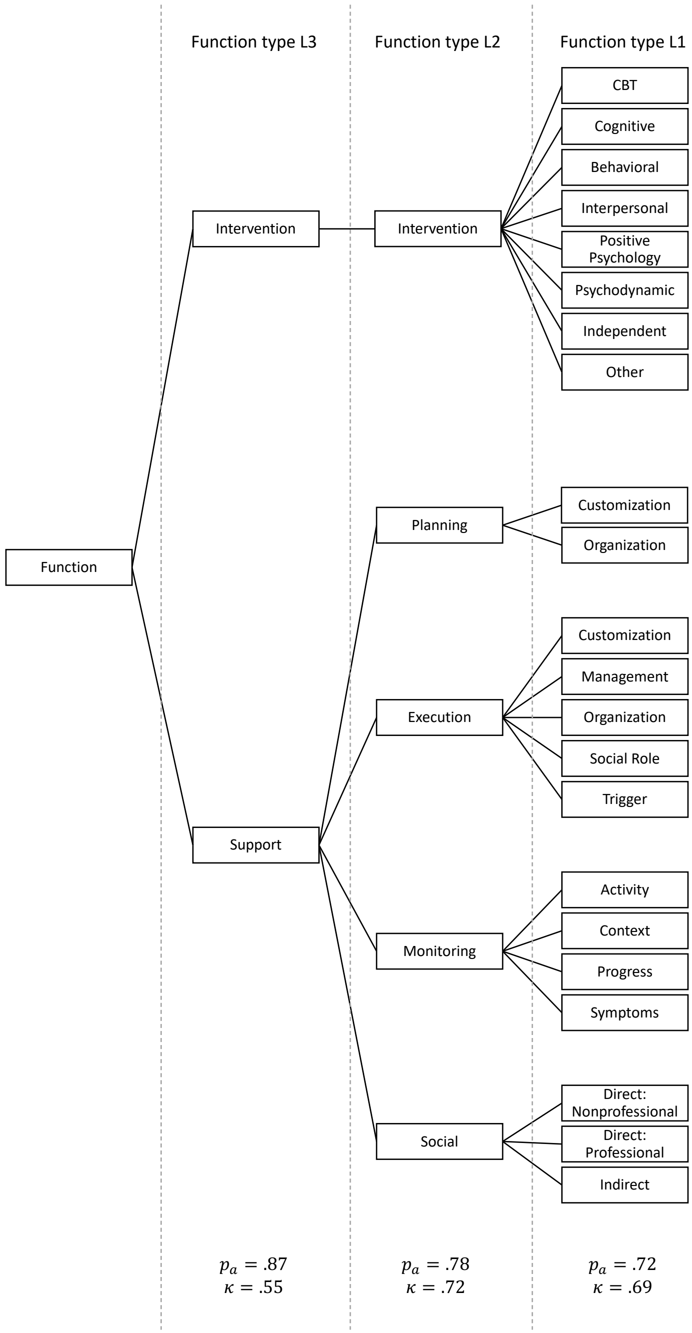

Figure 3.1: The top three levels of the function classification tree as well as the percent agreement and Cohen's kappa for the function classification task at each of the levels. Level 0 of the tree is specified in Appendix D (support functions) and Appendix E (intervention functions).

Abbreviation: CBT – *cognitive behavior therapy* (eHDTS) rating scales (Appendix F), were developed specifically for this task. They include one scale for intervention functions and four separate scales for each of the four types of support functions. Conceptually, the scales range from *offline* to *responsive on content* (Table 3.1). Although the emphasis in the interpretation of the eHDTS scales throughout this description is placed on the interactivity aspect, the actual scales are broader, also covering aspects such as responsiveness, personalization, data analysis, and data presentation. From here on, when directly describing the technological realization of systems or functions as measured by the scale, we refer to it as *technological sophistication*. In coding, a conservative approach ensured that the lower degree was assigned in case of doubt. Reliability levels were acceptable, with a mean correlation of 0.66 between coders. Furthermore, concurrent validity of the scales was supported by on-average moderate correlations ($\bar{r} = 0.53$) between ratings on these scales and ratings on an unlabeled ordinal scale, that is, leaving it open to coders to decide what the different levels of technological sophistication entail.

Table 3.1: The degrees of the e-mental health degree of technological sophistication rating scale, abstracted over the five different instantiations of this scale. A diary function serves as a hypothetical example. It should be noted that this is an abstract summary of the levels across several scales. It therefore does not capture the entire technological breadth of the different scales.

Degree	**Definition**	**Example**
0: Offline	The function is not provided through the system at all or is fully carried out by a human.	Diary sent by postal mail
1: Informational	The function is provided in an informational manner.	Diary can be downloaded as PDF
2: Data entry	The function is provided in an interactive manner but without processing of input from the user.	Diary can be filled on the Web and saved
3: Form response	The function is provided in an interactive manner with processing of meta-information.	Web-based diary that responds to the duration of typing
4: Content response	The function is provided in an interactive manner with processing of the content of user input.	Web-based diary that responds to the sentiment of text, which the user has written, for example, “It appears that this was a very negative experience for you."

Finally, one coder was provided with a list of function descriptions from all function types, without the function type label, and asked to assign a rating of technological sophistication to these on an unlabeled ordinal scale from 0 to 5 (uninformed). After two weeks, he was again invited to code the same functions with the appropriate scale for each function and each scale level defined (informed). The correlation between the uninformed and the informed rating ($r = 0.47$) provided some indication that, although each function type had its own eHDTS scale, the five scales were sufficiently similar to allow for aggregation and cautious comparisons on a system level.

Three low-inference attributes coded were the system version, the system build year, and the evaluation quality. A version was defined as a modification of the system offering different functionality. For example, Lemma *et al.* [22] created a version with human support and a version without it, whereas Currie *et al.* [23] offer different versions to support female or male patients by providing extra content for women. However, a system with an adaptive user interface based on gender was not regarded as two versions; it was regarded as one with a tailoring function. The system build year denotes the year in which systems were finalized, that is, the earliest year of operation mentioned in the earliest publication on the earliest version. Versions and systems are simply referred to as *systems* or *software* for legibility from here on. Most analyses to follow were conducted on the body of versions rather than systems. It is made explicit when this is not the case. Finally, the evaluation quality received a binominal coding of high and low. A high quality meant that the system was evaluated in a comparative trial, whereas a low quality meant that it was evaluated in a single-group trial. Comparative trials encompassed randomized controlled trials, randomized comparative trials, and nonrandomized comparative trials.

3.2.2. STATISTICAL ANALYSIS

We conducted quantitative analyses with R version 3.5. All data and the full analysis script are permanently stored for public access on a national database for research data with the 4TU Center for Research Data in the Netherlands [24]. Where distributions deviated markedly from normality, nonparametric tests were used. Furthermore, we report two estimated R^2 effect size measures where R^2 cannot be calculated exactly. For logistic regression models, Nagelkerke pseudo R^2 [24] was chosen, whereas, for multilevel models, the Level1 R^2, as proposed by Snijders and Bosker [25], was computed. When used, these are indicated as *Nagelkerke* R^2 and *Level1* R^2, respectively.

CHARACTERIZATION

To characterize systems, we regarded their composition in terms of functions and how systems differed depending on such factors as *guidance* or *system purpose*, that is, prevention or treatment. A Wilcoxon rank sum test compared the number of intervention functions with the number of support functions per system. In addition, two logistic regression models were fit. One determined whether a certain system purpose was more commonly occurring within a certain therapy type. The other tested whether autonomous or guided systems are represented to different degrees depending on purpose. Systems that include human guidance naturally have more functions, as guidance needs to be facilitated by the system somehow. This takes place by way of the direct social support functions. Thus, to allow for a fair comparison of the number of functions in autonomous versus guided systems, direct social support functions were excluded for the following three analyses. First, a linear regression examined the relationship between guidance and the number of functions of a system. Second, two more detailed analyses in the form of Wilcoxon rank sum tests considered this relationship separately for intervention and support functions.

TECHNOLOGICAL SOPHISTICATION

Technological sophistication was compared among the different types of functions, different types of systems, and different evaluation qualities. A correlation assessed the relationship between system size and technological sophistication, and linear regression models gave insight into the link between technological sophistication on the one hand and evaluation quality, guidance, or system purpose on the other. To contrast support and intervention functions, a multilevel linear model was fit using the function type as a fixed effect and allowing for random intercepts per system. Similarly, a one-way analysis of variance checked for differences in technological sophistication among the four different support types.

DEVELOPMENTS OVER TIME

Changes over time could take place both across and within systems. Two linear regression models examined development in size and technological sophistication across systems. Moreover, three multilevel linear models allowed studying development within systems. They determined whether size, technological sophistication, and evaluation quality changed across versions. Random intercepts modeled the nested relationship of versions within systems.

3.3. RESULTS

3.3.1. CHARACTERIZATION

In total, 133 systems with 259 versions were identified. Coding these systems on their key attributes led to the characterization presented in Table 3.2.

VERSIONS

Systems had two versions on average, but more than two-thirds (69.2 %, 92/133) only had one version. Thus, most systems seem to have been developed for a single research project. Only ten systems had five or more versions, for example, The Sadness Program with 13 versions, MoodGYM with 15 versions, and the Well-being Course with 18 versions.

INFORMATION AND COMMUNICATION TECHNOLOGY PLATFORMS

The World Wide Web was the most frequently employed platform, with 75.7 % (196/259) of the systems providing functionality on the Web and 6.2 % (16/259) of the systems providing responsive website content that could also be displayed appropriately on mobile phones. Emails were sent or received in 43.2 % (112/259) of systems. Following email, telephone (20.5 %, 53/259) and text messages (6.6 %, 17/259) were frequently used to reach out to users. Only 1.9 % (5/259) of the systems made use of storage media, such as CD and DVD and just as few exhibited technologies such as virtual agents (1.9 %, 5/259), virtual reality (0.8 %, 2/259), or connected to social media services (2.3 %, 6/259).

GUIDANCE

E-health software can include various types of human guidance or be entirely autonomous. Approximately half of all systems classified as the latter (47.5 %, 123/259). In the remaining systems, guidance was mostly provided by health care professionals. These were therapists in 24.3 % (63/259) of cases and practitioners of related professions, such as coaches, nurses, social workers, or clinical psychology students in 12.4 % (32/259) of cases. Less than 10 % (24/259) of guided systems were offered as adjunct systems, that is, systems that support face-to-face therapy. A total of 5.4 % (14/259) of systems were supported by technicians and other administrators, and only 1.2 % (3/259)

Table 3.2: The distributions over technology-related, therapy-related, evaluation-related key attributes of depression support system versions.

Technology Attribute	Value	**Therapy** Attribute	Value	**Evaluation** Attribute	Value
Total versions[a](N=133), mean(SD)	2.0(2.5)	*Comorbidity (N=259), n(%)*		Total studies, mean(SD)	1.2(0.9)
Technology (N=259), n(%)		None	181(69.9)	*Quality (N=259), n(%)*	
Offline	69(26.6)	Anxiety	43 (16.6)	Comparative	200(77.2)
World Wide Web	196(75.7)	Physical	24(9.3)	Noncomparative	74(28.6)
Email	112(43.2)	Nonspecific	7(2.7)	*Control group types (N=259), n(%)*	
Telephone	53(20.5)	Addiction[b]	5(1.9)	Attention controlled	73(41.7)
Computer	28(10.8)	Insomnia[c]	2(0.8)	Waitlist	69(39.4)
Text message	17(6.6)	*Purpose (N=259), n(%)*		TAU[d]	50 (28.6)
Mobile	16(6.2)	Treat	180(69.5)	*Measures (N=259), n(%)*	
App	14(5.4)	Prevent	76(29.3)	PHQ[e]	90(34.7)
Sensors	7(2.7)	After-care	3(1.2)	BDI[f]	74(28.6)
Social media	6(2.3)	Duration (weeks; N=210), mean(SD)	8.7(9.1)	CES-D[g]	65(22.0)
Virtual agent	5(1.9)	Total modules (N=218), mean(SD)	5.9(3.5)	Other depression measure	57(25.1)
Interactive voice response	5(1.9)	*Therapy class (N=259), n(%)*		Nondepression measure	31(12.0)
CD/DVD	5(1.9)	Independent	194(78.9)		
Virtual reality	2(0.8)	Behavioral	154(62.6)		
Undefined	4(1.5)	Cognitive	145(58.9)		
Support type (N=259), n(%)		Cognitive behavioral	124(50.4)		
Autonomous	123(47.5)	Interpersonal	43(17.5)		
Therapist	63(24.3)	Positive psychology	43(17.5)		
Professional	32(12.4)	Psychodynamic	7(2.0)		
Adjunct	24(9.3)	Other	5(2.8)		
Admin	14(5.4)				
Lay person	3(1.2)				
Total functions (N=259), mean(SD)	8.4(4.5)				
Function type (N=259), n(%)					
Intervention	246(95.0)				
Execution	214(82.6)				
Social	175(67.6)				
Monitoring	103(39.8)				
Planning	22(8.5)				
Sophistication (N=259), mean(SD)	1.6(0.6)				
Intervention	1.5(0.8)				
Execution	1.7(0.9)				
Social	1.5(0.9)				
Monitoring	2.1(1.1)				
Planning	1.8(1.0)				

[a] Conducted on systems instead of versions.
[b] Addiction is separated from physical illness, as it can be regarded as both a physical and mental illness.
[c] Insomnia is separated from physical illness, as insomnia is also a symptom of depression.
[d] TAU – treatment as usual
[e] PHQ – Patient Health Questionnaire
[f] BDI – Beck Depression Inventory
[g] CES-D – Center for Epidemiological Studies - Depression

of systems asked for support by a layperson, typically a peer, friend, or family member of the user.

SIZE AND FUNCTIONALITY

In terms of size, the average system offered eight functions (median= 8), with a range from one to 21. Furthermore, systems had, on average, six modules (median= 6) and an intended usage duration of slightly less than nine weeks (median= 8). Although nearly all software contained some intervention functions (95.0 %, 246/259) and some support functions (91.5 %, 237/259), the four support function types were not equally represented. A total of 82.6 % (214/259) of systems included execution support, such as reminders via text message. Social support functionality was provided by 67.6 % (175/259) of systems. This was either direct, whereby the user communicated with a human, or indirect, whereby the user could, for example, see that other people had performed the program before them. The least represented support function type (8.5 %, 22/259) was planning support. A typical example of a planning support function is setting up a treatment schedule at the outset of the intervention. Within systems, intervention functions were dominant: systems contained, on average, twice as many intervention functions as support functions ($V = 15079$, $p < .001$, $r = 0.09$). In addition, unguided and guided systems differed in their composition, with the former only having 63 % of the number of functions of the latter ($F_{1,233} = 51.34$, $p < .001$, $R^2 = 0.18$). This effect showed for both intervention ($U = 3467$, $p < .001$, $r = 0.41$) and support ($U = 3839.5$, $p < .001$, $r = 0.28$) functions.

THERAPEUTIC ASPECTS

Although the literature search and filtering focused on systems aiming to reduce depressive symptoms, only 69.9 % (181/259) of the identified software targeted depression exclusively. A total of 9.3 % (24/259) of the remaining systems specifically targeted users with a comorbid physical illness (e.g., cancer, multiple sclerosis, and diabetes). A few systems supported comorbidities in general (nonspecific, 2.7 %, 7/259). Of all systems, 16.6 % (43/259) also considered anxiety. However, other mental comorbidities were excluded from the reviewed literature, as they typically formed the primary treatment objective (e.g., in systems targeting psychotic conditions and depression simultaneously).

The most prominently represented intervention functions, present in 78.9 % (194/259) of systems, were unrelated to specific therapies, that is, they could be categorized with many or all different therapies, such as learning to recognize one's own symptoms or preventing relapses. A large percentage of software made use of behavioral (62.6 %, 154/259), cognitive (58.9 %, 145/259), and cognitive behavioral (50.4 %, 124/259) functions. Taken together, techniques related to cognitive behavioral therapy (CBT) were present in 84.9 % (209/259) systems. Techniques from psychodynamic approaches were rare (2.0 %, 5/259), as were life reviewing or hypnosis techniques (together present in 2.8 %, 7/259, denoted by others in Table 3.2). A total of 69.5 % (180/259) of systems had the purpose of treating depression and 29.3 % (76/259) of systems had the purpose of preventing it. Only 1.2 % (3/259) of the systems aimed to support patients in maintaining a depression-free state. The system purpose was related to the therapeutic approach ($\chi^2_7 = 34.1$, $p < .001$, Nagelkerke $R^2 = 0.24$). Systems with Positive Psychology techniques were more often intended for prevention than for treatment. This was not the case for systems with techniques from other therapies (Figure 3.2).

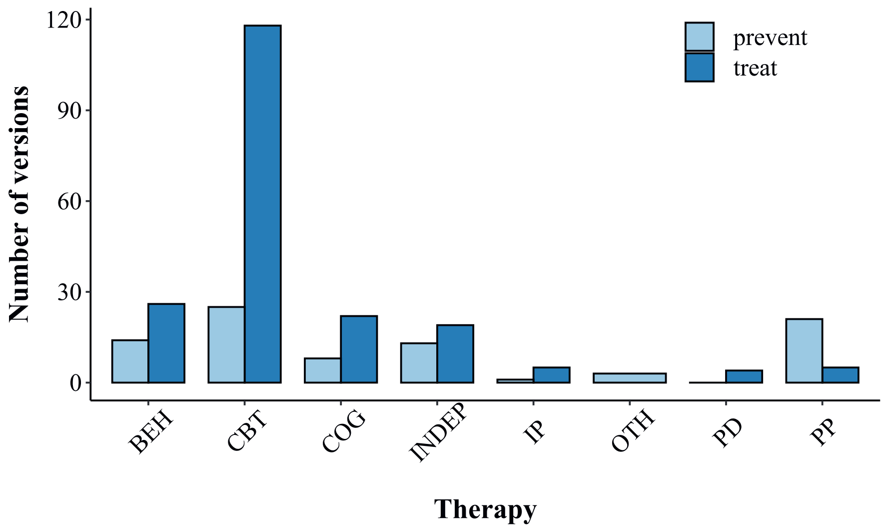

Figure 3.2: The number of versions with the purpose of preventing or treating depression per therapy. A detailed list of the therapy subtypes for each of the therapy categories listed here can be found in Appendix E

Abbreviations: BEH – *behavioral therapy*, CBT – *cognitive behavior therapy*, COG – *cognitive therapy*, INDEP – *independent of specific therapeutic theory*, IP – *interpersonal therapy*, OTH – *other*, PD – *psychodynamic therapy*, PP – *positive psychology*.

Similarly, guided systems ($\chi^2_1 = 10.0$, $p = .002$, Nagelkerke $R^2 = 0.05$) were more often used in treatment systems ($n_{guided} = 105$, $n_{unguided} = 75$), whereas unguided ones were used more in prevention ($n_{guided} = 28$, $n_{unguided} = 48$).

EVALUATION

Systems were often evaluated only once with end users (86.9 %, 225/259) and, for the largest part, in comparative trials (77.2 %, 200/259). In controlled trials, attention control (41.7 %, 73/175) and waitlist (39.4 %, 69/175) were similarly common, whereas treatment as usual (28.6 %, 50/175) was less frequent (Table 3.2). In total, 72.2 % (187/259) of systems were evaluated in controlled trials. Appendix G comprises two tables ranking systems according to the number of evaluations and the total number of participants who participated in these studies.

Although 21 different measures assessed depressive symptomatology across studies, the most frequent by far were the Patient Health Questionnaire [25], Beck's Depression Inventory [26], and the Center for Epidemiological Studies Depression Scale [27]. An additional 11 measures were depression related, determining such things as fatigue, rumination, stress, or quality of life. Finally, 12.0 % (31/259) of systems were evaluated in studies having primary outcomes other than depression, such as usability.

DESCRIPTION OF A FICTIONAL, PROTOTYPICAL SYSTEM

For illustration purposes, we outline here a fictional, prototypical depression treatment system by combining insights from the qualitative reading of the articles and the quantitative analyses. This is intended to serve as a narrative description of the taxonomy provided in Appendix C. However, it must be noted that this is a simplification and much variation exists among the systems. A prototypical system takes a CBT approach and might comprise six modules, one of which is released every week. The modules can be accessed on a website. The participant is made aware of the presence of a new module via email; thus, the participant is reminded to adhere to the treatment. Modules might cover topics such as activity scheduling, learning to detect automatic thoughts, cognitive restructuring, problem solving, psychoeducation concerning depression and the therapeutic approach, and relapse prevention. Each module comes with exercises that are submitted to be checked by a therapist or similar, who again provides feedback via email. The website might include a small calendar application for the purposes of activity scheduling and a diary application for the purposes of thought recording. In these applications, the user can enter and save information. Once a week, the participant is asked to complete a depression scale, and the therapist is notified if suicidal ideation is detected. The remaining questions are averaged and presented to the user as a mood graph on the landing page. This sketched system would have an average eHDTS score of around 2. For each of the eHDTS levels, a similar, fictional description of possible functions scoring at this level can be found in Appendix H. This is intended to provide a more concise and tangible description than Appendix F can and to further concretize the taxonomy presented in Appendix C.

3.3.2. TECHNOLOGICAL SOPHISTICATION

SYSTEMS

The average system comprised mostly functions providing information to the user without collecting and interpreting information from the user. This is further detailed in Figure 3.3. Almost all interventions had the majority of their functions delivered through technology, that is, hardly any system scored below 1 on technological sophistication. However, only 21.1 % (28/133) of systems had a sophistication level above 2, indicating that they were responsive to activities and information coming from the user. These systems comprised, for the most part, interventions inspired by CBT or closely related therapies. In fact, CBT systems lead the list of the most technologically advanced systems, even when adjusting for the number of functions (Table 3.3). The top two systems in both rankings are Help4Mood [28] and Deprexis [29]. The latter is a commercial system aiming to mimic the structure of face-to-face CBT therapy, whereas the former is a self-monitoring system that includes a virtual conversational agent. Both presented high-tech solutions according to the eHDTS scale, as they adapted the intervention to the users' indicated interests and needs (Deprexis) or to the self-monitoring data from users (Help4Mood). To allow researchers to compare their own system, Appendix I provides the eHDTS score per cumulative percentage decile of systems for both the weighted and unweighted system means. That is, when knowing the average weighted or unweighted eHDTS score of their system, researchers can use the table to determine which decile of systems their system scores at, below, or above.

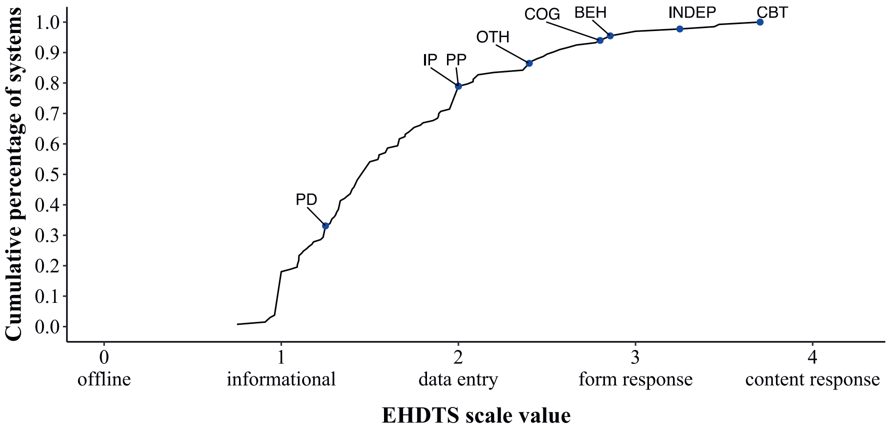

Figure 3.3: Cumulative density plot of all systems over the e-mental Health Degree of Technological Sophistication (eHDTS) scale. This analysis was conducted on the unweighted average of technological sophistication of the systems. Labeled dots show the highest scoring system within a specific therapy, as indicated by the label.

Abbreviations: BEH – *behavioral therapy*, CBT – *cognitive behavior therapy*, COG – *cognitive therapy*, eHDTS – *e-mental Health Degree of Technological Sophistication*, INDEP – *independent of specific therapeutic theory*, IP – *interpersonal therapy*, OTH – *other*, PD – *psychodynamic therapy*, PP – *positive psychology*.

Technological sophistication was not linked to the number of functions ($r_{257} = 0.01$, $p = .83$), the system purpose ($\chi^2_1 = 0.2$, $p = .69$), or guidance ($\chi^2_1 = 3.0$, $p = .08$). However, it did relate to the evaluation quality ($\chi^2_1 = 6.1$, $p = .01$, Nagelkerke $R^2 = 0.03$). More technologically sophisticated systems were less likely ($OR = 0.59$) to have been evaluated in comparative trials than less technologically sophisticated systems. Furthermore, when regarding randomized controlled trials (RCTs) in particular, we found that 80.8 % (139/172) of RCTs evaluate systems that score below data entry level on average, with the respective percentage of RCTs per eHDTS interval being the following: [0,1)–4 %; [1,2)–77 %; [2,3)–16 %; and [3,4)–3 %.

FUNCTIONS

Support functions ($mean = 1.73$, $SD = 1.06$) scored higher in technological sophistication than intervention functions ($mean = 1.43$, $SD = 0.88$), although this effect was small ($F_{1,1903} = 38.11$, $p < .001$, Level1 $R^2 = 0.03$). An equally small effect was observed when comparing the four types of support functions on their technological sophistication ($F_{3,619} = 8.46$, $p < .001$, Level1 $R^2 = 0.04$). Monitoring support functions had the highest average degree of technological sophistication ($mean = 2.1$, $SD = 1.1$, see Table 3.2). This indicates that monitoring functions were mostly technologically sophisticated to the extent that they reported data back to the user, but they neither interpreted data nor used the data to adapt the intervention. Social support ($mean = 1.5$, $SD = 0.9$) and

Table 3.3: Ranking of the 10 systems with the highest degree of technological sophistication in the database, first based on average e-mental Health Degree of Technological Sophistication (eHDTS) score (M) and then based on a weighted eHDTS score (M_w), trading off eHDTS against the number of functions in a system. The analyses were conducted on the basis of systems rather than versions. We advise some caution in taking this table at face value, as it is based on the aggregated eHDTS scores with some of the scales only having moderate interrater agreement.

Rank	Unweighted System	Therapy	n_f [a]	M [b]	Weighted System	Therapy	n_f	M	M_w [c,d]
1	Help4Mood [28]	CBT [e]	13.5	3.70	Help4Mood [28]	CBT	13.5	3.70	2.31
2	Deprexis [29]	CBT	14.3	3.47	Deprexis [29]	CBT	14.3	3.47	2.31
3	MOSS App [30]	CBT	9	3.44	Buhrman [31]	CBT	20	1.95	1.85
4	Ahmedani [32]	MI [f], CBT	4	3.25	Building a Meaningful Life through BA [33]	BA [g]	20	1.90	1.81
5	DCAT ATA [34]	SM [h]	5	3.00	Shamekhi [35]	MFN [i]	13	3.00	1.80
6	Shamekhi [35]	MFN	13	3.00	Living to the full [8]	ACT [j]	14.5	2.62	1.77
7	Panoply [36]	CBT	7	2.86	MindBalance [37]	CBT	14	2.57	1.67
8	MyPAA [38]	PhA [k]	7	2.86	Space from Depression [39]	CBT	15	2.20	1.54
9	EVO [40]	CCT [l]	5	2.80	Mobilyze! [41]	BA	13	2.54	1.52
10	Daybuilder [42]	SM	6.5	2.77	MOSS App [30]	CBT	9	3.44	

[a] n_f – number of functions
[b] M – unweighted average
[c] M_w – weighted average
[d] To obtain the weighted average (M_w), the unweighted average (M) is weighted with the feature-scaled number of functions (n_f): $M_w = (n_f - min(n_f)/(max(n_f) - min(n_f))M$, with $min(n_f) = 1$ and $max(n_f) = 21$
[e] CBT – cognitive behavior therapy
[f] MI – motivational interviewing
[g] BA – behavioral activation
[h] SM – symptom monitoring
[i] MFN – mindfulness
[j] ACT – acceptance and commitment therapy
[k] PhA – physical activity
[l] CCT – cognitive control training

intervention functions (*mean* = 1.5, *SD* = 0.8) ranked the lowest in terms of technological sophistication (Table 3.2). In social support, the score translates to technology being typically either used to simply provide contact information to the user or to serve as a communication medium between human support and user. Intervention functions often took an informational form, possibly with a limited amount of interactivity, for example, clicking through pages or filling in a Web-based diary.

The most frequently implemented support functions were execution support pertaining to the management of user progress and risk, triggers, indirect social support, professional direct social support, and symptom monitoring (Figure 3.4). However, only management execution support and indirect social support were present at least once in systems of all different therapies. A barely implemented function type was planning support. Shifting the focus to intervention functions, most stem from CBT or related therapies or are independent of a specific therapeutic framework. CBT systems clearly dominate the field, with most of the different function types being implemented in numerous such systems (Figure 3.4). Yet, the average technological sophistication of functions (Figure 3.5) was not related to how frequently they were implemented ($r_{154} = 0.12$, $p = .12$). Thus, functions that are often implemented are neither more nor less technologically sophisticated, on average, than functions that are rarely implemented. However, the more often a function was implemented, the more often at least 1 of these implementations was responsive to interaction activity of the user, for example, time spent on platform, or even to the content of information provided by the user ($r_{154} = 0.43$, $p < .001$). For interested readers, Appendix J finally also demonstrates that nearly all of the different functions were implemented in a highly sophisticated manner in at least one system.

3.3.3. DEVELOPMENTS OVER TIME

In the past two decades, the field of e-mental health for depression has seen marked growth, with five times as many systems developed in 2014 as in 2000 (Figure 3.6). As several years typically lie between development and the publication of study outcomes, less emphasis may be given to numbers after 2014. The figure also demonstrates that systems were being reused and extended to a substantial degree only from approximately 2009 onward. This is further supported, when examining systems with at least five versions more closely (Figure 3.7). Only MoodGYM had evolved multiple versions before 2009. Different versions developed within the same year are an indication that they were created for the same study, often differing in only one function as an experimental manipulation.

Despite growth in the field in general, systems seemed to neither get larger ($F_{1,257} = 0.25$, $p = .62$) nor more sophisticated ($F_{1,257} = 1.88$, $p = .17$) with time. Within systems, growth was observed across versions, with each new version of a system having half of a function more than the previous one ($b = 0.50$, $F_{1,125} = 11.60$, $p < .001$, Level1 $R^2 = 0.06$). However, technological sophistication seemed to remain the same ($F_{1,125} = 1.96$, $p = .16$). Finally, the evaluation quality showed no relationship with the version number ($F_{1,136} = 0.07$, $p = .79$). Later versions therefore appeared to be no more or less frequently evaluated in comparative trials than earlier ones.

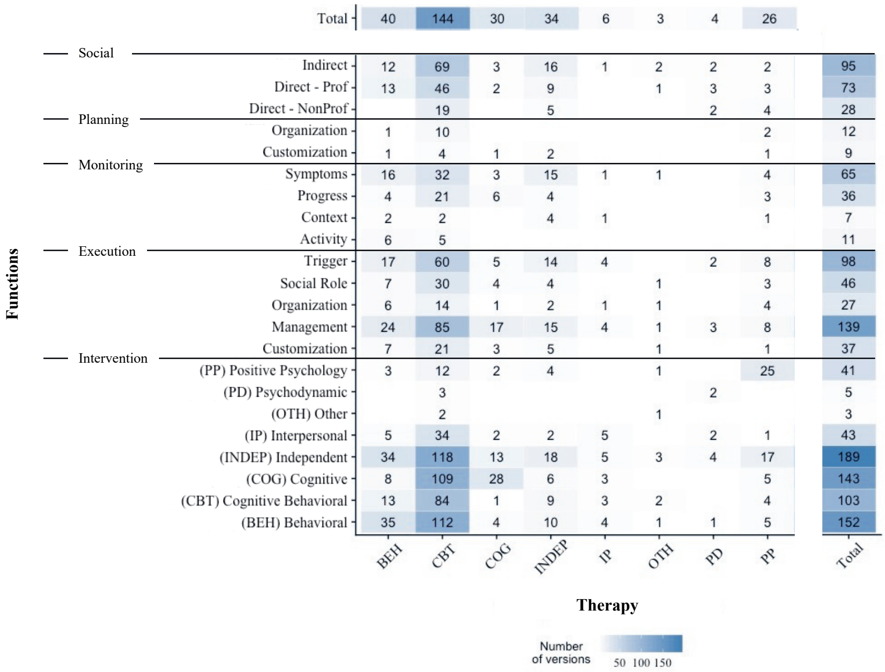

Figure 3.4: Heatmap of the frequency with which a specific type of function was implemented in a therapy across all systems of that therapy.

Abbreviations: BEH – *behavioral therapy*, CBT – *cognitive behavior therapy*, COG – *cognitive therapy*, INDEP – *independent of specific therapeutic theory*, IP – *interpersonal therapy*, OTH – *other*, PD – *psychodynamic therapy*, PP – *positive psychology*.

3.4. DISCUSSION

3.4.1. PRINCIPAL FINDINGS

Some limitations should be kept in mind when considering implications of the results. The first pertains to the coding of technological sophistication. Namely, the one-dimensional nature of the eHDTS scales can limit them in covering the full extent of the degree to which they reflect how technologized a function is. In monitoring support functions, for example, the scale captures how the system deals with the collected information but not how the data are obtained in the first place. Thus, whether monitoring data are collected via self-report or sensing does not influence the level of technological sophistication. However, as sensors and data analysis methods are becoming increasingly reliable, sensing will likely begin to play a crucial role in more automated, that is, more technologically advanced, systems [43]. In addition to monitoring, this is particularly to be expected in diagnosis and assessment systems [44, 45], which we have excluded in this review. Thus, in the future and especially when

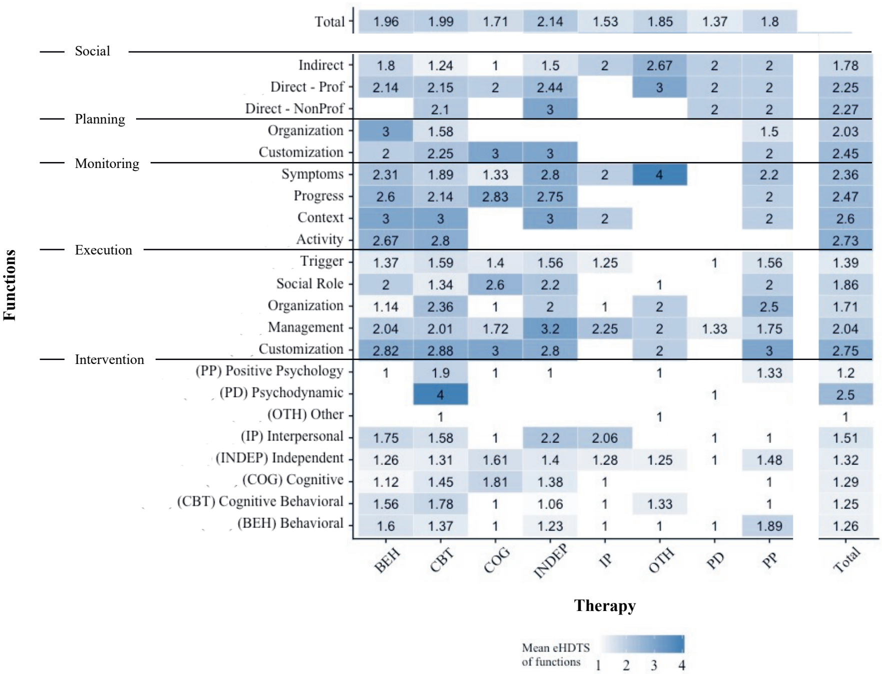

Figure 3.5: Heatmap of the average degree of technological sophistication per function type and therapy.

Abbreviations: BEH – *behavioral therapy*, CBT – *cognitive behavior therapy*, COG – *cognitive therapy*, INDEP – *independent of specific therapeutic theory*, IP – *interpersonal therapy*, OTH – *other*, PD – *psychodynamic therapy*, PP – *positive psychology*.

wishing to also study such systems, the manner in which data are collected should receive more attention in the monitoring scale. An additional point to consider when interpreting the results is the moderate reliability of some high-inference attributes. Although this is a limitation that might influence more detailed findings, such as the exact ranking of the systems according to their eHDTS score, we do not expect it to substantially affect the larger patterns found. However, by double coding samples, we have insight into the reliability of the estimates. In selecting a sample size for double coding, we have aimed for a 10 % margin of error for the reliability estimates, as suggested by Gwet [46] (see [12] for 95 % CI information for each estimate). Two final limitations concern the scope of the reviewed systems, as well as the scope of the reviewed functions. Systems developed for children and adolescents, women with depression during or following pregnancy, and those with comorbid psychiatric conditions, as well as systems developed before 2000, were excluded. In addition, we did not consider commercial systems that are not reported in the scientific literature. How well our findings generalize to these types of systems is therefore open to further investigation. As far as the scope of the reviewed functions is concerned, functions

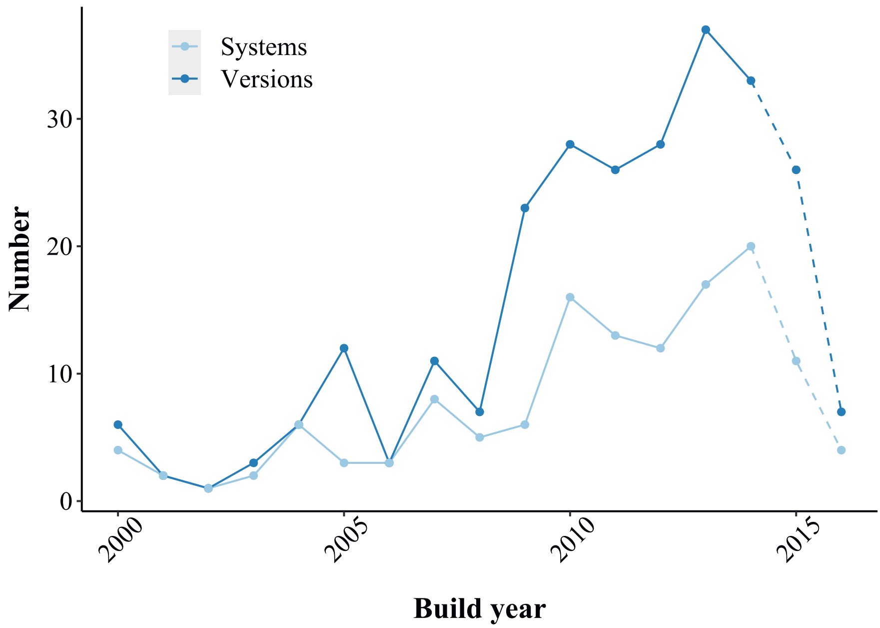

Figure 3.6: The number of systems and versions developed per year between 2000 and 2016..

pertaining to data security or the integration of the system with existing health software were not covered. Such aspects of the interventions were typically not found to be reported in the publications. As data security is becoming an important concern of software development and usage, we see a need for more consistent system reporting guidelines and an opportunity for reviews of future systems to subsequently investigate such functionality. In spite of these limitations, the outcomes of the analyses highlight some of the challenges and opportunities for the field of e-mental health for depression.

First, no clear progress in terms of system sophistication was observed between 2000 and 2017, within or across systems. A possible challenge for progress might lie in the short-term approach to system development in the field. In a long-term approach, multiple versions with substantial changes in functionality could be expected. Early versions would be tried in pilot studies, improved, and only eventually tested in an RCT. However, this is not what we found. Despite often proving effective in RCTs, two-thirds of the systems are not evolved and retested (e.g., [47–49]). In addition, in systems that do have multiple versions, systems are often extended only by a function for hypothesis testing among versions, and versions do not differ in technological sophistication. Finally, there was also no association between the evaluation level and the version number for systems that had more than one version.

Another challenge for the field is posed by the spread in technological sophistication. Our analyses confirmed what has been hinted at in previous reviews and meta-analyses [3, 7, 8]: systems developed within a research context vary in their implementation and

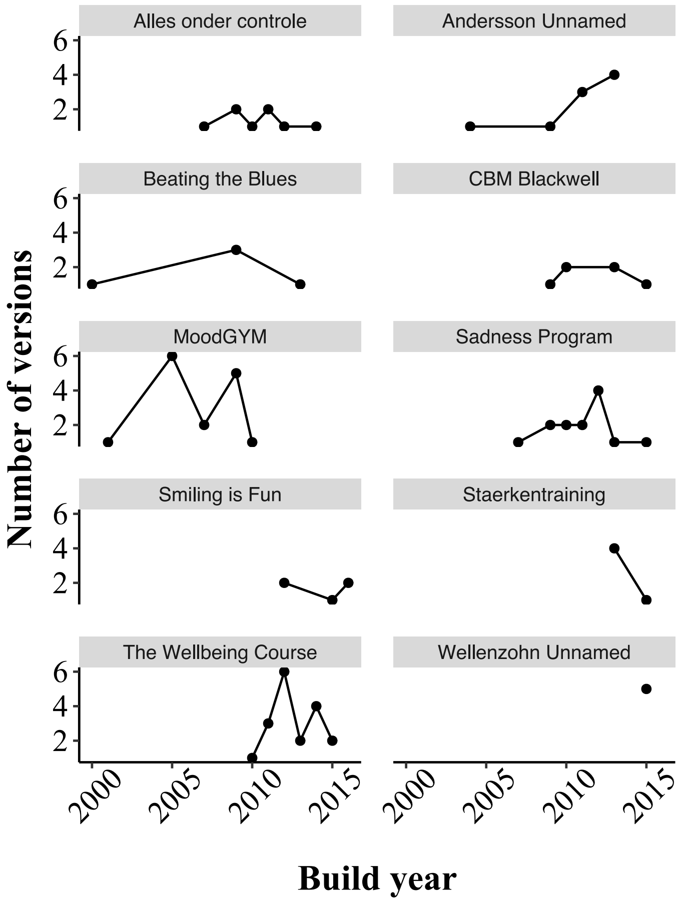

Figure 3.7: The number of versions developed per year between 2000 and 2016 for the ten systems having five or more versions.

in their technological sophistication. By and large, they are not very technologically advanced, and those systems that are mostly informational in nature account for 81 % of what is evaluated in RCTs. Only approximately one-fifth of the systems have a substantial amount of functions that are responsive to input from the user. These differences in technological realization have, thus far, been neglected in literature syntheses taking a clinical psychology perspective. For example, two effects identified in such syntheses are that both adherence and effect size appear to increase with higher levels of human guidance (no guidance vs administrative guidance vs therapist guidance) [4]. Although this has been hypothesized to be linked to missing therapeutic alliance or accountability, our results indicate another possibility. We found the lack of guidance to be neither compensated with more content or technological support nor with a more responsive and, thus potentially more engaging, system. It is therefore possible that guidance plays a role especially when systems are not very responsive. As, according to our analyses, this applies to approximately 80 % of the systems, the results of meta-analyses over all systems may not generalize to more technologically advanced solutions. This notion finds some support in a system-specific meta-analysis of the Deprexis system [50], which ranked second in our ranking of the most technologically advanced systems. Across different studies with Deprexis, dropout ranged from 6 % to 50 %, contrasting with the average dropout rate of 74 % found for other unguided systems in general [4]. Furthermore, it was not only observed that unguided Deprexis had an average effect size across trials comparable with that of other systems, including administrative guidance [4], but also that adding guidance did not influence the magnitude of the effect. However, it must also be emphasized at this point that the potential of more technologically advanced systems leading to higher adherence is merely a hypothesis that is in need of further investigation.

Aside from these challenges, we also see opportunities. Systems developed for depression, to date, are hardly making use of the full bandwidth of available technology. In fact, empirically evaluated systems are mostly delivered on the World Wide Web. Only a very few take a mobile form as either native apps or cross-platform Web applications. This is surprising considering that smartphones became a ubiquitous and highly used technology approximately mid-way of the examined time period. In a review from 2015 on the state of the app marketplace for depression apps, 82 apps had been identified for the treatment of depression [51]. A later review (2017) found that only five apps for depression treatment had been empirically evaluated in effectiveness trials [52]. Therefore, an abundance of apps exists, but most apps are commercial and few have been scientifically studied. However, the empirically evaluated apps included in this review fared well in technological sophistication, such as Mobilyze! [41] and Mobile Sensing and Support (MOSS) [30]. Both apps attempt to learn how to provide context-sensitive interventions on the basis of phone sensor readings. The former uses models trained before delivering the interventions, whereas the latter continuously learns user preferences as it intervenes. In addition to mobile apps, there are several other underexplored innovative technologies, such as social media, conversational agents, and virtual reality. Yet, where these were implemented, some technologically interesting solutions emerged. In social media systems, Panoply [36] can be considered a technological forerunner. It integrates social networking between Panoply users and crowdsourcing from Amazon Mechanical Turk to ensure high-quality content, both in terms of users' thought-recording posts and in terms of responses to these posts. Woebot [53], a fully autonomous chatbot provided on social media, was developed after our search; therefore, it was not included in the analyses. Through short, daily conversations using Facebook instant messenger, Woebot continuously checks in with users and tailors short intervening information and empathic replies to their reported mood. Finally, a creative attempt to alleviate depression is presented by the only virtual reality system that we found [54]. Users are first asked to comfort a virtual avatar with the embodiment of a child. They then take on the perspective of this child in virtual reality to hear their own comforting words said back to them, with the effect of increasing their self-compassion. However, innovative technology solutions, such as the ones mentioned, are scarce. Thus, there still are many opportunities for the field to explore such directions.

3.4.2. CONCLUSIONS

The e-mental health field for the treatment and prevention of depressive disorders is large and consist of a very active research community, as evidenced by the vast body of literature that could be identified for this study. In line with our research questions, three main conclusions can be drawn. First, although the system landscape is overall varied, there are clear trends: three quarters of the systems implement therapeutic techniques related to CBT, three quarters are delivered on the World Wide Web, and three quarters have been evaluated in comparative trials. Second, most systems do not get close to the full technological potential of e-mental health. However, some do. On the level of functions, we have further found that nearly all functions have been implemented in a responsive manner in at least one system, showing that the high end of the scale is obtainable across the board. Third, there appears to be no clear technological development across systems between 2000 and 2017. Furthermore, within systems that have multiple versions, a small increase in size with each new version showed, but this was not the case for technological sophistication. Consequently, it can be argued that, from a technological perspective, there is still room for improvement. Future research investigating the relationship between software implementation and clinical outcomes will need to show whether such improvement is beneficial and cost-efficient with regard to development and maintenance.

To conclude, the scientific contribution of this research is its provision of a comprehensive overview of the technological state of the art of e-mental health systems for the prevention and treatment of adult major depressive disorder developed and studied since the year 2000. This is further accompanied by EHealth4MDD, an open-access database containing all extracted and coded information from the literature used in this writing. Together, the review and database are intended to serve as inspiration for the development of new systems on the one hand and as facilitators for the study of hypotheses related to system composition on the other.

Bibliography

1. Burger, F., Neerincx, M. A. & Brinkman, W.-P. Technological State of the Art of Electronic Mental Health Interventions for Major Depressive Disorder: Systematic Literature Review. *Journal of Medical Internet Research* **22,** e12599 (2020).
2. World Federation For Mental Health (WFMH). Depression: A Global Crisis. *World Mental Health Day* (2012).
3. Schröder, J., Berger, T., Westermann, S., Klein, J. P. & Moritz, S. Internet interventions for depression: New developments. *Dialogues in Clinical Neuroscience* **18,** 203–212 (2016).
4. Richards, D. & Richardson, T. Computer-based psychological treatments for depression: A systematic review and meta-analysis. *Clinical Psychology Review* **32,** 329–342 (2012).
5. Foroushani, P. S., Schneider, J. & Assareh, N. Meta-review of the effectiveness of computerised CBT in treating depression. *BMC Psychiatry* **11** (2011).
6. Karyotaki, E., Riper, H., Twisk, J., Hoogendoorn, A., Kleiboer, A., Mira, A., MacKinnon, A., Meyer, B., Botella, C., Littlewood, E., Andersson, G., Christensen, H., Klein, J. P., Schröder, J., Bretón-ópez, J., Scheider, J., Griffiths, K., Farrer, L., Huibers, M. J., Phillips, R., Gilbody, S., Moritz, S., Berger, T., Pop, V., Spek, V. & Cuijpers, P. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms a meta-analysis of individual participant data. *JAMA Psychiatry* **74,** 351–359 (2017).
7. Wildeboer, G., Kelders, S. M. & van Gemert-Pijnen, J. E. The relationship between persuasive technology principles, adherence and effect of web-Based interventions for mental health: A meta-analysis. *International Journal of Medical Informatics* **96,** 71–85 (2016).
8. Kelders, S. M., Kok, R. N., Ossebaard, H. C. & Gemert-Pijnen, J. E. V. Persuasive system design does matter: A systematic review of adherence to web-based interventions. *Journal of Medical Internet Research* **14** (2012).
9. Titov, N. Internet-delivered psychotherapy for depression in adults. *Current Opinion in Psychiatry* **24,** 18–23 (2011).
10. Newman, M. G., Szkodny, L. E., Llera, S. J. & Przeworski, A. A review of technology-assisted self-help and minimal contact therapies for anxiety and depression: Is human contact necessary for therapeutic efficacy? *Clinical Psychology Review* **31,** 89–103 (2011).
11. Barak, A., Hen, L., Boniel-Nissim, M. & Shapira, N. A comprehensive review and a meta-analysis of the effectiveness of internet-based psychotherapeutic interventions. *Journal of Technology in Human Services* **26,** 109–160 (2008).

3

12. Burger, F., Neerincx, M. A. & Brinkman, W. P. Ehealth4MDD: A database of e-health systems for the prevention and treatment of depressive disorders. *Annual Review of CyberTherapy and Telemedicine* **16,** 18–24 (2018).
13. Burger, F. *Website of the eHealth4MDD Database* `http://insyprojects.ewi.tudelft.nl:8888/`. (2018).
14. Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., Altman, D., Antes, G., Atkins, D., Barbour, V., Barrowman, N., Berlin, J. A., Clark, J., Clarke, M., Cook, D., D'Amico, R., Deeks, J. J., Devereaux, P. J., Dickersin, K., Egger, M., Ernst, E., Gøtzsche, P. C., Grimshaw, J., Guyatt, G., Higgins, J., Ioannidis, J. P., Kleijnen, J., Lang, T., Magrini, N., McNamee, D., Moja, L., Mulrow, C., Napoli, M., Oxman, A., Pham, B., Rennie, D., Sampson, M., Schulz, K. F., Shekelle, P. G., Tovey, D. & Tugwell, P. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine* **6** (2009).
15. Microsoft Patterns & Practices Team. *Microsoft Application Architecture Guide, 2nd Edition* (Microsoft Press, 2009).
16. Oinas-Kukkonen, H. & Harjumaa, M. Persuasive systems design: Key issues, process model, and system features. *Communications of the Association for Information Systems* **24,** 485–500 (2009).
17. Morrison, L. G., Yardley, L., Powell, J. & Michie, S. What design features are used in effective e-health interventions? A review using techniques from critical interpretive synthesis. *Telemedicine and e-Health* **18,** 137–144 (2012).
18. Fogg, B. J. *Persuasive Technology: Using Computers to Change What We Think and Do* (Elsevier Inc., 2003).
19. Wendel, S. *Designing for Behavior Change* (O'Reilly, 2014).
20. Faßbinder, E., Klein, J. P., Sipos, V. & Schweiger, U. *Therapie Tools Depression* (Beltz Verlag, 2015).
21. Grant, M. J., Button, C. M. & Snook, B. An Evaluation of Interrater Reliability Measures on Binary Tasks Using d-Prime. *Applied Psychological Measurement* **41,** 264–276 (2017).
22. Lemma, A. & Fonagy, P. Feasibility study of a psychodynamic online group intervention for depression. *Psychoanalytic Psychology* **30,** 367–380 (2013).
23. Currie, S. L., McGrath, P. J. & Day, V. Development and usability of an online CBT program for symptoms of moderate depression, anxiety, and stress in post-secondary students. *Computers in Human Behavior* **26,** 1419–1426 (2010).
24. Burger, F. *Analysis of technological state of the art in e-mental health for the prevention and treatment of depression.* 4TU.ResearchData `https://doi.org/10.4121/uuid:623b37bc-3ea7-48e3-ab09-575aa3b1b9ac`. (2019).
25. Spitzer, R. L., Kroenke, K. & Williams, J. B. Validation and utility of a self-report version of PRIME-MD: The PHQ Primary Care Study. *Journal of the American Medical Association* **282,** 1737–1744 (1999).

26. Beck, A. T., Ward, C. H., Mendelson, M., Mock, J. & Erbaugh, J. An Inventory for Measuring Depression. *Archives of General Psychiatry* **4,** 561–571 (1961).

27. Radloff, L. S. The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Applied Psychological Measurement* **1,** 385–401 (1977).

28. Burton, C., Tatar, A. S., McKinstry, B., Matheson, C., Matu, S., Moldovan, R., Macnab, M., Farrow, E., David, D., Pagliari, C., Blanco, A. S. & Wolters, M. Pilot randomised controlled trial of Help4Mood, an embodied virtual agent-based system to support treatment of depression. *Journal of Telemedicine and Telecare* **22,** 348–355 (2016).

29. Meyer, B., Berger, T., Caspar, F., Beevers, C. G., Andersson, G. & Weiss, M. Effectiveness of a novel integrative online treatment for depression (Deprexis): Randomized controlled trial. *Journal of Medical Internet Research* **11** (2009).

30. Wahle, F., Kowatsch, T., Fleisch, E., Rufer, M. & Weidt, S. Mobile sensing and support for people with depression: A pilot trial in the wild. *JMIR mHealth and uHealth* **4** (2016).

31. Buhrman, M., Syk, M., Burvall, O., Hartig, T., Gordh, T. & Andersson, G. Individualized guided internet-delivered cognitive-behavior therapy for chronic pain patients with comorbid depression and anxiety: A randomized controlled trial. *Clinical Journal of Pain* **31,** 504–516 (2015).

32. Ahmedani, B. K., Crotty, N., Abdulhak, M. M. & Ondersma, S. J. Pilot feasibility study of a brief, tailored mobile health intervention for depression among patients with chronic pain. *Behavioral Medicine* **41,** 25–32 (2015).

33. Spates, C. R., Kalata, A. H., Ozeki, S., Stanton, C. E. & Peters, S. Initial Open Trial of a Computerized Behavioral Activation Treatment for Depression. *Behavior Modification* **37,** 259–297 (2013).

34. Wu, S., Vidyanti, I., Liu, P., Hawkins, C., Ramirez, M., Guterman, J., Gross-Schulman, S., Sklaroff, L. M. & Ell, K. Patient-centered technological assessment and monitoring of depression for low-income patients. *Journal of Ambulatory Care Management* **37,** 138–147 (2014).

35. Shamekhi, A., Bickmore, T., Lestoquoy, A. & Gardiner, P. Augmenting Group Medical Visits with Conversational Agents for Stress Management Behavior Change. in *Persuasive Technology: Development and Implementation of Personalized Technologies to Change Attitudes and Behaviors* (eds de Vries, P. W., Oinas-Kukkonen, H., Siemons, L., Jong, N. B.-d. & van Gemert-Pijnen, L.) **10171** (Springer International Publishing, 2017), 55–67.

36. Morris, R. R., Schueller, S. M. & Picard, R. W. Efficacy of a web-based, crowdsourced peer-to-peer cognitive reappraisal platform for depression: Randomized controlled trial. *Journal of Medical Internet Research* **17** (2015).

37. Doherty, G., Coyle, D. & Sharry, J. Engagement with online mental health interventions: An exploratory clinical study of a treatment for depression. in *CHI '12: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Association for Computing Machinery, 2012), 1421–1430.

3

38. Rebar, A. L., Boles, C., Burton, N., Duncan, M. J., Short, C. E., Happell, B., Kolt, G. S., Caperchione, C. M., Rosenkranz, R. R. & Vandelanotte, C. Healthy mind, healthy body: A randomized trial testing the efficacy of a computer-tailored vs. interactive web-based intervention for increasing physical activity and reducing depressive symptoms. *Mental Health and Physical Activity* **11,** 29–37 (2016).

39. Richards, D., Timulak, L., O'Brien, E., Hayes, C., Vigano, N., Sharry, J. & Doherty, G. A randomized controlled trial of an internet-delivered treatment: Its potential as a low-intensity community intervention for adults with symptoms of depression. *Behaviour Research and Therapy* **75,** 20–31 (2015).

40. Anguera, J. A., Gunning, F. M. & Areán, P. A. Improving late life depression and cognitive control through the use of therapeutic video game technology: A proof-of-concept randomized trial. *Depression and Anxiety* **34,** 508–517 (2017).

41. Burns, M. N., Begale, M., Duffecy, J., Gergle, D., Karr, C. J., Giangrande, E. & Mohr, D. C. Harnessing context sensing to develop a mobile intervention for depression. *Journal of Medical Internet Research* **13** (2011).

42. Nørregaard, L. B., Løventoft, P. K., Lauritsen, L., Olsson, E. C., Andersen, L., Rauff, S., Frøkjær, E. & Martiny, K. Patient expectations and experiences from a clinical study in psychiatric care using a self-monitoring system. in *NordiCHI '14: Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational* (Association for Computing Machinery, New York, NY, USA, 2014), 991–994.

43. Riva, G., Banos, R., Botella, C., Gaggioli, A. & Wiederhold, B. K. Personal health systems for mental health: The European projects. *Studies in Health Technology and Informatics* **163,** 496–502 (2011).

44. Glenn, T. & Monteith, S. New Measures of Mental State and Behavior Based on Data Collected From Sensors, Smartphones, and the Internet. *Current Psychiatry Reports* **16** (2014).

45. Liu, Z., Hu, B., Yan, L., Wang, T., Liu, F., Li, X. & Kang, H. Detection of depression in speech. in *2015 International Conference on Affective Computing and Intelligent Interaction, ACII 2015* (Institute of Electrical and Electronics Engineers Inc., 2015), 743–747.

46. Gwet, K. L. *Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring the Extent of Agreement Among Raters* (Advanced Analytics, LLC, 2010).

47. Hunkeler, E. M., Hargreaves, W. A., Fireman, B., Terdiman, J., Meresman, J. F., Porterfield, Y., Lee, J., Dea, R., Simon, G. E., Bauer, M. S., Unützer, J. & Taylor, C. B. A web-delivered care management and patient self-management program for recurrent depression: A randomized trial. *Psychiatric Services* **63,** 1063–1071 (2012).

48. Imamura, K., Kawakami, N., Furukawa, T. A., Matsuyama, Y., Shimazu, A., Umanodan, R., Kawakami, S. & Kasai, K. Effects of an internet-based cognitive behavioral therapy (iCBT) program in manga format on improving subthreshold depressive symptoms among healthy workers: A randomized controlled trial. *PLoS ONE* **9** (2014).

49. Ruwaard, J., Schrieken, B., Schrijver, M., Broeksteeg, J., Dekker, J., Vermeulen, H. & Lange, A. Standardized web-based cognitive behavioural therapy of mild to moderate depression: A randomized controlled trial with a long-term follow-up. *Cognitive Behaviour Therapy* **38,** 206–221 (2009).

50. Twomey, C., O'Reilly, G. & Meyer, B. Effectiveness of an individually-tailored computerised CBT programme (Deprexis) for depression: A meta-analysis. *Psychiatry Research* **256,** 371–377 (2017).

51. Shen, N., Levitan, M. J., Johnson, A., Bender, J. L., Hamilton-Page, M., Jadad, A. R. & Wiljer, D. Finding a depression app: A review and content analysis of the depression app marketplace. *JMIR mHealth and uHealth* **3** (2015).

52. van Ameringen, M., Turna, J., Khalesi, Z., Pullia, K. & Patterson, B. There is an app for that! The current state of mobile applications (apps) for DSM-5 obsessive-compulsive disorder, posttraumatic stress disorder, anxiety and mood disorders. *Depression and Anxiety* **34,** 526–539 (2017).

53. Fitzpatrick, K. K., Darcy, A. & Vierhile, M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. *JMIR Mental Health* **4** (2017).

54. Falconer, C. J., Rovira, A., King, J. A., Gilbert, P., Antley, A., Fearon, P., Ralph, N., Slater, M. & Brewin, C. R. Embodying self-compassion within virtual reality and its effects on patients with depression. *BJPsych Open* **2,** 74–80 (2016).

4

NATURAL LANGUAGE PROCESSING FOR COGNITIVE THERAPY: EXTRACTING SCHEMAS FROM THOUGHT RECORDS

This chapter is based on Burger, F., Neerincx, M. A. & Brinkman, W. P. Natural language processing for cognitive therapy: Extracting schemas from thought records. *PLoS ONE* **16** (2021)

ABSTRACT

The cognitive approach to psychotherapy aims to change patients' maladaptive schemas, that is, overly negative views on themselves, the world, or the future. To obtain awareness of these views, they record their thought processes in situations that caused pathogenic emotional responses. The schemas underlying such thought records have, thus far, been largely manually identified. Using recent advances in natural language processing, we take this one step further by automatically extracting schemas from thought records. To this end, we asked 320 healthy participants on Amazon Mechanical Turk to each complete five thought records consisting of several utterances reflecting cognitive processes. Agreement between two raters on manually scoring the utterances with respect to how much they reflect each schema was substantial (Cohen's $\kappa = 0.79$*). Natural language processing software pretrained on all English Wikipedia articles from 2014 (GLoVE embeddings) was used to represent words and utterances, which were then mapped to schemas using k-nearest neighbors algorithms, support vector machines, and recurrent neural networks. For the more frequently occurring schemas, all algorithms were able to leverage linguistic patterns. For example, the scores assigned to the* Competence *schema by the algorithms correlated with the manually assigned scores with Spearman correlations ranging between 0.64 and 0.76. For six of the nine schemas, a set of recurrent neural networks trained separately for each of the schemas outperformed the other algorithms. We present our results here as a benchmark solution, since we conducted this research to explore the possibility of automatically processing qualitative mental health data and did not aim to achieve optimal performance with any of the explored models. The dataset of 1600 thought records comprising 5747 utterances is published together with this article for researchers and machine learning enthusiasts to improve upon our outcomes. Based on our promising results, we see further opportunities for using free-text input and subsequent natural language processing in other common therapeutic tools, such as ecological momentary assessments, automated case conceptualizations, and, more generally, as an alternative to mental health scales.*

4.1. INTRODUCTION

E-mental health—delivering therapeutic interventions via information and communication technology—is regarded as a promising means of overcoming many barriers to traditional psychotherapeutic care. Yet, in a review of more than 130 scientifically evaluated e-mental health systems for depression, it was found that the technological state of the art of these systems is limited: even in recently developed systems, technology is often only used as a platform for delivering information to the patient. When the patient is asked to provide open, unconstrained textual information to the system, this information is typically either processed by a human in the case of guided systems or not processed at all in the case of autonomous systems [2]. Although both methods are arguably very robust to misunderstanding, human processing is costly while no processing offers no advantage over traditional paper-based workbooks. However, developments in data-driven natural language understanding are increasingly able to reliably interpret unconstrained qualitative user input. Here, we explore this opportunity for a specific therapeutic task in cognitive therapy: determining underlying maladaptive schemas from the information contained in thought record forms.

Thought record forms provide patients with a structured format for monitoring their thoughts, consisting of descriptions of the thought eliciting situation, the experienced emotion, the first cognitive appraisal of the situation, and the resulting behavior. Thought records are commonly employed in cognitive therapy, a form of psychotherapy based on Beck's cognitive theory [3]. The theory posits that not the situations but the way in which we appraise them causes our emotions. For example, it is not the fact that we are not invited to a party that makes us upset but rather the fear or understanding that this says something about us or our relationship with the host. Our immediate and unreflected appraisal of a situation is called an automatic thought. Automatic thoughts are in turn determined by schemas, the cognitive structures that make up our world view. A specific schema can be activated given the right trigger. In people with certain mental illnesses, it is theorized that pathogenic schemas have a particularly low activation threshold [4]. Consequently, a core part of cognitive psychotherapy involves teaching patients to monitor thoughts for insight into underlying schemas. Starting from the automatic thought noted down in the thought record, the downward arrow technique (DAT) [5] helps to determine the causative maladaptive schema. It consists of repeatedly asking *why it would be upsetting* or *what would be the worst that could happen* if the idea stated in the previous step was true. An example thought record that we collected in our experiment is shown in Table 4.1. The DAT is illustrated by the final three rows (in cursive font). Since the majority of thought records in our dataset include the DAT, hereinafter the term *thought record* refers to both the core thought record and DAT unless explicitly stated otherwise. Also extending beyond the nomenclature typically used in clinical psychology, we define as a thought record *utterance* the automatic thought or any completed step of the DAT. Each of the final four rows of the Participant Response column of Table 4.1 reflects an utterance. As can be seen in the response to the second downward arrow step, i.e., *I want friends. I will be lonely otherwise.*, an utterance can consist of multiple sentences.

Unlike automatic thoughts, schemas have received little attention in empirical research to date [6]. When considered, they have typically been explored in a top-down manner with measurement instruments developed on the basis of cognitive theory and validated with exploratory factor analyses (for example, [6, 7]). To the best of our knowledge, only one classification rubric for schemas exists that was not exclusively derived from theory but created from a content analysis of a set of thought records (also including DAT) collected with an online self-help cognitive behavioral therapy (CBT) program, namely the schema rubric of Millings *et al.* [8].

In this work, we develop the natural language processing (NLP) foundation for a task-oriented conversational agent (CA) that motivates users to regularly complete thought recording homework exercises. Most CAs used in practice to date are frame-based [9, Ch. 24]. To be able to parse the semantics of a user input (e.g. "I want to take my girlfriend to the theater next weekend.") and fill the slots in a frame (e.g. day, show, theater, time, number of tickets), the agent needs to classify broadly the intent of the entire input phrase (e.g. book theater tickets) and extract specifically the information corresponding to empty slots. When all slots are filled, the agent can complete the task. Up until recently, intent classification and slot filling were mostly done using a hand-written, domain-specific semantic grammar, often prescribing possible synonyms as well as a certain order for the information (e.g. {I want | Could I | It would be great if I could} * {book | reserve | get} * {tickets | cards} * {movies | theater} *). Systems using such grammars are expensive in terms of engineering time and prone to errors and misunderstandings [9, Ch. 24]. Both drawbacks have been largely eliminated with the advent of deep learning in the past decade. Rather than hand-crafting large sets of rules, deep learning allows for the acquisition of synonyms and word usage in context from large sets of data, such as Wikipedia. As two recent literature reviews show, these developments are slowly finding their way into CAs for health care [10, 11]. Laranjo *et al.* [10] found most of the CAs allowing for unconstrained natural language input to have been developed after 2010. Yet, only one-half of the reviewed agents used frame-based or agent-based dialog management [12], while the other half implemented entirely system-driven and finite-state dialog management strategies. The authors therefore conclude that CAs in health care are not up to par with those in other fields. Of all 40 agents considered by Montenegro *et al.* [11], only six use state-of-the-art natural language understanding techniques [13–18].

While it is always important to limit user frustrations that arise from understanding errors on the part of the CA, this is particularly crucial in dialog systems for mental health treatment due to the highly emotionally sensitive domain. It is conceivable that language understanding errors as well as inconsistent or insensitive [19] responses could affect not only patients' experience and trust in the system but, in the worst case, also their mental health. Consequently, rule-based systems have been the norm [20]. Questions from the system are phrased so narrowly that they leave little room for unexpected responses (e.g., [21]). Since even therapy following a strict protocol is much less task-oriented than booking theater tickets, most systems fully or partially resort to providing multiple response options to the user (see, for example, [22]). The more recently developed Woebot [23], a chatbot for treating college students with symptoms of anxiety and depression, only uses natural language processing as an option for some nodes of Woebot's decision tree architecture, choosing the next node mostly based on user selection of one of several suggested replies.

Table 4.1: Example of one complete thought record from the dataset collected in this study.

TR Question	**Entry Type**	**Participant Response**
Describe the situation very briefly in your own words.	open text entry field	while walking down the street I see someone I know, wave at them and they don't acknowledge my wave.
How well can you imagine yourself in this situation?	slider from 0 (not at all) to 100 (as good as if you were in the situation at the very moment)	85
Describe your emotion in this situation in one word.	open text entry field	disappointment
How intensely would you be experiencing this emotion?	slider from 0 (a trace) to 100 (the most intense possible)	45
Which of the following four emotions corresponds best with the emotion that you wrote down above?	multiple choice: sadness, fear, anger, happiness	sadness
Which (automatic) thought might have caused you to feel this way in the described situation?	open text entry field	They don't like me enough to wave back
And why would it be upsetting to you if "They don't like me enough to wave back" were true? What would it mean to you? What would it say about you?	*open text entry field*	*I may be unlikeable.*
And why would it be upsetting to you if "I may be unlikeable" were true? What would it mean to you? What would it say about you?	*open text entry field*	*I want friends. I will be lonely otherwise.*
And why would it be upsetting to you if "I want friends so I won't be lonely." were true? What would it mean to you? What would it say about you?	*open text entry field*	*If I am unlikeable then I won't have friends and will be alone all my life.*
What would you do in the situation, if anything?	open text entry field	I would try to make better impressions on people I meet.

Steps of the downward arrow technique are presented in cursive font. Three downward arrow steps were completed in this thought record. After each downward arrow step (question + open text entry field), participants were asked the intermediate closed question of whether they wanted to continue with the downward arrow technique or not. Thus, after each step participants indicated that they wanted to continue until the final one, where they indicated that they wanted to stop, thereby completing the downward arrow technique. The intermediate question is omitted here. Following the downward arrow technique, participants completed the entire thought record by describing their behavior in the situation. The scenario description presented to the participant was "You are walking down the street. On the other side of the street you see an acquaintance whom you've liked the few times you've been in his company. You wave to him, and you get no response."

Thought recording exercises are often assigned as homework to patients in face-to-face treatment or included in self-help workbooks and treatment systems with only general instructions. Timely feedback or tailored support from a therapist are therefore usually not available when patients attempt the exercise. As the goal of thought recording is the discovery of thinking patterns, frequent completion of thought records is crucial for their success. It is for these reasons that we aim to build a CA to motivate and support people in regularly completing thought records. The CA can use knowledge about schemas to provide feedback, to respond understandingly, or to strategically ask for supplementary information. This work therefore addresses the following primary research question: Can the underlying maladaptive schema of a thought record utterance be scored by a machine?

4.2. HYPOTHESES

The objective of this study was to see whether identifying schemas from thought records is at all possible. Consequently, our first hypothesis is that schemas can be extracted automatically (H1). We investigate this with a future goal of implementing a conversational agent capable of providing useful feedback. For such practical applications, we were also interested in studying ways to potentially improve automatic schema identification. As a result, three additional hypotheses, informed by psychological theory, were also investigated: automatic predictions improve as the downward arrow technique progresses (H2), within individuals, similar situations will activate the same schemas (H3), across individuals, there is a relationship between the active schemas and scores on mental health scales (H4). We here motivate the hypotheses in turn.

4.2.1. H1: SCHEMAS CAN BE EXTRACTED AUTOMATICALLY

As outlined above, conversational agents in health care, and particularly in depression treatment, to date are employing grammar-based or no NLP more often than not. Yet, the field more generally has not been blind to state-of-the-art data-driven methods. Thus far, however, they are mostly used in clinical psychology research to perform psychological assessment. Social media platforms and forums provide a treasure trove of natural language data occurring in virtual social environments. This has resulted in a large body of literature searching for linguistic markers indicative of depression, crisis, or suicidal risk in the data (e.g. [24–30]). One such example is the crisis detection models developed in [28]. With a dataset of posts comprising on average three sentences collected through the mental health app *Koko*, the authors use a recurrent neural network (RNN) to detect crisis (binary classification task). They augment their RNN with *attention* [31] to display the parts of a post that the neural network pays attention to during classification. Their best model, an RNN without attention, detects crisis with an F1-score accuracy of 0.80. In another study [29], the task was to correctly identify which topic-based forum (or *subreddit*) on the social media website Reddit the posts of users belong to. The posts were drawn from eleven different manually selected mental health subreddits. The best performing algorithm achieved an F1-score accuracy of 0.71 with a convolutional neural network in this multi-class (more than two classes that are mutually exclusive) classification task. Benton *et al.* [30] study a similar problem as a multi-label (more than two classes that are not mutually exclusive) learning task. Using tweets posted on the social media platform Twitter, they simultaneously classify suicidal risk, atypical mental health, and seven mental health conditions. They observed a clear added benefit of leveraging possible correlations between the labels in the multi-label models compared to a set of nine single-class prediction models. Although the described research indicates that automatically identifying crisis or mental health conditions from social media corpora is feasible, it is unknown whether this applies to schemas as well. However, the fact that the schema rubric of Millings *et al.* [8] was obtained via content analysis from a corpus of thought records indicates that language and word usage differ between the schemas. If this is the case, a good model trained on sufficient data should be able to pick up on these differences. Additionally, schemas are not mutually exclusive and might therefore inform each other, possibly further improving prediction accuracy. On the basis of these considerations, we posit the following:

H1 The schema(s) underlying a thought record can be identified by an algorithm with an accuracy above chance.

4.2.2. H2: Downward arrow converges and H3: schema patterns are similar across thought record type

Thought records ask patients to first briefly describe the situation that resulted in the pathogenic emotion in their own words. The automatic thought is thus directly connected to the situation description and both are highly individual. Automatically analyzing such free-form open text without any further restrictions is an *open-domain* NLP task, similar to small-talk. For an artificial intelligence, this is notoriously difficult to deal with well as it requires a comprehensive world model of many topics. Such a model cannot feasibly be engineered by humans and, if it is at all possible, very large amounts of data would be required to construct it bottom-up. Models created in this manner are usually no longer transparent and may show unintended behavior (e.g., [32]).

From a clinical perspective, an alternative to open thought recording is to elicit schemas by means of imagined situations, using scripted situation vignettes as a basis for the thought records. Thought recording is typically assigned as homework for the patient in cognitive therapy, with the completed forms constituting an integral part of the face-to-face sessions. While leaving patients to their own devices provides them with freedom and ensures ecological validity, the various different steps of the thought recording method do not always come easy to patients [33]. When they struggle, therapists may guide the process by resorting to imagery or role-play so as to recreate the situation in the face-to-face session and evoke the automatic thought again [34]. For initial practice [35] or for the controlled assessment of cognitive errors [36–38] and cognitive restructuring skills [39], therapists may additionally restrict patients by asking them to envision themselves in certain scripted ambiguous scenarios. From a technical perspective, such a scripted scenario can delimit the natural language domain. Taking the scenario into account in a schema identification model should thus produce more reliable results. Despite scenarios being viable from a clinical perspective and the safer option from a technical perspective, two aspects of cognitive therapy give rise to the possibility of open classification models for this specific NLP task: the *downward arrow technique* and the categorization of situations into *situation types*.

Downward arrow technique

The theory behind the downward arrow technique (DAT) posits that as one progresses along the downward arrow, a schema will be reached. While automatic thoughts are specific appraisals of situations, schemas are general: the same schema can cause a large variety of specific automatic thoughts. From this, it should follow that the thoughts delineated with the DAT become increasingly independent of the situation description. For the NLP, this means that the language in utterances should converge to language that is more characteristic of the schema. We therefore hypothesize as follows:

H2 Schema identification accuracy increases as one proceeds along the downward arrow.

Categorization of situations

Two situation types that are commonly distinguished in cognitive therapy are interpersonal situations and achievement-related situations (e.g., [40]). *Interpersonal* situations pertain to one's self-worth in relation to other people, while *achievement-related* situations are such where one might perform poorly and one's self-esteem is at risk. Hence, a schema identification model might generalize to any real-world situation as long as it takes into account whether the situation type is more interpersonal or more achievement-related. Consequently, the following hypothesis is tested:

H3 Within an individual, the schema patterns of scenario-based thought records can predict those of the real-life thought record when they match in situation type (interpersonal or achievement-related).

4.2.3. H4: Mental illnesses have associated schemas

Lastly, cognitive theory argues for differences between depression and anxiety with regard to schemas. Depressed individuals are theorized to have overly negative views of the self, the world, and the future, while anxious individuals hold schemas related to personal danger [41]. However, Millings *et al.* [8] found that only the presence of the schema related to power and being in control differs between those with depression and those with anxiety, with particularly the anxious participants in their online CBT program presenting with the schema. If each mental illness were to show specific associated schemas, though, mental health data could inform a prior distribution over schemas in terms of their likelihood. This might improve a machine learning model. Using the coding scheme of [8], we therefore pose the following exploratory hypothesis:

H4 The schema patterns of an individual combined across thought records can predict his or her depression, anxiety, and cognitive distortions as self-reported using standard psychological questionnaires.

4.3. Methods

To test the hypotheses stated above, a dataset of completed thought records was needed. Copies of thought records from actual patients gathered through a therapeutic practice were not an option because we could not obtain access to such an existing corpus. We therefore chose to collect a new dataset of thought records through the online crowdsourcing platform Amazon Mechanical Turk. The Human Research Ethics Committee of Delft University of Technology granted ethical approval for the research (Letter of Approval number: 546).

4.3.1. DESIGN

The data collection process was designed as a cross-sectional observational study. This means that there were no independent variables manipulated and consequently no conditions.

4.3.2. MATERIALS

Three online platforms were used in the study: Amazon Mechanical Turk (MTurk) for recruitment, Qualtrics for data collection, and YouTube for hosting instructional videos on how to complete thought records. People who registered for the task on MTurk were redirected to Qualtrics. YouTube videos were embedded in Qualtrics.

The instructions for the thought recording task included psychoeducation on cognitive theory, a short description of the components of a thought record, and four video examples of how to complete the thought records using two scenarios and four fictional characters to emphasize that thought records are highly individual and that there are no incorrect answers as long as thought records are coherent.

Two types of thought records were used in the study: *closed* and *open* thought records. The closed thought records asked participants to imagine themselves in a certain prescripted scenario and to write thought records as if what is detailed in the scenario had happened to them. The open thought records, on the other hand, asked participants to write thought records using a recent situation from their own lives. The scenarios of the closed thought records for any participant were chosen from a set of ten possible scenarios. These were divided into two sets of five scenarios, one set comprising scenarios of an interpersonal nature, the other comprising scenarios of an achievement-related nature. The scenarios were taken from the Ways of Responding Questionnaire [39] and the Cognitive Error Questionnaire [37]. A complete list of the scenarios can be found in the data repository of this study [42]. The open thought record followed the exact same structure as the closed ones, except that participants had to briefly describe a situation that happened in their life instead of first imagining themselves in a given scenario and then describing it again in their own words.

The formulation of the downward arrow technique (DAT) questions depended on the emotion category that participants selected. When this was *happiness*, they were not directed to complete the DAT after stating the automatic thought. Therefore, all thought records in our dataset have at least one utterance: the automatic thought. When selecting *sadness* or *anger* the DAT consisted of repeatedly asking “And why would it be upsetting to you if [previously stated thought] were true? What would it mean to you? What does it say about you?" When selecting *fear*, on the other hand, the corresponding question was “And what would be the worst that could happen if [previously stated thought] were true? What would it mean to you? What does it say about you?" Just like the thought records, the DAT was altered slightly to better fit online administration: after each step, participants were asked whether they wanted to continue with the technique or not. This was necessary to eventually break the loop while giving participants the chance to complete as many steps as they wanted.

4.3.3. MEASURES

Three mental health questionnaires were used: the Hospital Anxiety and Depression Scale (HDAS) [43], the Beck Depression Inventory (BDI-IA) [44], and the Cognitive Distortions Scale (CDS) [40]. The HDAS is a diagnostic tool for depression and anxiety, while the BDI-IA only assesses symptoms of depression. The CDS measures to what degree someone suffers from cognitive distortions, such as black-and-white thinking, in achievement-related as well as in interpersonal situations.

The post-questionnaire comprised three items asking participants how difficult and how enjoyable they found it to complete a thought record, and to indicate how many thought records they think they would complete if they were asked to complete a thought record daily for a period of seven days. We collected this data as secondary measures in anticipation of follow-up research, in which we aim to implement a conversational agent to motivate users to regularly record their thoughts. The data from the post-questionnaire were collected for follow-up research and will not be discussed in this paper.

4.3.4. PARTICIPANTS

The only qualifications participants needed to access the task on MTurk was to be located in the USA, Canada, the UK, or Australia, to be at least 18 years of age, and to never have participated in the same study before. A total of 536 participants accepted the task on MTurk. Of these, 320 responses were usable. Hence, approximately 40 % of responses had to be excluded on the basis of participants failing at least one of the two instruction comprehension questions or not taking the task seriously (having filled in incomprehensible text or obviously having copied and pasted text from other websites into the text-entry fields). Excluded participants were not reimbursed. Participants who completed the experiment received $4 for their participation, based on an estimated 35 minutes needed to read the instructions and to complete the task and all questionnaires. This estimate was obtained from a pilot run with 10 participants. In choosing the reimbursement amount, we aimed to fairly reimburse participants' time. As a consequence, the Amazon Mechanical Turk workers, just like patients wishing to get healthier, had an incentive to do the task. However, we did not use the reimbursement to motivate our participants to put in extra effort, as all participants received the same reimbursement.

Of the 320 included participants, 148 were female, 171 were male, and 1 indicated *Other*. The mean age of 319 participants was 36.25 years ($SD = 10.99$) with the youngest being 19 and the oldest 71. Demographic questions were optional and one participant chose not to provide her age.

4.3.5. PROCEDURE

Participants fulfilling the qualification criteria could access the task in MTurk. There, they were presented with basic information about the study, such as a short description of the task and the expected time to complete it. Once having accepted the task, participants were redirect to Qualtrics for the experiment. Upon giving their explicit consent to six statements, they were forwarded to a short demographic pre-questionnaire followed by the task instructions. To ensure that participants would not rush through the instructions, two instruction comprehension questions completed the instructional part: one asking participants what they would have to do in the main task in general and the other concerning procedural aspects of how to complete the thought records as explained in the videos. Failing to answer at least one of the questions correctly resulted in the immediate exclusion of the participant. This was made clear to participants before reaching the questions and the questions were displayed on the same page as the instructions, allowing participants to re-read instructions or re-watch videos before giving their answer. Participants who answered both instruction comprehension questions correctly were forwarded to the thought recording task. This consisted of four closed and one open thought record in this order. For the closed thought records, they were asked to first read the short scenario description and imagine themselves in the situation. They were then directed to a new page with the first thought record form. Throughout the process of completing this, it was possible at any point for the participants to access a short version of the instructions again.

The thought record form was followed by the downward arrow technique. After each step of the DAT, participants were asked whether they wanted to continue with another step. This allowed repeatedly reminding them of the stopping criteria: repeating oneself or feeling that answers were becoming somewhat ridiculous. After indicating that they did not want to continue with the DAT or in case of having selected *happiness* as the emotional response to the situation, participants were presented with the final thought record question. This concerned the behavior they would expect themselves to exhibit in the situation. The post-questionnaire and the three mental health scales completed participation. The entire experimental flow is visualized in Appendix K.

4.3.6. DATA AND ANALYSIS STRATEGIES

To obtain a labeled dataset for training the schema identification models, the thought record utterances had to be scored manually. To this end, we used the schema rubric developed by Millings *et al.* [8]. This rubric comprises ten categories, of which nine are well-defined schemas, such as *Attachment* or *Meta-Cognition*. The final category, however, is an “other" category for all thought records that cannot be assigned one of the well-defined schemas. Schemas are not mutually exclusive, a thought record can therefore be labeled with multiple schemas. We made three modifications to the original rubric. The first modification pertains to the area of application: the original rubric is always applied to an entire thought record, while we apply it to thought record utterances. As a second modification, we dropped the *Other* category, but allowed utterances to have a 0-score for all of the nine schemas labels. As the final modification, we have altered the original rubric from an utterance being indicative of an underlying schema or not (binary schema label) to it being indicative of an underlying schema to a certain degree (ordinal schema score). The schema scores that we assign range from *has absolutely nothing to do with the schema* (0) over *corresponds a little bit with the schema* (1) and *corresponds largely with the schema* (2) to *corresponds completely with the schema* (3).

The schemas of thought record utterances and the scenario type of the open thought record had to be manually scored. Table 4.2 shows example thought record utterances from our dataset for each of the nine schemas and the nine scores assigned to each of the utterances. All manual scoring was conducted by the first author, who scored the utterances in random order. To obtain an indication of reliability, an additional coder, a graduate student of clinical psychology, scored a subset of the utterances. For this, three subsets of 50 randomly selected utterances were used to train the coder until agreement on the interpretation of definitions was reached. Any scoring deviation of more than one point on the ordinal scale was discussed. Then the second coder coded another subset of 100 randomly chosen utterances. Interrater agreement between the first and second coder on this subset was substantial (weighted Cohen's $\kappa = 0.79$). The first coder also recoded the same subset one year after completing the initial coding of all utterances with good intracoder agreement (weighted Cohen's $\kappa = 0.83$).

H1: Schemas can be automatically extracted

To test the first hypothesis, thought record utterances were studied taking a natural language processing perspective: using a machine learning model to score an utterance with regard to the nine well-defined schemas. This task can formally be described as an ordinal multi-label scoring task: an algorithm must assign each utterance a schema vector consisting of nine values ranging between 0 and 3. Assigning ordinal scores to data is generally not trivial and common simplifications are to either treat the ordinal scores as separate classes (nominal data) or as equidistant integers on a continuum (interval data) [45]. The former is otherwise known as classification and entails that the ordering information of scores is lost. The latter is regression and entails that the ordering is maintained, but information is added, such as that labels are equally spaced and that the space between labels can be meaningfully interpreted. Where specific algorithms have been created for ordinal data [45], these often assume that higher ordinal labels subsume lower ones (compare, for example, [46]), e.g., if something corresponds very much to a schema (score 3) it also automatically corresponds a little bit to the schema (score 1). This is not the case here, as we also have score 0 meaning that an utterance does not correspond to a schema. Another criterion for choosing algorithms was the ready availability of functional, well-maintained, and commonly used software packages. We assume this to work to the advantage of reproducibility and further development. As a result of these considerations, we opted to explore both approaches of treating the scores as nominal as well as treating them as interval rather than exploring specific ordinal methods.

Before automatically scoring, the data were linguistically preprocessed by lower-casing, replacing misspellings, contractions, and numbers, adding missing sentence end marks and comma space, and finally removing stop words and unnecessary white space. They were then divided into a training set, a validation set, and a test set, with the test set comprising 15 % of all data, the validation set comprising another 12.75 %, and the training set comprising the remaining 72.25 %. Samples to include in test and validation set were not selected at random but rather we ensured that three criteria were fulfilled: 1. similar distribution of schemas, 2. approximately the same proportion of open and closed scenarios, 3. approximately the same distribution over DAT depths as in the entire dataset. This was achieved by randomly sampling 1000 times from the entire distribution, determining the deviation in distribution between the sample and the population for each of the three criteria, summing these three deviation measures, and choosing the sample with the smallest result. The process was first done for the

Table 4.2: Example utterances for each schema taken from the dataset collected in this study. Utterances are thoughts and can be either automatic thoughts or any thought written as part of the downward arrow technique. Scores were manually assigned for each of the nine mental health schemas by the first author.

Utterance	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**S9**
S1: Attachment examples									
I am unlovable and less than other people. I will never find friends or a girlfriend.	3	0	3	0	0	0	0	1	0
I don't want to be alone.	3	0	0	0	0	0	0	0	0
I was a bad mom.	3	0	0	0	0	0	0	0	0
I failed at the relationship.	3	0	0	0	0	0	0	0	0
I won't be a good partner to others.	3	0	0	0	0	0	0	0	0
S2: Competence examples									
I feel like a failure at my job.	0	3	0	0	0	0	0	0	0
I'm unprepared for this task.	0	3	0	0	0	0	0	0	0
I can never go into a sales job.	0	3	0	0	0	0	0	0	0
I am not good enough to get a job.	0	3	0	0	0	0	0	0	0
I would be unable to produce saleable work.	0	3	0	0	0	0	0	0	0
S3: Global self-evaluation examples									
It would mean that I am lazy and I need to do better	0	0	3	0	0	0	0	0	0
I should never have been born.	0	0	3	0	0	0	0	3	0
I am selfish.	0	0	3	0	0	0	0	0	0
S4: Health examples									
I would become ill.	0	0	0	3	0	0	0	0	0
I feel exhausted and anxious.	0	0	0	2	1	0	0	1	0
I cannot lose weight no matter what I try.	0	0	0	3	2	0	0	0	0
It would be very depressing, it would say that I would need counseling to get through life.	0	0	0	3	1	0	0	1	0
I will have health issues	0	0	0	3	0	0	0	0	0
S5: Power and control examples									
I'm going to be stuck in my current situation forever.	0	0	0	0	3	0	0	1	0
The feeling of being pressured by my boss.	0	0	0	0	3	0	0	0	0
I was fired and not given a chance to succeed.	0	1	0	0	2	0	0	0	0
I am not in control of what I do or how I perceive myself	0	0	0	0	3	2	0	0	0
That I still have a target painted on my back for their abuse.	1	0	0	0	3	0	1	0	0
S6: Meta-Cognition examples									
My perception of people is off and that's why I have a difficulty creating new relationships.	1	0	0	0	0	3	1	0	0
That I can be more than a bit compulsive about investigating odd byways of thought.	0	0	0	0	0	3	0	0	0
I trick myself into believing I'm better than I am.	0	0	0	0	0	3	0	0	0
Because I hold myself to a high standard.	0	0	0	0	0	2	0	0	0
I get angry easily over small things.	0	0	0	0	3	1	0	0	0

Utterance	S1	S2	S3	S4	S5	S6	S7	S8	S9
S7: Other people examples									
People would rather avoid me than be in my presence.	0	0	0	0	0	0	2	0	3
It means that these people not care about anyone but themselves, and i have to suffer	0	0	0	0	0	0	3	0	0
People will mock me	0	0	0	0	0	0	3	0	3
I am not as selfish as other people.	0	0	0	0	0	0	3	0	0
It means that other people can do despicable things and not be accountable.	0	0	0	0	0	0	3	0	0
S8: Hopelessness examples									
I will stop trying in life and give up	0	0	0	0	1	0	0	3	0
I should never have been born.	0	0	3	0	0	0	0	3	0
Depression makes me think I'd be better off dead.	0	0	0	2	0	0	0	3	0
I will never have a life I enjoy	0	0	0	0	0	0	0	3	0
I'll never feel like I have a purpose.	0	0	0	0	0	0	0	3	0
S9: Others views about self examples									
My friends don't like me.	2	0	0	0	0	0	0	0	3
Because I want people like him to like me.	0	0	0	0	0	0	0	0	3
I could not make him see that I am a responsible person.	0	0	0	0	0	0	0	0	3
I must not be his type of person.	0	0	0	0	0	0	0	0	2
It would say that she did not feel like she was able to talk to me.	0	0	0	0	0	0	0	0	3

test set and then repeated with the remaining data samples to obtain the validation set. We used normalized, 100-dimensional GLoVE embeddings [47] trained on all English Wikipedia articles existent in 2014 to represent the words in utterances.

Three types of algorithms of varying levels of complexity were chosen for the task: k nearest neighbors classification (kNN-C) and regression (kNN-R), support vector machine classification (SVC) and regression (SVR), and a multi-label recurrent neural net (RNN) as well as a set of separate RNNs per schema. All three types of algorithms are supervised-learning algorithms, meaning that they learn from labeled examples. The k-nearest neighbors algorithms work as follows: for each new utterance that the algorithm has to label, a distance is calculated between this utterance and each of the utterances of the training set. The distance indicates how similar, i.e. *close* in representation space, the new utterance is to the utterances the algorithm has seen before. In our case, the distance metric was calculated by first linguistically preprocessing each utterance, then representing each word of an utterance as a GLoVE-embedded word-vector, normalizing the vectors, averaging all word-vectors of an utterance, and finally computing the cosine similarity between the utterance and each utterances of the training set. The k then determines the number of closest training utterances (neighbors) that will be taken into account when calculating the label for the new utterance, i.e., if $k = 5$, the five closest training utterances will be considered. In the case of kNN-C, we combine the scores of the k neighbors with a conservative *mode* function, i.e., the unseen utterance is assigned the score that the majority of neighbors carry and the one with the lowest value if multiple exist. In the case of kNN-R, we combine the values by averaging the scores of the nearest neighbors. The kNN algorithms serve as a baseline as they are not trainable, i.e., for each new utterance all distances to all training examples must be computed again and thus all training data must be stored.

The second set of algorithms we applied to the data are support vector machines (SVMs). Unlike kNN algorithms, SVMs build a model from the training data, after which the data can be discarded. They are particularly suited for high-dimensional feature spaces. The core idea of SVMs for classification lies in finding a linear separation boundary between classes such that the space between the closest training examples on either side of the decision boundary (the margin) is maximized. To this end, they can leverage kernel functions to map classes that are not linearly separable in a lower-dimensional space to a higher-dimensional space. In SVMs for regression, on the other hand, a regression is fit to the data. The aim to maximize the margin around the regression line such that the error remains below a certain threshold. For the SVM algorithms, we again represented the utterances as averages of word-vectors. These were then standardized and used to train separate SVMs for each schema.

The final set of algorithms we used to model the data were recurrent neural networks (RNNs). Neural networks commonly consist of an input and an output layer and any number of hidden layers. The input layer holds nodes that simply pass on the numerical representation of the data. Each further layer is comprised of nodes and connections that transform the input. Each node combines all the signals coming in from the previous layer (transfer function) and decides whether or not to pass a signal on (activation function). Nodes of one layer are connected with the nodes of the next layer via weighting functions that amplify or discount the signal by means of multiplication. The output layer holds nodes that transform the signal to the desired type of output value, e.g., a value between 0 and 1. Neural networks become deeper with each additional hidden layer. While feedforward networks, in which the signal travels only in one direction from input to output layer, cannot deal with sequential input data, RNNs are a type of deep neural network specifically designed for this purpose. Thus, unlike the kNN and the SVM approaches, they can account for the temporal aspect of utterances as sequences of words. They do this by retaining a memory of the previous words, i.e., the output of the RNN for the previous word is fed back into the RNN together with the current word. Again, two ways of modelling the data were explored in this research: a set of separate RNNs per schema and a multi-label RNN. The per-schema RNNs allow for assessing the potential added benefit of the deep neural network architecture. For these models, we treat the ordinal scores as separate classes, ignoring the ordering. Each of the nine RNNs in the set outputs a vector of four values between 0 and 1, each value expressing the confidence of the algorithm that the utterance should be assigned a score of 0, 1, 2, or 3 for the specific schema. To obtain the schema score, the score with the highest confidence is selected. The multi-label RNN, on the other hand, can leverage interdependencies between the schemas as it has knowledge of all schema scores at the same time. It predicts all nine schemas simultaneously and outputs a value between 0 and 1 for each schema. In preparing our analysis script for publication, we encountered the challenge that despite setting all random seeds as required, the trained RNNs showed a small degree of variability in the output when re-running the script. We have therefore chosen a stochastic approach: for both RNN approaches, we first train the models 30 times, we then predict all items of the test set with all 30 models, and finally, we select for the median model in terms of performance. All results reported below are based on the median multi-label RNN and the median per-schema RNN set.

It must be stressed at this point that we only test whether a machine is able to detect patterns at all and do not strive to obtain the best scoring performance. As a consequence, a number of refinement possibilities, such as sequence to vector models or extensive hyperparameter tuning, were not explored.

H2: DOWNWARD ARROW CONVERGES

To examine whether utterances developed with the downward arrow technique converge to a schema, we aimed to predict the algorithm's *scoring accuracy* from the *depth* of utterances. We assigned *depth*= 1 for the automatic thought and increased it incrementally with every downward arrow technique step. Figure 4.1 shows the number of thought records in our dataset with a specific depth. To determine *scoring accuracy*, we used the predictions made on the test set with the median set of per-schema RNNs of Hypothesis 1. For each utterance, the Spearman correlation between the algorithmically predicted and manually assigned scores serves as the measure. Thus, if an utterance such as "I will never be loved" was scored as $[3,0,0,0,0,0,0,1,0]$ manually on the nine schemas and received the scores $[2,0,0,0,1,0,0,1,2]$ by the RNN, the resulting *scoring accuracy* for this utterance would be $\rho = 0.59$, i.e., the Spearman correlation between the two vectors of scores. To study the effect of depth on scoring accuracy, we conducted a multilevel analysis; the data structure required a three-level linear model with the *depth* as a fixed effect and the *automatic scoring accuracy* as the dependent variable. For each participant (Level 3), there are several thought records (Level 2) and for each thought record, there are several utterances (Level 1). The null model predicts the scoring accuracy from the mean scoring accuracy per participant and thought record. The model therefore has random intercepts at Level 3 and at Level 2 nested within Level 3 (thought records nested in participants). For Model 1, the fixed effect *depth* was added to the null model. We expected to see an increase in automatic scoring accuracy as utterance depth increases.

H3: SCHEMA PATTERNS ARE SIMILAR ACROSS THOUGHT RECORD TYPES

The next analyses tested whether schemas observed in the scenario-based (closed) thought records are predictive of schemas observed in real-life (open) thought records. For this, only the manually assigned scores were used. Each participant completed two achievement-related and two interpersonal closed thought records. The first author labeled all open thought record scenarios as either interpersonal or achievement-related (intercoder agreement with a second independent coder on all open thought records was substantial with Cohen's $\kappa = 0.68$). Nine linear regression models were fit with *schema presence in closed thought records* as the only predictor and *schema presence in the open thought record* as the only outcome variable. Thus, we fit one model for each schema. For example, in the Health schema model, the presence of the Health schema in closed thought records predicts the presence of the Health schema in open thought records. To determine *schema presence in closed thought records*, we identified the two closed thought records with the same situation type (interpersonal or achievement-related) as

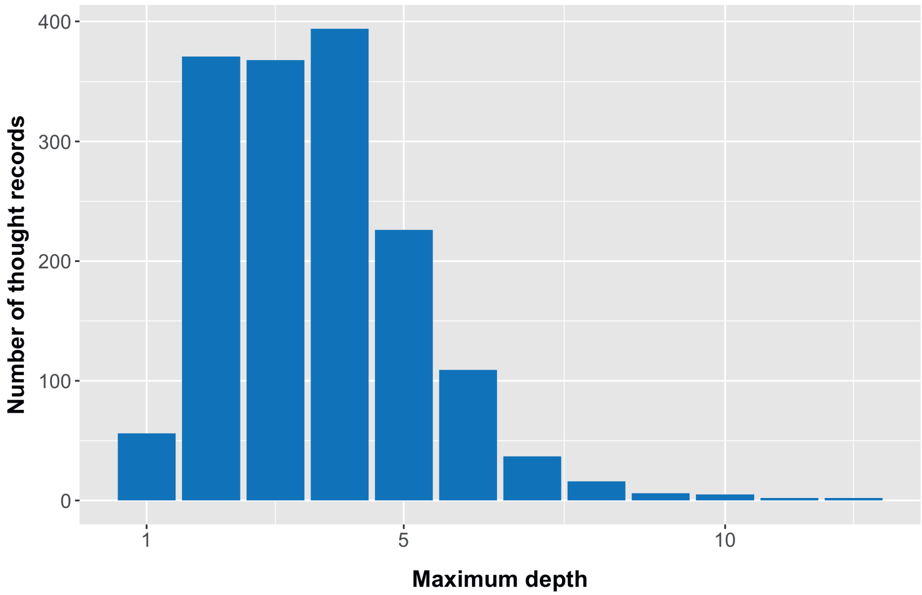

Figure 4.1: Number of thought records having a certain depth, the depth is the number of downward arrow steps + 1 for the automatic thought.

the open thought record. For each of the nine schemas, we then took the highest score across utterances of both closed thought records and average these two values together. For example, let us assume that a participant described an interpersonal situation in the open thought record. To calculate the predictor for the *Health* schema, the two interpersonal closed thought records of this participant were identified and from each the highest score obtained on the *Health* schema across utterances was taken, leading to two scores, which were then averaged together. We followed the same procedure for the outcome variable, *schema presence in the open thought record*. However, since there is only one such thought record for each participant, no averaging was needed. Appendix L illustrates the procedure with a concrete example for clarification.

H4: MENTAL ILLNESSES HAVE ASSOCIATED SCHEMAS

The final hypothesis is an exploratory investigation of whether the outcomes from the mental health questionnaires can be predicted from the schema patterns. To this end, we created a summary score per schema and participant. The summary score was calculated by first taking per participant, thought record, and schema the maximum score (0-3) across utterances. This gives one value for each schema for the five thought records a participant completed. These values were then re-coded into a binomial value, with all values smaller or equal to 2 mapping to 0 and 3 mapping to 1. Thus, we only considered schemas that were clearly and unambiguously present. Finally, the binomial values were summed within a participant. Each participant could therefore obtain a maximum value of 5 for a schema if the schema was clearly present in all five completed thought records of the participant. We then created five linear models, each one taking one of the mental health measures (HDAS Depression, HDAS Anxiety, BDI, Cognitive Distortions Relatedness, Cognitive Distortions Achievement) as outcome variable. Every model has the nine schemas as predictors. Since the same data were used to predict five different outcomes, we used a Bonferroni correction to adjust the significance threshold to $\alpha = 0.05/5 = 0.01$. Just like for Hypothesis 3, the manually assigned scores were used, as this was both suited for testing the hypothesis and less susceptible to errors than the automatically assigned scores.

4.4. RESULTS

To gain insight into the collected data, Table 4.3 shows the frequencies of each score per schema. In total, there were 5747 utterances.

Table 4.3: Number of utterances with a specific score per schema as manually scored by the first author. Percentages are provided in parentheses. Schemas are sorted as in the article by Millings *et al.* [8]

Schema	**Score**			
	0 (has absolutely nothing to do with schema)	1 (corresponds a little bit with schema)	2 (corresponds largely with schema)	3 (corresponds completely with schema)
Attachment	4047 (70.42 %)	446 (7.76 %)	272 (4.73 %)	982 (17.09 %)
Competence	4151 (72.22 %)	314 (5.46 %)	157 (2.73 %)	1125 (19.58 %)
Global self-evaluation	4548 (79.14 %)	226 (3.93 %)	280 (4.87 %)	693 (12.06 %)
Health	5428 (94.45 %)	56 (0.97 %)	46 (0.80 %)	217 (3.78 %)
Power and Control	5089 (88.55 %)	390 (6.79 %)	154 (2.68 %)	114 (1.98 %)
Meta-cognition	5626 (97.89 %)	61 (1.06 %)	41 (0.71 %)	19 (0.33 %)
Other people	5593 (97.32 %)	92 (1.60 %)	44 (0.31 %)	18 (0.31 %)
Hopelessness	4931 (85.80 %)	582 (10.13 %)	174 (3.03 %)	60 (1.04 %)
Other's views on self	4688 (81.57 %)	129 (2.24 %)	639 (11.11 %)	29 1(5.06 %)

4.4.1. H1: SCHEMAS CAN BE AUTOMATICALLY EXTRACTED

For the majority of schemas, all algorithms could assign scores to the utterances that correlated with the human scores well above what would be expected by chance alone (see Table 4.4). Furthermore, for all schemas, there was at least one effective algorithm.

As determined with the validation set, the best parameter choice for kNN-C was $k = 4$, while for kNN-R, it was $k = 5$. Both support vector approaches performed best with a radial basis function kernel. The best-performing multi-label RNN was trained in batches of 32 utterances and with 100 epochs. It consists of two hidden layers: an embedding layer, performing the GLoVE embeddings, and a bidirectional long short-term memory layer of 100 nodes. It was trained with a dropout probability of 0.1

Table 4.4: Spearman correlation and bootstrapped confidence intervals of predicted scores with manually assigned scores per model and schema. The result of the best model per schema is shown in bold font.

Schema	Model Outcome					
	kNN-C	kNN-R	SVM	SVR	per-schema RNNs	multi-label RNN
Attachment	0.55 [0.51,0.60]	0.63 [0.59,0.65]	0.65 [0.61,0.68]	0.68 [0.65,0.70]	**0.73 [0.70,0.76]**	0.67 [0.66,0.72]
Competence	0.69 [0.64,0.73]	0.66 [0.63,0.69]	0.68 [0.65,0.72]	0.64 [0.61,0.67]	**0.76 [0.72,0.79]**	0.66 [0.64,0.69]
Global self-evaluation	0.40 [0.33,0.46]	0.41 [0.36,0.46]	0.36 [0.31,0.40]	0.49 [0.45,0.52]	**0.58 [0.54,0.63]**	0.49 [0.45,0.53]
Health	0.74 [0.65,0.81]	0.53 [0.44,0.60]	0.73 [0.65,0.81]	0.35 [0.31,0.40]	**0.75 [0.65,0.82]**	0.35 [0.31,0.39]
Power and Control	0.11 [0.02,0.18]	0.23 [0.17,0.27]	nan [0.00,1.00]	0.31 [0.26,0.35]	0.28 [0.20,0.35]	**0.31 [0.27,0.34]**
Meta-cognition	nan [0.00,1.00]	0.10 [0.01,0.20]	nan [0.00,1.00]	0.11 [0.06,0.16]	-0.01 [0.00,-0.01]	**0.11 [0.06,0.14]**
Other people	0.28 [0.00,1.00]	0.24 [0.17,0.31]	nan [0.00,1.00]	0.19 [0.14,0.24]	**0.22 [0.07,0.33]**	0.16 [0.10,0.20]
Hopelessness	0.48 [0.44,0.55]	0.51 [0.47,0.56]	0.49 [0.43,0.53]	0.54 [0.51,0.57]	**0.63 [0.56,0.68]**	0.53 [0.50,0.56]
Other's views on self	0.45 [0.41,0.51]	0.46 [0.42,0.50]	0.48 [0.43,0.53]	0.52 [0.48,0.55]	**0.58 [0.52,0.63]**	0.50 [0.47,0.54]

The abbreviation *nan* that resulted for some schemas and some models stands for *not a number* and is caused by the absence of variance in the prediction (see text for details).

4

and categorical cross-entropy loss. The nine nodes of the output layer use a sigmoid activation function. The metric for choosing the best model was the mean absolute error. The individual models, we set up differently, but adopted some of the hyperparameters of the multi-label model (namely, the batch size, the number of LSTM nodes, the dropout rate, and the loss function). For each schema, the individual models have four outputs, one for each of the four possible scores. The activation function of the final layer is a softmax, to express the likelihood with which a certain utterance has each of the scores.

It can be seen from Table 4.4 that the per-schema RNNs perform best overall. They take the structure of the data most closely into account, both in terms of the utterances (sequential input) and in terms of the scores (one output neuron per score), and were also able to produce the best predictions for most of the schemas. Any possible advantage of exploiting relationships between schemas was not observable in the results, since the multi-label RNN did not clearly outperform all the other models for any one schema. Interestingly, the *Health* schema is consistently better identifiable by the classification algorithms (kNN-C, SVM, and the per-schema RNNs), while the *Power and Control* schema could be better identified by the regression algorithms (kNN-R, SVR, and multi-label RNN). Nan-values for some algorithms and schemas can be explained by the algorithms predicting 0-values for all items of the test set due to not having seen enough non-zero training examples. Similarly, consistently low correlations for certain schemas (*Meta-cognition*, *Power and Control*, and *Other people*) are the result of a combination of few non-zero training examples (compare Table 4.3) and variations in the words used within those non-zero training examples. The *Health* schema, for example, could be predicted fairly well, despite few non-zero training examples, because the non-zero training examples had similar wording, frequently including words related to *dieting* and *weight-loss* caused by a scenario with this theme.

4.4.2. H2: Downward arrow converges

The mean correlation between the predicted schema scores and the manually labeled schema scores was found to be 0.75 ($b = 0.75, t(220.76) = 46.97, p < 0.001$) when the nesting structure of utterances nested within thought records and thought records nested within participants is taken into account via random intercepts. Steps at a deeper level could not be scored better by the best model of H1 than steps at a more shallow level. The scoring accuracy, as measured by the Spearman correlation, did not improve with additional steps of the downward arrow technique ($\chi^2(1) = 1.21, p = 0.27$).

4.4.3. H3: Schema patterns are similar across thought record types

Fig 4.2 shows the percentage of utterances having a certain schema (manually assigned score > 0) for the open and closed thought records in our dataset. It can be seen that, across participants, schemas are similarly distributed in the two TR types: the mean difference over all schemas is 3.69 % with the *Other people* schema having the smallest difference (0.02 %) and the Competence schema the largest one (9.24 %). Some schemas are more present in open TRs (e.g., the *Power and Control* schema) and others in closed ones (e.g., the *Health* or *Competence* schemas).

On the level of the individual, a series of linear regression models tested whether the

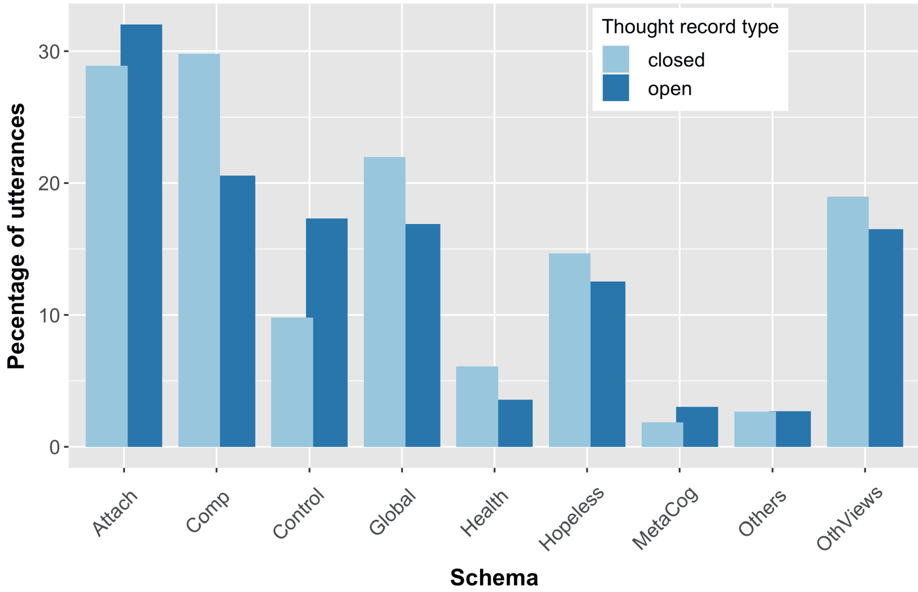

Figure 4.2: Percentage of utterances that reflect a certain schema (score > 0) in open and closed thought records respectively.

active schemas in closed thought records could predict the active schemas in the open thought record of the same scenario type (interpersonal or achievement-related). The outcome variable was the maximum schema score of the open thought record, while the predictor variable was the average of the maximum schema score of the two closed thought records of the same scenario type. Table 4.5 presents the results of the models. For the *Competence* schema, 43 % of the variance in the open thought record could be predicted from the closed thought records of the same scenario type, while for the *Attachment* schema this was the case for 20 % of the variance.

4.4.4. H4: MENTAL ILLNESSES HAVE ASSOCIATED SCHEMAS

Five linear regression models tested whether there is a link between the active schemas of participants as indicated in thought records and the outcomes on five mental health inventories (see Appendix M for the table of results). The Bonferroni-corrected α of 0.01 serves as the significance threshold. For both symptom-based mental health inventories for depression, i.e., the HDAS–Depression and the BDI-IA, none of the schemas was a significant predictor of the outcome scores. However, for the anxiety inventory (HDAS–Anxiety) and the two Cognitive Distortion scales, we found that the *Global Self-Evaluation* schema was linked to these measures: all other schemas being equal, any additional thought record with a clearly present *Global Self-Evaluation* schema resulted in a 0.63 ($\beta = 0.18$) point increase on the HDAS–Anxiety ($t = 2.98, p = 0.003$),

Table 4.5: Outcomes of the per-schema linear regression models to test whether participants show similar schema patterns in open as in closed thought records of the same scenario type (interpersonal vs. achievement-related).

Schema	b	95 % CI	t	p	F(1,316)	Adj. R^2
Attachment	0.53	[0.42,0.65]	9.07	< 0.001	82.28	0.20
Competence	0.79	[0.69,0.86]	15.58	< 0.001	242.7	0.43
Global self-evaluation	0.31	[0.18,0.45]	4.50	< 0.001	20.23	0.06
Health	0.21	[0.07,0.35]	2.96	< 0.01	8.77	0.02
Power and Control	-0.08	[-0.31,0.14]	-0.73	0.47	0.53	0.00
Meta-cognition	0.17	[-0.04,0.37]	1.60	0.11	2.57	0.00
Other people	0.12	[-0.03,0.28]	1.54	0.12	2.38	0.00
Hopelessness	0.30	[0.16,0.44]	4.18	< 0.001	17.49	0.05
Other's views on self	0.24	[0.12,0.37]	3.77	< 0.001	14.22	0.04

a 2.11 (β = 0.24) point increase on the Cognitive Distortions – Relatedness measure ($t = 3.83, p < 0.001$), and a 2.06 ($\beta = 0.21$) point increase on the Cognitive Distortions – Achievement measure ($t = 3.64, p < 0.001$). Finally, the number of thought records with a clearly present *Power and Control* schema also significantly predicted the Cognitive Distortions – Relatedness measure ($b = 3.44, \beta = 0.15, t = 2.69, p = 0.007$).

4.5. DISCUSSION AND CONCLUSION

As the first and core hypothesis, we posited that utterances of thought records could be automatically scored with respect to their underlying schemas. With all three machine learning algorithm types (kNN, SVM, and RNN) that were tried, we found affirmative evidence for this. Even when only representing utterances as averages of word vectors, linguistic patterns could be learned (as in the case of the kNN and SVM models). The best-performing algorithm across schemas were the per-schema RNNs. Although for many schemas there is only a small difference in outcome between the best and second-best algorithms, no *one* second-best algorithm emerges. Yet, the fact that the per-schema RNNs outperformed the other algorithms on several schemas provides an indication that the information contained in the word order may be useful for optimal scoring performance. Looking at the best outcomes for each schema, correlations between predicted scores and actual scores ranged from $\rho = 0.11$ to $\rho = 0.81$. The schemas for which the algorithms saw many training examples with non-zero scores (*Attachment* and *Competence*) could be classified well by all. However, the *Health* schema also exhibits good classification potential. This is probably due to very distinctive language as a result of one specific scenario related to dieting and weight loss, i.e., many utterances scored on the *Health* schema contained words such as "fat," "gain," "overweight," "diet," or "skinny." These words are likely to be within close proximity of each other in the word vector space, possibly leading to similar utterance representations and hence a clear linguistic pattern. Although the outcomes from the models cannot be compared directly to the interrater (weighted Cohen's $\kappa = 0.79$) and intrarater (weighted Cohen's $\kappa = 0.83$) reliability scores we obtained on a sample of the data, the reliability scores give a good indication that the nature of the data and the scoring method are limiting the level of agreement that can be achieved as there is some room for interpretation of utterances, schema definitions, and even in the scale points. This, in turn, means that automatic scoring accuracy cannot be expected to exceed the human performance, since the algorithm only has the human-labeled data to learn from. As our goal was only to see whether scoring was feasible and not to obtain the best possible performance, we did not explore many of the other available options for data representation, data augmentation, or modeling. These include looking into more state-of-the-art ways of representing utterances, such as BERT [48] or GPT-3 [49], making better use of the ordering information in the scores, creating a corpus-specific word vector space, or trying to generate more training examples with neural networks. Together with this article, however, we make our collected dataset publicly available and invite other researchers or machine learning enthusiasts to improve upon our results.

As our second hypothesis, we predicted an algorithm trained on utterances of varying downward arrow technique (DAT) depths to be able to better score the utterances as the depth increases. This is because the DAT was specifically developed to aid patients in identifying their maladaptive schemas, taking the automatic thought from the completed thought record as starting point. After applying the technique, a schema formulation should be reached. In our dataset and with the best performing algorithm of Hypothesis 1, we did not find support for Hypothesis 2. This may be due to only very few participants completing more than three steps. Since our participants were drawn from a non-clinical population and had never practiced thought recording before, it is possible that they did not reach the same level of introspection as a clinical, therapist-guided group would. Additionally, motivations differ between this group of participants (motivated by financial gains) and a clinical group (motivated by mental health gains). Further research might therefore compare our results to those obtained in a clinical setting.

As our third hypothesis, we expected that the dysfunctional schemas that were active when completing scenario-based (closed) thought records would be able to predict those active when completing a real-life personal (open) thought record within participants, given that the closed and open thought records matched in scenario type, that is, both revolved around either an interpersonal or an achievement-related situation. In our study, we relied mostly on prescripted scenarios and asked participants to respond to these as if they were real. We found support for our third hypothesis. For two schemas, we even observed that 20 % (Attachment schema) and 43 % (Competence schema) of the variance in the open thought record score could be predicted by the scores in the closed thought records. This corresponds to the central idea of schema theory: if a person holds a certain schema, this may be activated in various situations of a similar kind and influence how the person appraises the situation [50]. Consequently, we regard it as a viable option to use prescripted scenarios instead of real-life ones when needed. However, it can be argued that the *Attachment* schema may be particularly relevant in *interpersonal* scenarios, while the *Competence* schema plays more in *achievement-related* scenarios. The medium to large (as defined by Cohen [51, p. 413]) effect that shows for these two schemas may therefore be the result of labeling the open thought records as belonging to one of these two scenario types and splitting the dataset accordingly. On the basis of these considerations, the scenarios should be carefully chosen and varied enough to be able to unveil all possible schemas when substituting closed scenarios for open ones. Therefore, a larger number of thought records may be needed than when using open thought records.

Lastly, as our fourth hypothesis, we proposed that the schema patterns across all thought records of a person can predict outcomes on depression, anxiety, and cognitive distortion scales. We found partial support for this hypothesis. Concerning the link between schemas and mental health outcomes, we found no relationship between the schemas and outcomes on both depression inventories. While Millings *et al.* [8] observed a higher prevalence of the *Power and Control* schema in people with anxious tendencies, we observed higher scores on the HDAS – Anxiety scale when participants had a negative *Global Self-Evaluation*. This schema was also a good predictor of cognitive distortions linked to relatedness and achievement. We could not replicate the finding reported in [8] that higher anxiety scores were linked to a less frequently active *Attachment* schema either. This may, however, be a population effect, as we did not work with a clinical population. Yet, an active *Power and Control* schema was related to more cognitive distortions pertaining to relatedness in our dataset. On the whole, we found more links between schemas and cognitive distortions than schemas and mental health inventory outcomes. This may have to do with thought records being a cognitive task concerned with unveiling dysfunctional cognitions, which connects directly to the cognitive distortion measure and less to the symptom-based nature of the mental health inventories. Additionally, a single thought record presents a snapshot of a person's thought processes at best and typically many are completed in the course of therapy before a certain schema emerges as clinically relevant [52]. Thus, more extensive experimentation looking for recurring schemas and thought patterns in a clinical population over an extended period of time may paint a clearer picture with regard to the usefulness of the automatic schema-labeling method for therapeutic or diagnostic purposes. On a more practical note, our results indicate that a software application striving to construct a long-term user model might benefit from assigning a higher a priori probability to the activation of the *Global Self-Evaluation* schema after an initial assessment of the user's anxiety levels and cognitive distortions. Still, making a choice on this requires trading off the collection of such sensitive mental health scale data against the added benefit of improving the prediction model. One limitation that must also be considered here is the fact that our results are based on a particular method for combining the utterance labels across all thought records of a participant. This method was a choice and various other methods are conceivable, potentially leading to other outcomes.

The core finding of this research is that it is possible to interpret rich natural language data from the psychotherapy domain using a computer algorithm. The applicability of this finding extends especially to various kinds of psychological assessment. For example, one of the common applications of e-mental health in research are ecological momentary assessments. To date, these typically employ multiple choice response items for self-report measures, which may be combined with sensor readings from handheld devices or wearables (compare [53] for depression). Our findings are promising for effectively using more open response formats and journaling, thus allowing participants to better describe their thoughts, feelings, and behaviors in their own words while minimizing analysis effort. This is also interesting in light of new methodological developments in mental health assessment as a result of big data, such as studying symptom dynamics of individuals with network analyses [54]. Such dynamic networks of symptoms may be augmented with the schemas as determined from thought records to better understand how the activation and co-activation of schemas and other symptoms predicts mental well-being over time. Another possible area of application are cognitive case conceptualizations [55]. These are comprehensive outlines of the patient's problems as first drafted during the intake conversation between patient and therapist. They are continually refined throughout therapy, often on the basis of homework assignments [56]. With the possibility of automatically interpreting thought record data, it may be possible to sketch a first CCC before therapy by collecting and analyzing thought records over the period of time the patient spends on a waiting list and to then collaboratively update this CCC with the therapist as new thought records are completed during therapy. Moreover, Schema Therapy [50] presents a further thought classification system to that of schemas, namely that of schema modes. Furthermore, it proposes a much larger set of schemas than the ones used in this research. With a background in Schema Therapy, it may be possible to use our collected dataset and re-label the data with respect to these other schemas or the schema modes. Beyond psychological assessment, Millings *et al.* [8] propose future work to compare the derivation of schemas using the downward arrow technique in an online setting to a face-to-face therapy setting. We would be interested in adding the algorithmically derived schemas to this comparison in a long-term study.

In conclusion, we have presented an algorithmic benchmark solution for automatically scoring utterances extracted from thought records with respect to the underlying schema. We expect the model and the opportunities resulting from the positive results to be of relevance for the field of clinical psychology. For the field of computer science, we make the dataset of collected thought records publicly available. Especially the complexity of the outcome variables (ordinal multi-label) may be intriguing for those looking to develop new algorithms or test existing ones. Lastly, for both fields, clinical psychology and computer science, the dataset could be used to study and advance automatically generated explanations of the algorithmic schema identification. In so doing, it can contribute to diagnoses and explainable artificial intelligence (XAI) technology, which is seen as an important requirement for responsible and effective AI-implementation (e.g., [57]).

4

BIBLIOGRAPHY

1. Burger, F., Neerincx, M. A. & Brinkman, W. P. Natural language processing for cognitive therapy: Extracting schemas from thought records. *PLoS ONE* **16** (2021).
2. Burger, F., Neerincx, M. A. & Brinkman, W.-P. Technological State of the Art of Electronic Mental Health Interventions for Major Depressive Disorder: Systematic Literature Review. *Journal of Medical Internet Research* **22,** e12599 (2020).
3. Beck, A. T. Cognitive therapy: A 30-year retrospective. *American Psychologist* **46,** 368–375 (1991).
4. Beck, A. T. Thinking and Depression: II. Theory and Therapy. *Archives of General Psychiatry* **10,** 561–571 (1964).
5. Burns, D. D. *The Feeling Good Handbook* (Plume, 1999).
6. Osmo, F., Duran, V., Wenzel, A., de Oliveira, I. R., Nepomuceno, S., Madeira, M. & Menezes, I. The negative core beliefs inventory: Development and psychometric properties. *Journal of Cognitive Psychotherapy* **32,** 67–84 (2018).
7. Wong, Q. J., Gregory, B., Gaston, J. E., Rapee, R. M., Wilson, J. K. & Abbott, M. J. Development and validation of the Core Beliefs Questionnaire in a sample of individuals with social anxiety disorder. *Journal of Affective Disorders* **207,** 121–127 (2017).
8. Millings, A. & Carnelley, K. B. Core belief content examined in a large sample of patients using online cognitive behaviour therapy. *Journal of Affective Disorders* **186,** 275–283 (2015).
9. Jurafsky, D. & Martin, J. *Speech and Language Processing (3rd ed. draft)* `https://web.stanford.edu/~jurafsky/slp3/`. Accessed in 2019.
10. Laranjo, L., Dunn, A. G., Tong, H. L., Kocaballi, A. B., Chen, J., Bashir, R., Surian, D., Gallego, B., Magrabi, F., Lau, A. Y. & Coiera, E. Conversational agents in healthcare: A systematic review. *Journal of the American Medical Informatics Association* **25,** 1248–1258 (2018).
11. Montenegro, J. L. Z., da Costa, C. A. & da Rosa Righi, R. Survey of conversational agents in health. *Expert Systems with Applications* **129,** 56–67 (2019).
12. McTear, M. F. Spoken Dialogue Technology: Enabling the Conversational User Interface. *ACM Computing Surveys* **34,** 90–169 (2002).
13. Ni, L., Lu, C., Liu, N. & Liu, J. MANDY: Towards a smart primary care chatbot application. in *International symposium on knowledge and systems science* (Springer Verlag, 2017), 38–52.
14. Amato, F., Marrone, S., Moscato, V., Piantadosi, G., Picariello, A. & Sansone, C. Chatbots meet ehealth: Automatizing healthcare. in *CEUR Workshop Proceedings Vol-1982* (CEUR-WS, 2017), 40–49.

15. Jin, L., White, M., Jaffe, E., Zimmerman, L. & Danforth, D. Combining cnns and pattern matching for question interpretation in a virtual patient dialogue system. in *Proceedings of the EMNLP 2017 - 12th Workshop on Innovative Use of NLP for Building Educational Applications* (Association for Computational Linguistics (ACL), 2017), 11–21.
16. Ochs, M., de Montcheuil, G., Pergandi, J.-M., Saubesty, J., Pelachaud, C. I., Mestre, D. & Blache, P. An architecture of virtual patient simulation platform to train doctors to break bad news. in *Proceedings of the 2017 Conference on Computer Animation and Social Agents (CASA)* (2017).
17. van Heerden, A., Ntinga, X. & Vilakazi, K. The potential of conversational agents to provide a rapid HIV counseling and testing services. in *Conference Proceedings of the 2017 International Conference on the Frontiers and Advances in Data Science, FADS 2017* (Institute of Electrical and Electronics Engineers Inc., 2017), 80–84.
18. Rizzo, A. A., Kenny, P. & Parsons, T. D. Intelligent Virtual Patients for Training Clinical Skills. *Journal of Virtual Reality and Broadcasting* **8** (2011).
19. Miner, A. S., Milstein, A., Schueller, S., Hegde, R., Mangurian, C. & Linos, E. Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. *JAMA Internal Medicine* **176,** 619–625 (2016).
20. Abd-alrazaq, A. A., Alajlani, M., Alalwan, A. A., Bewick, B. M., Gardner, P. & Househ, M. An overview of the features of chatbots in mental health: A scoping review. *International Journal of Medical Informatics* **132** (2019).
21. Bresó, A., Martínez-Miranda, J., Botella, C., Baños, R. M. & García-Gómez, J. M. Usability and acceptability assessment of an empathic virtual agent to prevent major depression. *Expert Systems* **33,** 297–312 (2016).
22. Shamekhi, A., Bickmore, T., Lestoquoy, A., Negash, L. & Gardiner, P. Blissful agents: Adjuncts to group medical visits for chronic pain and depression. in *Lecture Notes in Artificial Intelligence* **10011 LNAI** (Springer Verlag, 2016), 433–437.
23. Fitzpatrick, K. K., Darcy, A. & Vierhile, M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. *JMIR Mental Health* **4** (2017).
24. Guntuku, S. C., Yaden, D. B., Kern, M. L., Ungar, L. H. & Eichstaedt, J. C. Detecting depression and mental illness on social media: an integrative review. *Current Opinion in Behavioral Sciences* **18,** 43–49 (2017).
25. Al-Mosaiwi, M. & Johnstone, T. In an Absolute State: Elevated Use of Absolutist Words Is a Marker Specific to Anxiety, Depression, and Suicidal Ideation. *Clinical Psychological Science* **6,** 529–542 (2018).
26. Edwards, T. & Holtzman, N. S. A meta-analysis of correlations between depression and first person singular pronoun use. *Journal of Research in Personality* **68,** 63–68 (2017).

27. Newell, E. E., McCoy, S. K., Newman, M. L., Wellman, J. D. & Gardner, S. K. You Sound So Down: Capturing Depressed Affect Through Depressed Language. *Journal of Language and Social Psychology* **37,** 451–474 (2018).

28. Kshirsagar, R., Morris, R. & Bowman, S. Detecting and Explaining Crisis. in *Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology—From Linguistic Signal to Clinical Reality* (Association for Computational Linguistics (ACL), 2017), 66–73.

29. Gkotsis, G., Oellrich, A., Velupillai, S., Liakata, M., Hubbard, T. J., Dobson, R. J. & Dutta, R. Characterisation of mental health conditions in social media using Informed Deep Learning. *Scientific Reports* **7** (2017).

30. Benton, A., Mitchell, M. & Hovy, D. Multi-task learning for mental health using social media text. *arXiv preprint arXiv:1712.03538* (2017).

31. Bahdanau, D., Cho, K. H. & Bengio, Y. Neural machine translation by jointly learning to align and translate. in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings* (2015).

32. Wolf, M. J., Miller, K. W. & Grodzinsky, F. S. Why we should have seen that coming: comments on microsoft's tay "experiment," and wider implications. *The ORBIT Journal* **1,** 1–12 (2017).

33. Dobson, D. & Dobson, K. S. *Evidence-based practice of cognitive-behavioral therapy.* (Guilford Press, 2017).

34. Barlow, D. H. *Clinical handbook of psychological disorders: A step-by-step treatment manual* (Guilford Press, 2021).

35. Cook, M. N. *Transforming teen behavior: Parent-teen protocols for psychosocial skills training.* (Academic Press, 2015).

36. Schoth, D. E. & Liossi, C. A systematic review of experimental paradigms for exploring biased interpretation of ambiguous information with emotional and neutral associations. *Frontiers in Psychology* **8** (2017).

37. Lefebvre, M. F. Cognitive distortion and cognitive errors in depressed psychiatric and low back pain patients. *Journal of Consulting and Clinical Psychology* **49,** 517–525 (1981).

38. Pössel, P. Cognitive Error Questionnaire (CEQ): Psychometric Properties and Factor Structure of the German Translation. *Journal of Psychopathology and Behavioral Assessment* **31,** 264–269 (2009).

39. Barber, J. P. & Derubeis, R. J. The ways of responding: A scale to assess compensatory skills taught in cognitive therapy. *Behavioral Assessment* **14,** 93–115 (1992).

40. Covin, R., Dozois, D. J., Ogniewicz, A. & Seeds, P. M. Measuring cognitive errors: Initial development of the cognitive distortions scale (CDS). *International Journal of Cognitive Therapy* **4,** 297–322 (2011).

41. Beck, A. T. & Alford, B. A. *Depression: Causes and treatment* (University of Pennsylvania Press, 2009).

4

42. Burger, F. *Dataset and Analyses for Extracting Schemas from Thought Records using Natural Language Processing.* 4TU.ResearchData `https://doi.org/10.4121/16685347`. (2021).

43. Zigmond, A. S. & Snaith, R. P. The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica* **67,** 361–370 (1983).

44. Beck, A. T. *Cognitive therapy of depression* (Guilford Press, 1979).

45. Gutiérrez, P. A., Pérez-Ortiz, M., Sánchez-Monedero, J., Fernández-Navarro, F. & Hervás-Martínez, C. Ordinal Regression Methods: Survey and Experimental Study. in *IEEE Transactions on Knowledge and Data Engineering* **28** (IEEE Computer Society, 2016), 127–146.

46. Cheng, J., Wang, Z. & Pollastri, G. A neural network approach to ordinal regression. in *Proceedings of the International Joint Conference on Neural Networks* (IEEE *Xplore*, 2008), 1279–1284.

47. Pennington, J., Socher, R. & Manning, C. D. GloVe: Global vectors for word representation. in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014* (Association for Computational Linguistics (ACL), 2014), 1532–1543.

48. Devlin, J., Chang, M. W., Lee, K. & Toutanova, K. BERT: Pre-training of deep bidirectional transformers for language understanding. in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL HLT 2019* **1** (Association for Computational Linguistics (ACL), 2019), 4171–4186.

49. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I. & Amodei, D. Language models are few-shot learners. *Advances in Neural Information Processing Systems* **2020-December** (2020).

50. Young, J. E., Klosko, J. S. & Weishaar, M. E. *Schema therapy: A practitioner's guide* (Guilford Press, 2006).

51. Cohen, J. *Statistical power analysis for the behavioural sciences.* (Lawrence Earlbaum Associates, Inc., 1988).

52. DeRubeis, R. J., Webb, C. A., Tang, T. Z. & Beck, A. T. Cognitive therapy. in *Handbook of cognitive-behavioral therapies* (ed Dobson, K. S.) (Guilford Press, 2010), 277–316.

53. Colombo, D., Fernández-Álvarez, J., Patané, A., Semonella, M., Kwiatkowska, M., García-Palacios, A., Cipresso, P., Riva, G. & Botella, C. Current state and future directions of technology-based ecological momentary assessment and intervention for major depressive disorder: A systematic review. *Journal of Clinical Medicine* **8** (2019).

54. Cramer, A. O., Borkulo, C. D. V., Giltay, E. J., Maas, H. L. V. D., Kendler, K. S., Scheffer, M. & Borsboom, D. Major depression as a complex dynamic system. *PLoS ONE* **11** (2016).

55. Needleman, L. D. *Cognitive case conceptualization: A guidebook for practitioners* (Lawrence Erlbaum Associates, Inc., 1999).

56. Cronin, T. J., Lawrence, K. A., Taylor, K., Norton, P. J. & Kazantzis, N. Integrating between-session interventions (Homework) in therapy: The importance of the therapeutic relationship and cognitive case conceptualization. *Journal of Clinical Psychology* **71,** 439–450 (2015).

57. Peeters, M. M., van Diggelen, J., van den Bosch, K., Bronkhorst, A., Neerincx, M. A., Schraagen, J. M. & Raaijmakers, S. Hybrid collective intelligence in a human–AI society. *AI and Society* **36,** 217–238 (2021).

5

USING A CONVERSATIONAL AGENT FOR THOUGHT RECORDING AS A COGNITIVE THERAPY TASK: FEASIBILITY, CONTENT, AND FEEDBACK

This chapter is based on Burger, F., Neerincx, M. A. & Brinkman, W.-P. Using a conversational agent for thought recording as a cognitive therapy task: Feasibility, content, and feedback. *Frontiers in Digital Health* **4** (2022)

Abstract

E-mental health for depression is increasingly used in clinical practice, but patient adherence suffers as therapist involvement decreases. One reason may be the low responsiveness of existing programs: especially autonomous systems are lacking in their input interpretation and feedback-giving capabilities. Here, we explore (a) to what extent a more socially intelligent and, therefore, technologically advanced solution, namely a conversational agent, is a feasible means of collecting thought record data in dialog, (b) what people write about in their thought records, and (c) whether providing content-based feedback increases motivation for thought recording*, a core technique of cognitive therapy that helps patients gain an understanding of how their thoughts cause their feelings. Using the crowd-sourcing platform Prolific, 308 participants with subclinical depression symptoms were recruited and split into three conditions of varying feedback richness using the minimization method of randomization. They completed two thought recording sessions with the conversational agent: one practice session with scenarios and one open session using situations from their own lives. All participants were able to complete thought records with the agent such that the thoughts could be interpreted by the machine learning algorithm, rendering the completion of thought records with the agent feasible. Participants chose interpersonal situations nearly three times as often as achievement-related situations in the open chat session. The three most common underlying schemas were the Attachment, Competence, and Global Self-evaluation schemas. No support was found for a motivational effect of providing richer feedback. In addition to our findings, we publish the dataset of thought records for interested researchers and developers.*

5.1. INTRODUCTION

Software systems increasingly help to prevent and treat depressive disorders. However, Richards *et al.* [2] have shown that the more users are left to their own devices, the higher the dropout rates. Similarly, participants in face-to-face therapy often struggle with adhering to homework assignments [3–7]. We therefore explore (a) whether collecting thought record data when mimicking the conversational style of in-person care with a conversational agent is feasible, (b) what users write about in their thought records, and (c) whether offering feedback that demonstrates an understanding of the situation in response to textual input is motivating.

Depression poses a serious liability to global public health: it has a high lifetime prevalence and takes a greater toll on people's quality-adjusted life-expectancy than many other chronic conditions [8]. Although depression can be treated effectively with medication, psychotherapy, or a combination [9] and possibly even prevented entirely with a toolkit of psychotherapeutic techniques, numerous barriers to seeking and obtaining help exist [10]. One way to address the resulting treatment gap [11] is with *e-mental health for depression*, delivering treatment or prevention programs via electronic devices. As a result of the COVID-19 pandemic, e-mental health for depression is increasingly finding its way into standard clinical practice [12].

The landscape of technology-delivered depression treatment and prevention systems is varied, ranging from video-conferencing with a counselor[1] to fully automated software programs. A literature review on the state of the art of software systems, however, revealed that the majority of systems are low-tech implementations: most of their functional components could receive information from users but were not interpreting or reacting to this input autonomously [13]. This contrasts with face-to-face counseling, in which relational *micro-skills* of the counselor are thought to lead to a better alliance [14] or better rapport [15]. One such micro-skill is called *reflective listening* in the context of motivational interviewing. The counselor demonstrates understanding and empathy by paraphrasing or reflecting on what was said. The advances in various areas of information processing in recent years offer an opportunity to enrich autonomous systems with these micro-skills and observe their effects. Here, we study whether feedback that provides an interpretation of a user's textual input can suffice to motivate users.

One promising technology for use in healthcare contexts are conversational agents. Provoost *et al.* [16], for example, found that an agent that was simply mirroring users' mood in an ecological momentary assessment task already had an adherence-stabilizing effect. Similarly, users who received personalized messages from a conversational agent felt more *heard* by the agent and were more motivated to continue when symptoms worsened than those who did not [17]. And users of Woebot [18], a chatbot for depression treatment, most frequently reported a lack of understanding by the bot as the greatest nuisance when interacting with it.

A therapeutic exercise that might benefit from support by a conversational agent is *thought recording*. It is an integral part of Cognitive Therapy (CT), an evidence-based therapy form often used for the treatment [19] and sometimes used for the prevention

[1]We use the term *counselor* here to subsume primary care providers, coaches, counselors, and therapists.

(e.g. the Penn Resiliency Program [20]) of depressive disorders. CT rests on the idea that understanding, challenging, and changing problematic appraisals (cognitive restructuring) will improve affect. It lends itself to the dialog format because thoughts are often thought and expressed in natural language. Thought record forms provide patients with a structured format for monitoring their feelings, thoughts, and behavior in emotionally difficult situations to gain insight into *core beliefs* or *schemas*, the underlying causative patterns of thinking. Therapists ask patients to complete thought records as close in time to the negatively experienced situation as possible and thus outside of the face-to-face sessions. Patients then bring the records to the sessions to discuss with the therapist. As a consequence, the success of CT depends on patients' homework compliance [21, 22]. However, adherence to homework assignments is difficult for many patients [3–7]. Since depression commonly dampens motivation and a positive outlook on the future, those with symptoms may be particularly difficult to motivate [4].

In short, conversational agents are a promising technology for supporting individuals with depression symptoms in regularly completing thought records. In this work, we explore the feasibility of providing such automated conversational support for thought recording and report on the content of the thought records. In addition, we study whether the agent giving *richer* feedback, that is, feedback demonstrating a greater understanding of the user input, has a motivational effect and whether this effect is partially explained by the *insight* gained from receiving richer feedback. Finally, Grant *et al.* [23] found that people with a high need for self-reflection often keep diaries, indicating that this character trait motivates them to engage in self-reflection. Those who kept diaries, though, did not necessarily have more self-insight than those who did not, showing that self-reflection does not always lead to insight. If a conversational agent aids in the step from self-reflection to insight, however, those with a high need for self-reflection might be more motivated. Based on these considerations, we hypothesize (1) that as feedback richness expands, users are more motivated to engage with the conversational agent, (2) that this link is mediated by the insight that users gain from the exercise and the feedback, and (3) that the link between feedback richness and motivation is additionally moderated by users' need for self-reflection.

5.2. MATERIALS AND METHODS

We developed a conversational agent for the thought recording task and let participants interact with this agent to collect thought record data and to examine the motivating effect of richer feedback. For the latter objective, we chose a double-blind, between-subjects design. The independent variable, *feedback richness*, was designed to have three levels: acknowledging the reception of user input (low), feedback of low richness plus process-related feedback concerning the amount of input provided (medium), and feedback of medium richness plus content-related feedback, i.e., giving an interpretation of the input with regard to possible underlying schemas (high). As dependent variables, we used the *number of voluntarily completed thought records* in the second session with the conversational agent as well as the *engagement in self-reflection*. In addition, the mediating variable *insight* and the moderating variable *need for self-reflection* were assessed. We obtained ethical approval from the Human Research Ethics Committee of Delft University of Technology (Letter of Approval number: 1600) and pre-registered the study on the Open Science Framework (`https://osf.io/5vucg`).

5.2.1. MATERIALS

The materials, including the informed consent, data management plan, pre- and post-questionnaires, the task instructions, the scenarios, the measures, the power analysis simulation script, as well as all data relevant for the analyses and the dataset of thought records can be found in the data repository accompanying this article [24]).

CONVERSATIONAL AGENT AND SCHEMA-IDENTIFYING ALGORITHM

We developed the conversational agent that engaged participants in the thought record exercise using the chatbot development platform Rasa (version 2.6). The agent received a gender-neutral name (Luca). Luca had a deterministic conversational style that relied on buttons to obtain answers from the user for all interactions except within the thought record and the downward-arrow technique. The thought record form fields encompassed the four core elements of any thought record: what happened that caused the participant distress (situation), how they felt (emotion), what they thought (automatic thought), and what they did in response (behavior). In therapy, when the patient has learned to record their thoughts in this simple format, the form can be extended in various ways [25], for example, with the downward arrow technique. The agent implemented this technique by taking the automatic thought as a starting point and repeatedly asking the same question about the previously stated thought to ultimately arrive at a schema [26]. In line with the technique of reflective listening, the agent gave feedback of varying levels of richness on the delineated thoughts (Figure 5.1). *Low feedback richness* entailed that it thanked the participants for completing the thought record and reminded them that completing more thought records might provide insight into thought patterns. *Medium feedback richness* consisted of the low-level feedback but additionally presented participants with a diagram of the number of downward arrow steps they had completed in this thought record and all previous thought records and put this number in relation to the number of people who had completed as many steps in a previous study. For *high feedback richness*, finally, the medium-level feedback was extended with natural language processing to determine one or multiple schemas that may have been activated. A spider diagram illustrated the degree to which the algorithm deemed the schema(s) present in the thought record using blue dots along nine schema axes. Orange dots in the same diagram depicted the aggregated results from previous thought records of this participant. The schemas for this condition were determined using a set of nine neural networks, one for each possible schema (see Burger *et al.* [27] for details concerning how the networks were trained and tested and Goodfellow *et al.* [28] for details concerning the statistical foundations of recurrent neural networks). Millings *et al.* [29] first identified and described the schemas, which were obtained from a content analysis of thought records collected from a clinical population with depression and/or anxiety. The feedback of medium richness served as a control condition for the feedback of high richness, as it allowed separating the effect of giving feedback on participants' efforts from that of giving feedback that might generate insight.

5

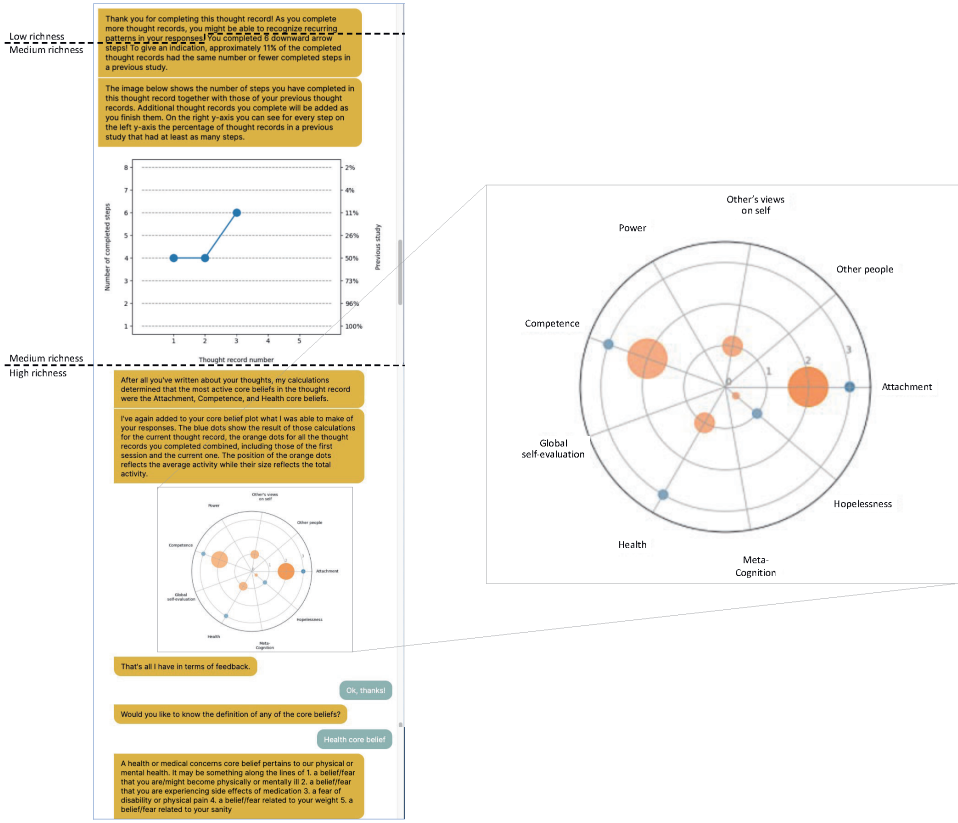

Figure 5.1: Example feedback for the third thought record in the high feedback condition. Participants in this condition saw all three feedback types combined. All participants received the first two sentences (low feedback richness). Participants in the medium feedback condition saw everything up to the end of the first plot, while participants in the high feedback condition saw also what is shown in the second plot and could optionally see the definitions for the schemas. In this thought record, three schemas were equally active and more so than the other ones as determined by the algorithm. They are shown as the blue dots on the spider plot. The activation pattern of schemas across all thought records of this participant is reflected in the size and location of the orange dots in the spider plot. New information was added to the plots after every completed thought record for the feedback of medium and high richness.

SCENARIOS

The agent used a set of ten scenarios to select from for the scenario-based thought records of the first session. These were taken mostly from the Ways of Responding scale [30] with two added from the Cognitive Error Questionnaire [31]. We divided the scenarios into two sets of five scenarios, one with situations that might be difficult on an interpersonal level (e.g., an acquaintance does not wave back at you) and one with situations that might be difficult on an achievement-related level (e.g., you were fired from your new job for not meeting your quota). The agent presented participants with one randomly chosen scenario from each of the two sets.

5.2.2. MEASURES

We used the three subscales of the Self-Reflection and Insight Scale [23] as measures: the *Engagement in Self-Reflection* subscale for self-reported motivation (outcome variable), the *insight* subscale for self-insight participants gain from thought recording with the agent (mediator variable), and the *Need for Self-Reflection* subscale for participants' general need to reflect on their thoughts, emotions, and behaviors (moderator variable). We modified the *Insight* and the *Engagement in Self-Reflection* subscales to measure state rather than trait variables. For example, the item "I am usually aware of my thoughts" (Insight) became "Completing the thought-recording task with the chatbot has made me more aware of my thoughts."

5.2.3. PARTICIPANTS

Participants were recruited from Prolific, a crowd-sourcing platform for research studies. We pre-screened participants on their depression symptoms using the 9-item patient health questionnaire (PHQ-9) [32]. In line with [33], we used the range of $4 < score < 8$ for selecting *subclinical* participants unlikely to meet diagnostic criteria for depression. A clinical population was not chosen for ethical reasons and a healthy population was not chosen because we expected a subclinical population to be more similar to a clinical population in terms of motivational barriers. Participants were not informed of their score or the selection criterion. To participate in the pre-screening, participants had to be at least 18 years of age and fluent speakers of English. We recruited 2899 participants. Participants with subclinical depression symptoms and those who did not fail more than one attention check (519 participants) were invited to participate in the next part of the study. With a power analysis simulation following the bias-corrected bootstrapping method [34] modified for a categorical predictor and a Poisson-distributed outcome variable, we determined that 306 participants would be needed for a medium effect size (in line with [35], at least 13 % of the variance in *feedback richness* estimated to be explained by *insight*, a-path, and at least 13 % of the variance in *motivation* estimated to be explained by *insight* when controlling for *feedback richness*, b-path) at $\alpha = .05$ and power of 80 %. We stopped recruitment after having complete data of 306 participants, but, due to participants still being in the pipeline when recruitment stopped, the final dataset contains the data of 308 participants (143 female, 164 male, 1 other). Their ages ranged from 18 to 75 with $mean_{age} = 30.97$ and $SD_{age} = 11.66$. Participants could be excluded for failing multiple attention checks, failing multiple instruction comprehension questions, not taking the task seriously (writing gibberish, copying and pasting content from other websites, writing incoherent responses to the agent), or technical problems. In total, thirty-six participants had to be excluded for one of these reasons of which only one was excluded for not taking the task seriously.

5.2.4. PROCEDURE

In the first part of the experiment, Prolific redirected participants to the survey tool Qualtrics to complete the *Need for Self-Reflection* scale (pre-questionnaire). Based on the result, Qualtrics divided them into one of three possible buckets (low, medium, and high need for self-reflection). Within a few hours after completing the pre-questionnaire, a message on Prolific invited participants to the next part of the experiment, which

5

consisted of instructions and the first thought recording session with the conversational agent. Participants were blindly assigned to the experimental conditions using the minimization method of randomization [36] with the need for self-reflection buckets as the only variate. The agent started the conversation in the first session with a brief onboarding message. It then repeated the main instructions. Upon presenting the first scenario to the user, it proceeded with the thought record and downward arrow form fields. Finally, it gave feedback depending on the condition and asked if the participant was ready for the next scenario. The first session always consisted of two scenario-based thought records to become familiar with the task and the feedback. Participants received an invitation via Prolific to participate in the second session between 24 and 48 hours after completing the first session. The second session proceeded as the first but with the agent moving directly to the thought record after an initial “welcome back" exchange. In the second session, the agent asked participants to complete at least one but as many additional thought records as they wanted. For the thought records of this session, they were taking day-to-day situations from their own lives. We compensated each session with 2 GBP based on an estimated completion time of 20 minutes. No extra monetary compensation was provided for more completed thought records to not interfere with motivation. Participants were informed at the beginning of the second chat session of the expected completion time for the post-questionnaire, which included the *Insight* and *Engagement in Self-Reflection* scales as well as any additional comments or feedback. In total, participants could receive 4.3 GBP for completing all parts of the study.

5.2.5. DATA AND ANALYSIS METHOD

To determine feasibility, we correlated the nine values of the frequency distribution over the schemas as assigned by the algorithm on this dataset with the distribution of two previously collected datasets [27, 29]. To this end, we recoded the labels for each utterance from ordinal (0–schema not present to 3–schema clearly present) to binomial (0–schema not present and 1–schema at least a little bit present). The same procedure was followed for the dataset of Burger *et al.* [27]. For the Millings *et al.* [29] dataset, however, we compare with the frequencies reported in the article, which are based on entire thought records rather than utterances and which were manually assigned.

Two independent coders (one male and one female computer science student) labeled all thought records of the second session using the DIAMONDS framework for psychologically relevant situation characteristics [37] to examine the content of participants' thought records in the second session. They were trained on ten example thought records in a joint session of one hour to clarify the definitions of the DIAMONDS. Since participants were asked to report only situations that caused a negative emotion, we dropped the Positivity characteristic. Two further labeling categories were added: *COVID19-related* and *situation type (achievement-related vs. interpersonal)*. All labels were binomial (situation *has* or *does not have* characteristic). Coders were instructed together and coded 10 example situations (taken from the first chat session) together with the first author before coding the situations described by participants in the second chat session independently. While interrater agreement was mixed on the DIAMONDS, ranging from minimal $\kappa = 0.25$ (Negativity) via moderate $\kappa = 0.66$ (Interpersonal) to strong $\kappa = 0.83$ (Mating), the raters largely agreed on the frequency of labels within the dataset (Pearson $r = 0.89$ based on 10 values).

The reliability of the three subscales of the Self-Reflection and Insight scale was good (*Need for Self-Reflection*: Crohnbach's $\alpha = 0.85$ with $[0.85, 0.86]$ 95 % CI) and acceptable (*Insight*: $\alpha = 0.76$ with $[0.75, 0.77]$ 95 % CI and *Engagement in Self-Reflection*: $\alpha = 0.75$ with $[0.73, 0.76]$ 95 % CI). The items of each subscale were summed to obtain a summary score for the variables *need for self-reflection*, *insight*, and *engagement in self-reflection*. Engagement in self-reflection was negatively skewed (ceiling effect) and we consequently boxcox transformed the data with $\lambda = 1.97$ for use in the analyses.

We followed the Baron *et al.* [38] method to test for the mediated effect. For the direct effect, this entailed fitting a generalized linear model with a log-link function as the behavioral outcome variable *number of voluntarily completed thought records* was expected to be Poisson-distributed and fitting a second linear model for the self-report outcome variable *engagement in self-reflection*. A further linear model was fit to test whether *feedback richness* affected *insight* (mediator). Finally, we fit one generalized linear model and one linear model to assess the effect of the mediator on each of the two outcome variables. Due to the lack of mediation observed in these models, we did not test for moderated mediation. However, we checked with two linear regression models whether participants' *need for self-reflection* (moderator) moderated the direct link between the *feedback richness* and either of two outcome variables.

5.3. RESULTS

All 308 participants were able to complete thought records with the conversational agent. Of the 93 participants who chose to comment, 34 reported that they found the experiment insightful, with five participants specifically mentioning the added value of the agent and the feedback ("The chatbot makes the experience more friendly," "shows that chatbot can offer a sincere alternative to human response," "[...] get immediate feedback than on a paper which feels sometimes too much like homework," "[...] I felt like someone was paying attention to me," "this chatbot is really helpful in discovering my thought patterns"). However, another five participants also commented that they struggled with the downward arrow technique and would have liked more agent or even human support ("I know it's a chatbot, but I wish Luca could engage a little more when trying to work your way down the arrow," "it was really hard to go down the thought steps instead of in circles, i feel like maybe a human would've been able to help with that," "I also would prefer to do this activity with a real person rather than a chatbot," "It is not always easy to figure out what the next drill down should be," "Was somewhat confused to break my thought patterns down in the arrow scheme though"). Only two participants remarked negatively about the rich feedback ("I think Luca's overall assessment of my core beliefs was decent, but not perfect" and "I found the circular diagram a bit difficult to understand") and one in the low-level feedback condition about the lack thereof ("I did not see any feedback from the chatbot, it would be nice to.").

The relative frequency distribution with which the schema-labeling algorithm identified certain schemas in this dataset compared to that of the previous study by Burger *et al.* [27] correlated highly for both the scenario-based (Spearman's $\rho = 0.93$) and the personal thought records (Spearman's $\rho = 0.95$). The schema frequency distribution of both scenario-based and personal thought records taken together also correlated positively (Spearman's $\rho = 0.57$) with that reported by Millings *et al.* [29] (Figure 5.2). Across all three datasets, the most frequently occurring schemas were the Attachment, Competence, and Global Self-Evaluation schemas.

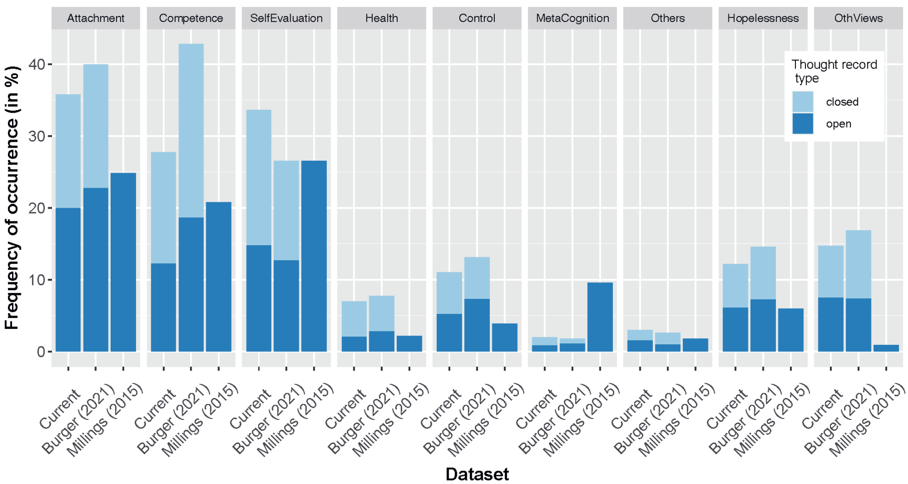

Figure 5.2: Frequency of occurrence of schemas in this dataset compared to a previously collected dataset [27], in which participants completed thought records in survey format, and that of Millings *et al.* [29]. In the current dataset and the one collected by Burger *et al.* [27] schemas are identified by an algorithm from thoughts (automatic thought or any downward arrow step), while in the dataset by Millings *et al.* [29], schemas were identified by the authors and from entire thought records. While the algorithm assigns ordinal codes corresponding to the degree to which a schema was present, for the purpose of this analysis, we recoded these scores to binomial scores with all values above 0 being coded as 1. *Closed* thought records are those based on scripted scenarios while *open* thought records are those in which participants report on situations from their lives.

We present a typical thought record situation for each content label in Table 5.1. Besides reporting mostly negative situations (98 % Negativity), participants opted for more interpersonal and social than achievement-related or intellectual situations.

Participants felt engaged in self-reflection when completing the thought records ($mean = 29.56$, $SD = 4.04$), but completed, on average, only 1.62 ($SD = 0.72$) thought records in the second session. The direct effect of *feedback richness* on either of these measures of motivation (Figure 5.3) was not observed. There was also no effect found for the *feedback richness* on the mediator variable *insight* (a-path). As a consequence, partial mediation was no longer relevant. Nonetheless, a significant link between the mediating variable *insight* was found for both measures of motivation (b-path): for every additional scale point of insight they report, participants complete 1.03 times as many thought records ($b = 0.03$, $z(304) = 2.12$, $p = 0.03$) and feel 4.91 scale points more engaged ($b = 11.14$, $t(304) = 11.30$, $p < 0.001$) on a scale ranging from 6 to 36.

The moderator *need for self-reflection* had no effect on the direct link between *feedback*

Table 5.1: Example thought record situations for each content label. The column *IRR* shows the InterRater Reliability while the column *MRF* shows the Mean Rater Frequency, i.e. the mean of how frequently the raters found a specific label to occur in the dataset. Labels were not mutually exclusive.

Label	**Explanation**	**IRR (κ)**	**MRF (%)**	**Example situation**
Achievement-related	Situations in which self-esteem is at risk because it is possible to perform poorly.	0.58	19	When I didn't get a job I was interviewed for.
Interpersonal	Social situations that can affect one's self-worth.	0.66	60	My colleague blamed me for their mistake.
COVID-related	Thought records in which participants mention COVID-19.	0.83	4	Staying indoors a lot due to the pandemic.
Duty	Situations that require executing a task conscientiously or dutifully.	0.43	27	I had to give a presentation.
Intellect	Situations that are cognitively stimulating.	0.46	19	I was worried about sitting an exam for university.
Adversity	Situations in which one is criticized, blamed, or dominated.	0.58	21	I was really sick and my then-boss made me work while I was sick.
Mating	Situations that involve potential or actual romantic partners.	0.83	19	My husband is stressed and moody because of it.
Negativity	Situations that are anxiety-inducing, stressful, frustrating, upsetting.	0.25	98	I rejected a holiday job offer because it paid too little and now I cannot find anything else.
Deception	Situations that can result in feelings of hostility due to deception or sabotage.	0.35	13	I found out I was being cheated on by my girlfriend.
Sociality	Situations that involve social interaction.	0.32	48	I was given a huge amount of rudeness and grief by a customer at work.

5

richness and either of the two motivation measures. Additionally, participants' need for self-reflection did not predict how many thought records they would do voluntarily. It did, however, explain their engagement in the task with participants feeling 4.42 scale points more engaged with every additional scale point of their self-reported need for self-reflection ($b = 8.97$, $t(302) = 3.69$, $p < 0.001$).

5.4. DISCUSSION

The findings show that thought recording with a conversational agent is feasible for a subclinically depressed population: 100 % of participants completed the thought records such that the machine learning algorithm trained on a similar dataset could label thoughts with regard to the underlying schemas. The distribution of the thus assigned schema labels not only closely resembles that of the dataset used for training the algorithm (healthy population) but also the manually labeled dataset of Millings *et al.* [29] (clinical population). In addition, participants frequently reported enjoying the experiment and finding it so valuable that they intended to continue using the technique in

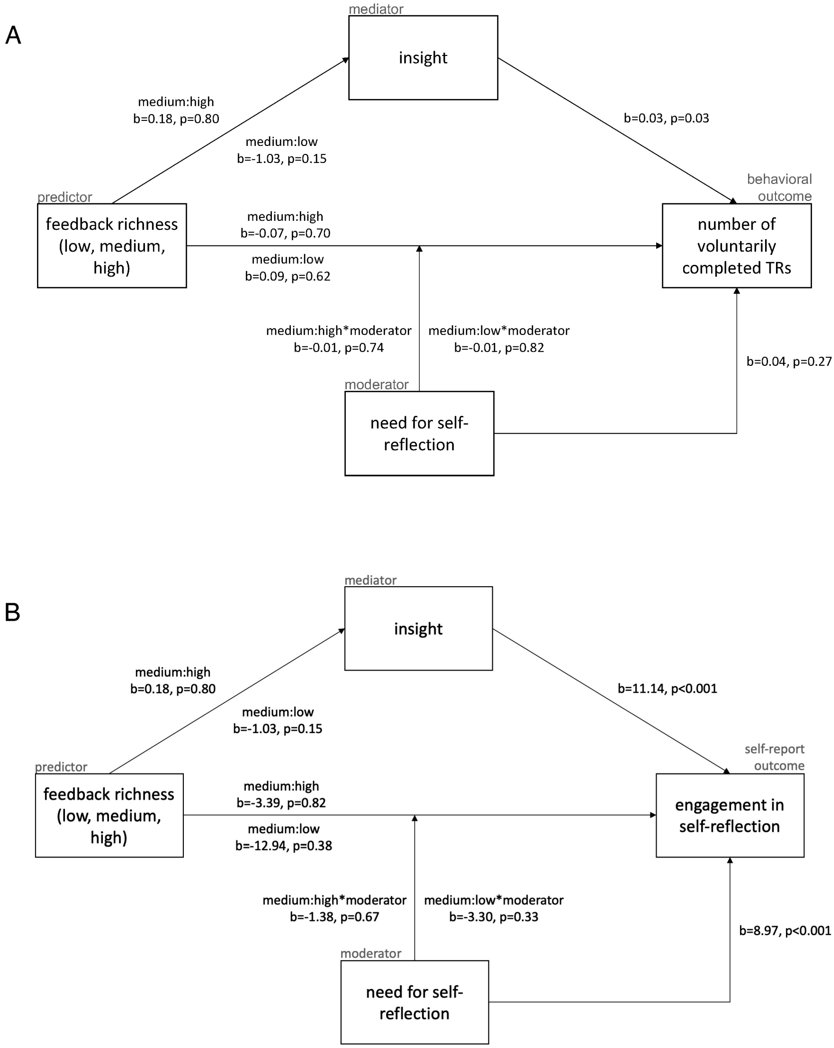

Figure 5.3: Results of all paths of the mediation and moderation analyses, with the behavioral outcome measure of motivation, number of voluntarily completed thought records in the second chat session, shown in **(A)** and the self-reported one, boxcox-tranformed ($\lambda = 1.97$) engagement in self-reflection, shown in **(B)**.

their day-to-day lives. Prior studies looking into the feasibility of using conversational agents for mental health interventions have found similar results concerning user satisfaction and ability to interact with the conversational agents [39, 40], but none had specifically studied thought record completion before. In terms of content, participants' personal thought records concerned interpersonal (58 %) or social situations approximately three times more often than achievement-related (19 %) or intellect-related ones, which is also reflected in the schemas, with the *Attachment* schema being identified by the algorithm more than the *Competence* schema. Despite around 4 % of situations mentioning the COVID-19 pandemic, the *Health* schema was not more active in this dataset than in the ones collected before the pandemic. This is likely due to the training dataset of the algorithm being biased towards dieting situations for this particular schema. It can also be seen from the frequency of the *Negativity* label that participants were able to choose negative situations (98 %) as instructed but sometimes put a positive spin on the meaning of the situation for themselves (e.g. "It says that I don't have to feel obliged to do anything for anyone, and I don't want to feel that way."). It is important to note here, however, that the participants were subclinically depressed and these findings concerning the feasibility and content of the thought records may not generalize to a clinical population.

We did not observe the hypothesized effect of feedback on motivation: the results did not show that the *feedback richness* influenced either the *motivation* of participants (direct effect) or the hypothesized mediating variable *insight participants gained from the task*. However, participants' gained insight positively related to both measures of motivation, and participants who reported a greater need for self-reflection also reported being more engaged in the task. When regarding these findings, limitations of the feedback on the one hand and of motivation on the other should be considered. For one, the spider plot and the academic definitions of the schemas in the rich-feedback condition might not have been as accessible as we had hoped and therefore did not add the expected value. This could be addressed in future research by following an cyclic design approach including both end users and graphical designers, simplifying the feedback, including measures of graphic literacy and health literacy as moderators, or conducting pilot studies to determine whether feedback is processed as desired. In line with this, articles concerning the design of graphical feedback in behavior change support systems argue for the importance of health literacy and usability as guiding principles [41, 42], and platforms that have successfully used complex informational feedback in graphical format have done so in collaboration with a design company and with an iterative refinement process [43]. Another possible limitation of the feedback is that participants may have perceived the richer feedback as discrepant with the otherwise limited conversational capabilities of the agent. As far as motivation is concerned, this was measured with just one session, such that small issues like participants misclicking, minor technical glitches, or external disturbances may have played a larger role than in a long-term study. Additionally, motivation may also have been adversely affected by the monetary compensation in the online context and may have panned out differently with patients being internally motivated by a desire to get healthy. Lastly, our participant sample included more males than females, which is noteworthy due to depression being more prevalent in women. Future research might therefore consider looking into a moderating effect of gender. And, when looking more closely at the distribution of schemas in the different populations (5.2), the clinical samples differs most markedly with respect to the *Global self-evaluation*, the *Meta-Cognition*, and the *Other's views on self* schemas to the non-clinical samples. Since self-evaluation and meta-cognition are likely to also be linked to one's need for self-reflection and one's engagement in self-reflection, it is possible that the results would play out differently in a clinical sample. Since the experiment was not underpowered, however, and some limitations pertain to all three conditions, we conclude from the null results that this type of feedback richness is unlikely to have a large effect on motivation regardless of the limitations.

In summary, people with subclinical depression symptoms are capable of thought recording with a conversational agent. Not only were the thoughts they recorded of sufficient richness to allow for automatic schema identification, but the three most frequently occurring schemas (Attachment, Competence, and Global Self-evaluation) in this sample of subclinically depressed participants were the same as in previous work with healthy [27] and with clinical [29] populations. However, no support could be found that richer feedback leads to a higher motivation to engage with thought recording. More research and perhaps participatory design are needed to determine engagement strategies for the agent that can lead to greater adherence. One possible route to explore is to combine the content-based feedback generated by reflective listening with additional communication strategies of motivational interviewing, such as establishing rapport or eliciting self-motivational messages [44]. Finally, the study could be repeated with a clinical population to determine the role that other (de-)motivational forces, such as dampened enjoyment of tasks and the wish to get healthy, play in this population. We contribute the dataset of collected thought records, all measures [24] and the conversational agent code[2] for researchers and developers interested in working with this data.

[2] `https://github.com/fvburger/ThoughtRecordingChatbot`

Bibliography

1. Burger, F., Neerincx, M. A. & Brinkman, W.-P. Using a conversational agent for thought recording as a cognitive therapy task: Feasibility, content, and feedback. *Frontiers in Digital Health* **4** (2022).
2. Richards, D. & Richardson, T. Computer-based psychological treatments for depression: A systematic review and meta-analysis. *Clinical Psychology Review* **32,** 329–342 (2012).
3. Russo, T. Cognitive counseling for health care compliance. *Journal of rational emotive therapy* **5,** 125–134 (1987).
4. Detweiler, J. B. & Whisman, M. A. The role of homework assignments in cognitive therapy for depression: Potential methods for enhancing adherence. *Clinical Psychology: Science and Practice* **6,** 267–282 (1999).
5. Helbig, S. & Fehm, L. Problems with homework in CBT: Rare exception or rather frequent? *Behavioural and cognitive psychotherapy* **32,** 291 (2004).
6. Kazantzis, N. & Shinkfield, G. Conceptualizing patient barriers to nonadherence with homework assignments. *Cognitive and Behavioral Practice* **14,** 317–324 (2007).
7. Kazantzis, N., Brownfield, N. R., Mosely, L., Usatoff, A. S. & Flighty, A. J. Homework in cognitive behavioral therapy: A systematic review of adherence assessment in anxiety and depression (2011–2016). *Psychiatric Clinics* **40,** 625–639 (2017).
8. Jia, H., Zack, M. M., Thompson, W. W., Crosby, A. E. & Gottesman, I. I. Impact of depression on quality-adjusted life expectancy (QALE) directly as well as indirectly through suicide. *Social psychiatry and psychiatric epidemiology* **50,** 939–949 (2015).
9. Marcus, M., Yasamy, M. T., van Ommeren, M., Chisholm, D. & Saxena, S. *Depression: A global public health concern.* World Health Organization. (2012).
10. Collins, K. A., Westra, H. A., Dozois, D. J. & Burns, D. D. Gaps in accessing treatment for anxiety and depression: challenges for the delivery of care. *Clinical psychology review* **24,** 583–616 (2004).
11. Kohn, R., Saxena, S., Levav, I. & Saraceno, B. The treatment gap in mental health care. *Bulletin of the World Health Organization* **82,** 858–866 (2004).
12. Wind, T. R., Rijkeboer, M., Andersson, G. & Riper, H. The COVID-19 pandemic: The 'black swan' for mental health care and a turning point for e-health. *Internet interventions* **20** (2020).
13. Burger, F., Neerincx, M. A. & Brinkman, W.-P. Technological State of the Art of Electronic Mental Health Interventions for Major Depressive Disorder: Systematic Literature Review. *Journal of Medical Internet Research* **22,** e12599 (2020).
14. Leahy, R. L. The therapeutic relationship in cognitive-behavioral therapy. *Behavioural and Cognitive Psychotherapy* **36,** 769–777 (2008).

15. Rogers, J. & Maini, A. *Coaching for Health: Why it works and how to do it* (McGraw-Hill Education, 2016).
16. Provoost, S., Ruwaard, J., Neijenhuijs, K., Bosse, T. & Riper, H. Mood mirroring with an embodied virtual agent: a pilot study on the relationship between personalized visual feedback and adherence. in *Proceedings of the 2018 International Conference on Practical Applications of Agents and Multi-Agent Systems* (Springer, 2018), 24–35.
17. Tielman, M. L., Neerincx, M. A. & Brinkman, W.-P. Design and evaluation of personalized motivational messages by a virtual agent that assists in post-traumatic stress disorder therapy. *Journal of medical Internet research* **21,** e9240 (2019).
18. Fitzpatrick, K. K., Darcy, A. & Vierhile, M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. *JMIR Mental Health* **4** (2017).
19. Beck, A. T. & Alford, B. A. *Depression: Causes and treatment* (University of Pennsylvania Press, 2009).
20. Gillham, J. E., Reivich, K. J., Freres, D. R., Chaplin, T. M., Shatté, A. J., Samuels, B., Elkon, A. G., Litzinger, S., Lascher, M., Gallop, R., *et al.* School-based prevention of depressive symptoms: A randomized controlled study of the effectiveness and specificity of the Penn Resiliency Program. *Journal of consulting and clinical psychology* **75,** 9 (2007).
21. Kazantzis, N. & Dattilio, F. M. Definitions of homework, types of homework, and ratings of the importance of homework among psychologists with cognitive behavior therapy and psychoanalytic theoretical orientations. *Journal of Clinical Psychology* **66,** 758–773 (2010).
22. Dobson, K. S. A Commentary on the Science and Practice of Homework in Cognitive Behavioral Therapy. *Cognitive Therapy and Research* **45,** 303–309 (2021).
23. Grant, A. M., Franklin, J. & Langford, P. The self-reflection and insight scale: A new measure of private self-consciousness. *Social Behavior and Personality: an international journal* **30,** 821–835 (2002).
24. Burger, F., Neerincx, M. A. & Brinkman, W.-P. *Dataset and Analyses for Using a conversational agent for thought recording as a cognitive therapy task: feasibility, content, and feedback* 4TU Research Data Repository `https://doi.org/10.4121/20137736`. 2022.
25. Wenzel, A. Basic strategies of cognitive behavioral therapy. *Psychiatric Clinics* **40,** 597–609 (2017).
26. Burns, D. D. *The Feeling Good Handbook* (Plume, 1999).
27. Burger, F., Neerincx, M. A. & Brinkman, W. P. Natural language processing for cognitive therapy: Extracting schemas from thought records. *PLoS ONE* **16** (2021).
28. Goodfellow, I., Bengio, Y. & Courville, A. *Deep Learning* (MIT Press, 2016).

29. Millings, A. & Carnelley, K. B. Core belief content examined in a large sample of patients using online cognitive behaviour therapy. *Journal of Affective Disorders* **186,** 275–283 (2015).

30. Barber, J. P. & Derubeis, R. J. The ways of responding: A scale to assess compensatory skills taught in cognitive therapy. *Behavioral Assessment* **14,** 93–115 (1992).

31. Lefebvre, M. F. Cognitive distortion and cognitive errors in depressed psychiatric and low back pain patients. *Journal of Consulting and Clinical Psychology* **49,** 517–525 (1981).

32. Kroenke, K., Spitzer, R. L. & Williams, J. B. The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine* **16,** 606–613 (2001).

33. Manea, L., Gilbody, S. & McMillan, D. Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis. *CMAJ* **184,** E191–E196 (2012).

34. Fritz, M. S. & MacKinnon, D. P. Required sample size to detect the mediated effect. *Psychological science* **18,** 233–239 (2007).

35. Cohen, J. *Statistical power analysis for the behavioral sciences* (Routledge, 2013).

36. Taves, D. R. Minimization: a new method of assigning patients to treatment and control groups. *Clinical Pharmacology & Therapeutics* **15,** 443–453 (1974).

37. Rauthmann, J. F., Gallardo-Pujol, D., Guillaume, E. M., Todd, E., Nave, C. S., Sherman, R. A., Ziegler, M., Jones, A. B. & Funder, D. C. The Situational Eight DIAMONDS: A taxonomy of major dimensions of situation characteristics. *Journal of Personality and Social Psychology* **107,** 677 (2014).

38. Baron, R. M. & Kenny, D. A. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology* **51,** 1173–1182 (1986).

39. Gaffney, H., Mansell, W. & Tai, S. Conversational agents in the treatment of mental health problems: mixed-method systematic review. *JMIR mental health* **6,** e14166 (2019).

40. Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S. & Torous, J. B. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. *The Canadian Journal of Psychiatry* **64,** 456–464 (2019).

41. Broderick, J., Devine, T., Langhans, E., Lemerise, A. J., Lier, S. & Harris, L. *Designing health literate mobile apps.* Institute of Medicine, Washington, DC. (2014).

42. Thomson, C., Nash, J. & Maeder, A. Persuasive design for behaviour change apps: Issues for designers. in *Proceedings of the Annual Conference of the South African Institute of Computer Scientists and Information Technologists* **43** (Association for Computing Machinery, 2016).

43. Western, M. J., Peacock, O. J., Stathi, A. & Thompson, D. The understanding and interpretation of innovative technology-enabled multidimensional physical activity feedback in patients at risk of future chronic disease. *PloS one* **10,** e0126156 (2015).

44. Hall, K., Gibbie, T. & Lubman, D. I. Motivational interviewing techniques: facilitating behaviour change in the general practice setting. *Australian family physician* **41,** 660–667 (2012).

6

CONCLUSION

This thesis investigated how artificial intelligence (AI) can guide people in e-mental health interventions for depression. Each chapter presented an important step towards answering this main research question. We conducted a literature review to analyze the technological state of the art of e-mental health for depression (Ch. 3). To this end, a measure to assess technological sophistication of functions and systems had to be created (Ch. 2). The main purpose of the literature review in the context of this thesis was to identify opportunities for AI to augment e-mental health interventions. One such opportunity was that of natural language processing (NLP). As reviewed in the introduction (Ch. 1), several additional considerations also played a role in choosing this direction for the second part of the thesis research: (1) processing natural language and conveying understanding is a key skill of therapists, which is relevant in light of findings that more human involvement in e-mental health for depression was linked to higher effectiveness and lower attrition, (2) the techniques of cognitive therapy, one of the most widely used approaches to psychotherapy for depression, critically hinge on natural language, (3) improving the natural language capabilities was identified as a key research challenge for conversational agents for mental health, and (4) the advances made in AI in the past decade, especially in deep learning, have been shown to be very powerful in natural language processing in various domains including mental health. We consequently designed a conversational agent capable of processing the thoughts that people delineate in a thought record, for which an NLP model was needed (Ch. 4). In evaluating this conversational agent, we tested whether the dialog format presented a feasible means to collecting thought record data (Ch. 5). Here, the main conclusions are presented. Additionally, this chapter discusses the limitations of the work, points out directions for future research, and states the scientific and practical contributions made. It ends with a reflection.

6.1. CONCLUSIONS

Two research questions and two hypotheses were introduced in Chapter 1 as important steps towards generating knowledge that can answer the main research question of how artificial intelligence can guide people in e-mental health interventions for depression. A number of key conclusions can be drawn from Chapters 2-5 that concern themselves with these questions and hypotheses.

6.1.1. ASSESSMENT OF TECHNOLOGICAL SOPHISTICATION WITH THE EHDTS SCALES

> RQ1: How can the technological sophistication of a software system for the prevention and treatment of major depressive disorder be assessed?

To answer the question of how to assess the technological sophistication of a software system for depression, we proposed to use a summary score of the technological sophistication scores of each constituent of the system that is delivering functionality. This requires that systems be decomposed into their functional components or *functions*. To this end, we used the system descriptions provided in the research articles, which largely concerned functionality on what is called the *application layer* in the Microsoft Application Architecture Guide [1]. Two types of functions could be distinguished: *intervention functions* and *support functions*. In leaning on the Persuasive System Design framework [2], we divided the latter into four subtypes: *planning support*, *execution support*, *monitoring support*, and *social support*. Taking as a starting point the notion that a system is more technologically sophisticated if it can respond more intelligently to the user and is more interactive, we developed a scale with five degrees to determine the technological sophistication of a function. These five degrees of sophistication ranged from human execution of the functionality (degree 0) via provision of information (degree 1), allowing for user input (degree 2), and processing high-level features of user input (degree 3) to processing the content of user input (degree 4). However, in this form, the scale was not readily applicable to all five function types. Consequently, the general definitions of each degree were adapted to better suit the specific function type. For example, for the function type *social support*, all five degrees include humans carrying out functionality, because social support is provided by other humans. The adaptations that were made in the degree definitions for each of the five function types resulted in a set of five scales: the e-mental Health Degree of Technological Sophistication (eHDTS) scales with one scale for each function type.

In a first reliability study, we tested whether two independent coders can identify the function type and apply the corresponding scales when given a mixed set of functions of all five function types. The two degrees assigned by the two coders correlated moderately with each other. In a second reliability study, two independent coders were given a set of functions from only one function type and the corresponding scale to assess the technological sophistication of this function type. There was a strong correlation between the degrees assigned by the two independent coders for all scales in this task with the exception of the eHDTS scale for *monitoring support* functions, for which the correlation was moderate. The technological sophistication of functions could thus be reliably coded. Finally, in a third study, coders were given a set of function descriptions and asked to assign a degree of technological sophistication from 0 to 4 to these descriptions without knowledge of the definitions of the degree scores (naïve). In a later step they were asked to assign a degree with the definitions to the same set of functions (informed). For four scales (all but the eHDTS scale for *monitoring support* functions, for which the correlation was weak), moderate to strong correlations were observed between the naïve and informed codings. We thus demonstrated concurrent validity for four of the five eHDTS scales, showing that the scales largely correspond to the intuitive understanding of *technological sophistication*. In conclusion, these findings indicate that the eHDTS scales may be a reliable and valid means of assessing the technological sophistication of functions of e-mental health systems for depression.

6.1.2. TECHNOLOGICAL STATE OF THE ART OF E-MENTAL HEALTH FOR DEPRESSION

RQ2: What is the technological state of the art of software systems for the prevention and treatment of major depressive disorder?

We answered the second research question with a systematic literature review on the technological sophistication of systems and the developments over time. Overall, we found that systems were lacking in technological sophistication, with approximately 80 % of systems scoring at or below level 2 on the eHDTS (average of eHDTS scores over all functions of a system). This means that the majority of systems reported in the literature between 2000 and 2017 are of a psychoeducational nature: they have functions that allow for user input, but only a few (if any) functions that process the user input and respond to it. Additionally, autonomous systems were neither more technologically sophisticated nor did they provide more functionality than their human-guided counterparts. It therefore seems that neither of these options are used to compensate for human support in autonomous systems. This is an important finding in light of the frequently reported low adherence rates to e-mental health as it invites the question of whether systems are perhaps just not engaging enough to sustain usage. While developers may not see their systems as competing with all other systems on the software marketplace due to addressing a very specific target group, the technological standards that users have as a result of using other software on a daily basis might need to be taken into consideration nonetheless. No increase in technological sophistication was observed despite substantial growth in the number of systems developed yearly in the covered time period. However, we also found that across all functions, there was always at least one system that implemented the function in a very technologically sophisticated manner. Moreover, some of the newly developed systems do take different approaches to the prototypical educational website system. For example, Woebot [3] is an autonomous chatbot that delivers entirely conversation-based treatment, meeting the users where they are at, namely on the social media website Facebook. The Panoply platform [4] uses crowd-sourcing to obtain relatively instant reappraisal support from peers and to practice reappraising by giving such support to peers. And, as a final example, the only virtual reality based system [5] for depression treatment included in the review, reflects users own consoling words back to them using perspective-taking. In conclusion, most e-mental health for depression systems reported in the literature are low-tech implementations that barely react to the input they receive from users but some promising steps are being taken towards more interactive solutions. Such solutions will be necessary to determine the adherence-enhancing capabilities of more technological sophistication in long-term evaluations.

6.1.3. RECURRENT NEURAL NETWORKS CAN IDENTIFY SCHEMAS FROM THOUGHTS.

> H1: Schemas can be automatically extracted from the thoughts people delineate in thought records including the downward arrow technique.

To test the hypothesis that it is possible to automatically label thoughts with respect to their underlying schemas we collected a dataset of thought records including the downward arrow technique. The thoughts noted down in the thought records of participants were manually labeled by a human coder. This labeling entailed that each thought was assigned nine scores for each of the nine schemas. Scores ranged from 0-3 and denoted the degree to which the thought corresponded with the schema. Three types of machine learning algorithms were trained to recognize the schemas underlying the thoughts: k-nearest neighbors, support vector machines, and recurrent neural networks. With all three algorithm types and for all schemas, we found evidence in support of our hypothesis. The correlation between the schema scores assigned by the algorithms and those assigned by a human ranged from $\rho = 0.11$ to $\rho = 0.76$. All algorithms were able to assign a schema score to the thoughts for the majority of schemas. Two data-related factors played a decisive role in how well a schema could be scored: (1) the amount of thoughts exhibiting a certain schema that were available for training, and (2) the idiosyncratic language use for a schema in this dataset. For example, the *Health* schema could be scored well despite comparatively few samples because it had very idiosyncratic language due to a weight loss scenario that formed the basis for a large number of thought records. A set of nine recurrent neural networks, one for each of the nine possible schemas, outperformed the other algorithms on six schemas, indicating that it is beneficial for this machine learning task to take the word order within thoughts into account. Two schemas benefited from a single recurrent neural network algorithm capable of scoring thoughts with respect to all schemas simultaneously. Such a model has information about the scores of all schemas and can therefore take simple and complex correlations between schemas into account. In conclusion, the evidence supported the hypothesis that the identification of schemas from thoughts is possible with a computer algorithm.

Possibly due to the participants completing only few steps of the downward arrow, however, we did not find that the RNNs were better able to label thoughts that were farther along the arrow. This was not expected, since the downward arrow technique is intended to arrive at a schema formulation and those should be the most unambiguously formulated thoughts with regard to the schema they reflect and should thus be most easy to label correctly. Additionally, we found that within a participant, similar situations activated similar schemas, i.e., the schemas of scenario-based thought records could predict those of real-life situations if the scenario and the real-life situation matched in situation type (interpersonal or achievement-related).

6.1.4. THOUGHT RECORDS CAN BE COMPLETED IN DIALOG WITH A CONVERSATIONAL AGENT.

> H2: An intelligent conversational agent is a feasible tool for completing thought records in dialog format.

We sought evidence for the hypothesis that it is possible for people to complete thought records in dialog with a computer program by developing a conversational agent and observing how people interact with it. The agent was capable of posing questions corresponding to the thought record open text entry fields. Participants with subclinical symptoms of depression recruited on the research crowd-sourcing platform Prolific interacted with the agent in two sessions. All participants were able to complete thought records in this setting. The thoughts that were noted down by the participants could largely be labeled with regard to the underlying schemas by the recurrent neural networks trained on the previously collected dataset. The frequencies with which schemas occurred in this dataset (subclinical population) were similar to both those of the dataset used for training the algorithms (healthy population) as well as those of Millings *et al.* [6] (clinical population). Of the participants who chose to give a comment (93 out of 308), five remarked that they would have appreciated more guidance from the conversational agent on the downward arrow technique. Many, however, stated that they found the experiment insightful and intended to continue applying the learned technique in their day-to-day lives. In conclusion, these findings substantiate the hypothesis that thought recording can be done successfully in dialog with a conversational agent.

6.2. LIMITATIONS

Some general limitations play a role in the research conducted for this thesis and should be taken into account when considering the main conclusions. The first limitation pertains to the choices we made in realizing the eHDTS scales and the conversational agent. For the eHDTS scales, we chose to focus on the interactivity of functions and their ability to process user input in explicating technological sophistication. However, one could also take a different starting position, such as the one that a system is particularly technologically sophisticated if it applies good data security protocols. This perspective may have resulted in a very different conclusion concerning the technological state of the art of systems. As another example, some results are based on summary scores, such as the eHDTS score of a system. These carry with them the issue that choosing a different method of summarizing may result in different conclusions. For the ranking of systems, we therefore reported results for two summary scores (the average and a weighted average over functions), but many more are conceivable. Similarly, for assigning a schema score to thought records in generating the feedback of the agent, we chose to assign the highest score that was obtained across all thoughts contained in a thought record for any given schema. How well the scores of this method and other summary methods match with those that a therapist would assign is an interesting question for further research. Finally, alternative routes to supporting thought recording with natural language processing are also possible. For example, these could have been more process-focused and used NLP to detect when and how users are struggling with formulating a good automatic thought, when they need more guidance in taking the next downward arrow step, or when they have arrived at a well-written core belief. As a result of the choices made along the way, we do not know how our conclusions generalize to other routes and we therefore only claim that this thesis provides *an* answer to the main research question but not *the* answer.

A second limitation concerns the voluntary parts to the thought recording task and using such tasks in a crowd-sourcing setting, as we did for testing the two hypotheses H1 and H2. Crowd-sourcing comes with its own suite of issues concerning the successful collection of high-quality data [7], one of which is the source of motivation of participants. In a study from 2015, Litman *et al.* [8] found that both US and India-based workers on Amazon Mechanical Turk reported being mostly motivated by the monetary compensation. In order to prevent recruits from cutting corners to maximize payment, researchers have devised various quality-improving methods, such as attention checks and comprehension checks. While we took such measures where possible, two key parts to the thought recording task in our experiment were voluntary: the number of downward arrow steps taken to arrive at a core belief in both experiments and the number of thought records participants complete in the second session of the experiment to test H2. Completing each of these voluntary parts to the task takes a considerable amount of time and not completing them or completing only the minimum therefore could significantly increase participants' hourly pay rate. Future research may investigate how thought records collected in our setting compare to those collected from an intrinsically motivated sample.

This connects to the more general limitation that despite developing for a depressed population, none of the data to test H1 and H2 was collected with a clinically depressed participant sample. While this was motivated by ethical reasons, it is important to note that maladaptive schemas do not necessarily lead to mental illness and, conversely, those who are healthy can hold maladaptive schemas as well. For example, studies aiming to identify a second-order schema structure have been conducted using a healthy population rather than a clinical one (e.g., [9]). The frequency with which schemas occurred in the thought records that we collected with the healthy sample then also resembled those of a clinical population [10]. In line with the common practice of conducting clinical research in phases to ensure that an intervention is safe before moving to a greater dose or more risky population [11], the successful collection of thought records with a healthy population allowed us to move to a subclinically depressed population for the second study. In this second study, we attempted to increase motivation with the feedback of the conversational agent (Ch. 5). It was important to get as close as possible to a depressed population, as the motivation threshold is particularly high in depressed people [12]. The machine learning algorithm was as capable of labeling thoughts in the thought records of the subclinical population as it was in those of the healthy population that it was trained on. Again, the frequency distribution of schemas in the subclinical sample resembled that of the healthy and that of the clinical population of Millings *et al.* [6]. These aspects of the data give some indication that our conclusions might readily generalize to a clinical population.

6.3. FUTURE WORK

Aside from the more narrow directions for further research that directly follow from the limitations and that are already mentioned above, additional interesting avenues are outlined here. An issue in coding the systems for depression prevention and treatment was the lack of a reporting standard for system functionality. Since many systems are not maintained after the research finishes and can therefore no longer be *experienced*, much information is lost in brief module descriptions. While we have attempted to open the "black boxes" [13] that these systems are, this is limited by what is described and how it is interpreted. This connects to the larger problem of reproducibility in the case of research conducted with computer systems. The system presents a specific solution to a problem and if neither the system nor a standardized means of reporting on its functionality are available after research has finished, it becomes difficult or even impossible to reproduce the findings. As a result of the reproducibility crisis in psychology [14] and science more generally [15], many approaches to improving research transparency are currently being developed and tested by scientific communities (e.g., data repositories, source code repositories, research pre-registrations). In human-agent interaction in general, a means to address this is by developing design patterns, i.e., interaction templates that generalize across different use cases (see, for example, [16–18]). Another approach specifically for conversational agent research is the *Artificial Social Agent Evaluation Instrument*, an assessment and reporting tool for artificial social agents developed by Fitrianie *et al.* [19]. A similar, concise, usable, and comprehensive reporting standard for e-mental health for depression systems might readily translate to other fields of e-health and could facilitate the application of measures (e.g., eHDTS scales) that assess possibly important explanatory variables (e.g., technological sophistication) for the outcomes of meta-analyses.

For the machine learning algorithm that labels thoughts with regard to their underlying schemas, future work might lie in improving this with active learning [20]. The current model may be used as a starting point and as participants complete more thought records, they will be asked to label the conversational agent feedback as accurate or inaccurate, thus updating the model. Even more powerful would be the feedback of a therapist fed to the algorithm when the conversational agent is used as an adjunct to face-to-face therapy, for example. Additionally, follow-up research might look into ways to balance the presence of the schemas in the dataset. To this end, two approaches can be considered, a human approach or an algorithmic approach. Humans might be asked to generate more thoughts reflecting these schemas, such as designing scenarios for thought records that manage to elicit them or instructing people to explicitly formulate them. Alternatively, one might look into algorithmic approaches to create synthetic samples from the thoughts that are already in the dataset [21].

A further interesting direction for future research is a meta-analysis to link clinical outcomes with the technological sophistication of systems. Since it has previously been found that e-mental health is more effective when it is better adhered to [22, 23], it is important to study possible determinants of adherence. One such determinant is the motivation to engage with the system [24, 25]. This is likely to be affected by features of user, system, and interaction context. However, the sphere of influence of developers pertains to the features of the system. With the technological sophistication measure we open a door to study whether increasing the technological sophistication of a system is worthwhile. The Deprexis [26] system with a high technological sophistication and low dropout rates is a promising example in support of the idea that technological sophistication may correlate negatively with dropout. However, this is only one example and a meta-analysis is needed to draw conclusions.

If a meta-analysis should find a link between technological sophistication and adherence, this would encourage a repetition of our final study (Ch. 5) with some improvements. As a first improvement, a participatory design approach with a depressed population could be taken to designing the feedback of the conversational agent, as is prescribed, for example, by the socio-cognitive engineering method [27]. Additionally, it would be of interest to study the effects of the conversational agent in the long run in a depressed population, since the feedback of the agent builds on previous sessions and gets increasingly rich as users complete more thought records. It may also be considered that the agent feedback can be regarded as a kind of *reflective listening* in the context of motivational interviewing [10]. Future research taking this perspective might investigate whether adding additional conversational techniques from motivational interviewing increases motivation. This might indicate that only the content of the feedback is not sufficient to increase motivation but rather that motivational interviewing as a holistic approach is needed.

Finally, future research might explore the use of the conversational agent in other contexts or larger systems. For example, one might study the combination of the conversational agent that supports thought recording with one that supports thought strengthening (e.g. [28]) to form a cognitive restructuring micro-intervention. Given the very positive feedback concerning the usefulness of the experiment that we obtained from the healthy and subclinically depressed participants in both of our studies, such a micro-intervention might hold much potential as a low-risk preventive. Another example is the application of the conversational agent in an ecological momentary assessment application, which would allow for the collection of long-term interactions and for linking the schema-model of the user with other symptoms of depression over time [29]. In a similar vein, the agent could be integrated into a larger stand-alone treatment system, into face-to-face therapy as an adjunct for homework assignments, or into a blended care approach. In the latter two cases, the agent may need to interact with both therapist and client and possibly even take on the role of mediator between the two stakeholders.

6.4. CONTRIBUTIONS

This thesis makes a number of contributions for both scientists and system developers. These consist of the generated insights and the resulting new research questions on the one hand and the methods and tools that were developed on the other hand. Several contributions were made with the literature review. The eHDTS scales quantify the degree of technological sophistication of system functions. They thus allow researchers and system developers to obtain an estimate of how technologically sophisticated the functions of their system are, how they compare to the same or similar functions implemented in other systems, and where possible points of improvement may lie. Additionally, researchers can study the technological sophistication of systems or functions by including this as a variable in analyses. For example, the quantification of technological sophistication makes it possible to study the link between technological sophistication and adherence in a meta-analysis. The coding of technological sophistication further enabled an overview of the technological state of the art of the field of e-mental health for depression with the conclusion that most systems were lacking in interactivity and artificial intelligence: they often take the form of psychoeducational interventions, i.e., delivering information and exercises to the user but not processing information coming from the user. Two additional insights gained from the review are that despite an ever-increasing number of systems being developed in the examined time frame, there was no observable progress in technological sophistication and that human support is neither compensated with more functions nor with more technological sophistication in autonomous systems. The database developed for the literature review describing systems, their characteristics, their versions, their functions, their evaluations, and the scientific articles that detail them is a key contribution of this research work due to systematically organizing the systems found in the literature. To the best of our knowledge, it is unprecedented in its comprehensiveness and covers 17 years of research. Coupling it with other databases in the field, such as the one concerning clinical trials of Cuijpers *et al.* [30], can enable researchers to efficiently answer research questions concerning these systems. Furthermore, the database is described in detail on its own website[1] and is also contained in the data repository for the literature review, such that researchers interested in extending it with additional variables or more recent systems can easily do so. Developers of future systems can used the database for inspiration regarding the functionality they would like their system to offer. The variables used to systematically describe the systems in the database are concisely presented in a taxonomy consisting of system and evaluation attributes.

Further scientific contributions followed from testing whether it is possible to automatically identify schemas from thoughts. For one, a dataset of thought records was collected and manually labeled. Generally, such a dataset of thought records can be of value to researchers as thought record data is hard to come by. With our dataset it was possible to train, validate, and test several machine learning algorithms with regard to their thought labeling capabilities. For most schemas, there was at least one algorithm capable of labeling the thoughts such that the correlation between algorithmically and manually assigned schema scores was moderate to strong, and all algorithms were able to label thoughts with regard to at least six of the schemas. All scripts and data used for training and testing the algorithms to answer our first hypothesis are available in a data repository. For us, a set of recurrent neural networks (RNNs) trained separately for each schema performed best overall and could be used by us in generating the feedback of the conversational agent. Improving the labeling accuracy may be a challenge for data scientists as the dataset is an unbalanced dataset of natural language utterances with multiple, not mutually exclusive labels and a complex nesting structure. Taken together with the unlabeled dataset of thought records collected to test our second hypothesis, additional avenues for labeling become possible (e.g., applying methods of semi-supervised learning). For clinical psychologists, a key contribution is the demonstration that qualitative, text-based input can be automatically analyzed for the purpose of psychological assessment on a large scale. This opens up new possibilities for using such methods in addition to, or even instead of [28], questionnaires. Specifically for cognitive behavioral therapists, a higher labeling accuracy of the algorithm may be an important step towards use in practice. They may also consider extending the coding with other labels, such as another set of schemas or the modes of schema therapy [31]. For system developers, the algorithms and our dataset of thought records may present inspiration or a starting point for including similar functionality in their own system. While we make our algorithms and data available for all purposes, we caution against out-of-the-box use in practice as we never evaluated any part of it with a clinical population.

Finally, we contribute the conversational agent that supports users in completing thought records and is capable of providing feedback on the process (number of

[1] http://insyprojects.ewi.tudelft.nl:8888/

completed downward-arrow steps) and on the content (active schemas) of the thought record and all completed thought records over time. This agent was found to be a feasible means of collecting thought record data in dialog format. Participants who chose to comment considered it an insightful experiment and intended to keep using thought recording in their day-to-day lives. The thoughts they denoted could, for the most part, be labeled by the RNNs, and the frequencies of schema occurrences in this sample of thoughts (subclinical population) was similar to those assigned by the RNNs on the dataset on which the RNNs were tested (healthy population) and similar to those manually assigned by Millings *et al.* [6] (clinical population). The conversational agent is available to researchers or developers to further improve upon, to integrate into their own system, or to use as is.

6.5. FINAL REMARKS

Depression is a debilitating mental illness that can cost years or even decades to overcome. In the best case, the affected person seeks help during the first episode, is treated effectively with psychotherapy, medication, or both, and no further episodes follow. In the worst case, the affected person sees no way out and does not seek help or does not obtain it. While e-mental health presents a good way of increasing accessibility to treatment, in various articles in the field of e-mental health for depression it is clearly emphasized that the systems and authors do not aim to replace therapists but rather that these systems present a viable step in stepped care models (e.g., [32]). Even in their model of supportive accountability, Mohr *et al.* [33] claim that this accountability necessarily requires the presence of another human and thus rule out the possibility of supportive accountability in a stand-alone system. In light of our findings that there is little technological progress in the field despite much technological progress outside of it, one is left to wonder what has hampered researchers' desire to experiment with more technologically advanced solutions. Feijt *et al.* [34] found that an important barrier to use of e-health in practice is the lack of therapists' experience with the technology. During the COVID19 pandemic, however, interacting with clients remotely by means of technology became a necessity for many therapists. In the Netherlands, this exposure has resulted in an attitude shift in healthcare providers with respect to e-health, albeit largely e-health in the sense of computer mediated communication with patients [35]. Obviously, technology or artificial intelligence should not be employed at all costs and comes with its own risks. As Darcy *et al.* [36] point out, however, the technological revolution of the field of medicine is already underway and clinicians are in a unique position to ensure that the technology will serve the good of patients. Rather than avoiding technology, we therefore second the opinion that clinical psychologists should work together with computer and data scientists to develop ethically responsible interventions that also move with the times technologically. From a conversational agent perspective, we see it as an interesting challenge to attempt to mimic what we know are effective therapist behaviors with conversational agents and to then study the effects. However, from a clinical perspective, one may consider that conversational agents can allow for studying aspects of human behavior in a controlled and isolated manner, possibly elucidating the effective behaviors of therapists in the first place and better than can be done in studies of human-human interaction. It is common practice in conversational agent research to go down the former path and take what is known from psychology research to see how it translates to agents. Research exploring the opposite direction critically hinges on the close collaboration of clinical psychologists with computer scientists and has the potential to enrich both fields.

Bibliography

1. Microsoft Patterns & Practices Team. *Microsoft Application Architecture Guide, 2nd Edition* (Microsoft Press, 2009).
2. Oinas-Kukkonen, H. & Harjumaa, M. Persuasive systems design: Key issues, process model, and system features. *Communications of the Association for Information Systems* **24,** 485–500 (2009).
3. Fitzpatrick, K. K., Darcy, A. & Vierhile, M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. *JMIR Mental Health* **4** (2017).
4. Morris, R. R., Schueller, S. M. & Picard, R. W. Efficacy of a web-based, crowd-sourced peer-to-peer cognitive reappraisal platform for depression: Randomized controlled trial. *Journal of Medical Internet Research* **17** (2015).
5. Falconer, C. J., Rovira, A., King, J. A., Gilbert, P., Antley, A., Fearon, P., Ralph, N., Slater, M. & Brewin, C. R. Embodying self-compassion within virtual reality and its effects on patients with depression. *BJPsych Open* **2,** 74–80 (2016).
6. Millings, A. & Carnelley, K. B. Core belief content examined in a large sample of patients using online cognitive behaviour therapy. *Journal of Affective Disorders* **186,** 275–283 (2015).
7. Pickering, D. & Blaszczynski, A. Paid online convenience samples in gambling studies: Questionable data quality. *International Gambling Studies* **21,** 516–536 (2021).
8. Litman, L., Robinson, J. & Rosenzweig, C. The relationship between motivation, monetary compensation, and data quality among US-and India-based workers on Mechanical Turk. *Behavior research methods* **47,** 519–528 (2015).
9. Mącik, D. & Mącik, R. Are four maladaptive schema domains a better option than five? Recommendations based on comparison of the latent structure of schemas on a large group of healthy adults. *Behavioural and Cognitive Psychotherapy,* 1–11 (2022).
10. Miller, W. R. & Rollnick, S. *Motivational Interviewing: Preparing People to Change Addictive Behaviour.* (Guilford Press, 1991).
11. Pasqualetti, G., Gori, G., Blandizzi, C. & Del Tacca, M. Healthy volunteers and early phases of clinical experimentation. *European journal of clinical pharmacology* **66,** 647–653 (2010).
12. Detweiler, J. B. & Whisman, M. A. The role of homework assignments in cognitive therapy for depression: Potential methods for enhancing adherence. *Clinical Psychology: Science and Practice* **6,** 267–282 (1999).

13. Schröder, J., Berger, T., Westermann, S., Klein, J. P. & Moritz, S. Internet interventions for depression: New developments. *Dialogues in Clinical Neuroscience* **18,** 203–212 (2016).

14. Open Science Collaboration. Estimating the reproducibility of psychological science. *Science* **349,** aac4716 (2015).

15. Baker, M. 1,500 scientists lift the lid on reproducibility. *Nature* **533** (2016).

16. Ligthart, M. E., Neerincx, M. A. & Hindriks, K. V. Design patterns for an interactive storytelling robot to support children's engagement and agency. in *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (Association for Computing Machinery, 2020), 409–418.

17. Schoonderwoerd, T. A., Jorritsma, W., Neerincx, M. A. & Van Den Bosch, K. Human-centered XAI: Developing design patterns for explanations of clinical decision support systems. *International Journal of Human-Computer Studies* **154,** 102684 (2021).

18. van Diggelen, J., Neerincx, M., Peeters, M. & Schraagen, J. M. Developing effective and resilient human-agent teamwork using team design patterns. *IEEE intelligent systems* **34,** 15–24 (2018).

19. Fitrianie, S., Bruijnes, M., Li, F. & Brinkman, W.-P. Questionnaire Items for Evaluating Artificial Social Agents-Expert Generated, Content Validated and Reliability Analysed. in *Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents* (Association for Computing Machinery, 2021), 84–86.

20. Monarch, R. M. *Human-in-the-Loop Machine Learning: Active learning and annotation for human-centered AI* (Simon and Schuster, 2021).

21. Kaur, H., Pannu, H. S. & Malhi, A. K. A systematic review on imbalanced data challenges in machine learning: Applications and solutions. *ACM Computing Surveys (CSUR)* **52,** 1–36 (2019).

22. Hilvert-Bruce, Z., Rossouw, P. J., Wong, N., Sunderland, M. & Andrews, G. Adherence as a determinant of effectiveness of internet cognitive behavioural therapy for anxiety and depressive disorders. *Behaviour research and therapy* **50,** 463–468 (2012).

23. Donkin, L., Christensen, H., Naismith, S. L., Neal, B., Hickie, I. B. & Glozier, N. A systematic review of the impact of adherence on the effectiveness of e-therapies. *Journal of medical Internet research* **13,** e52 (2011).

24. Achilles, M. R., Anderson, M., Li, S. H., Subotic-Kerry, M., Parker, B. & O'Dea, B. Adherence to e-mental health among youth: Considerations for intervention development and research design. *Digital health* **6,** 2055207620926064 (2020).

25. Eccles, H., Nannarone, M., Lashewicz, B., Attridge, M., Marchand, A., Aiken, A., Ho, K., Wang, J., *et al.* Barriers to the use of web-based mental health programs for preventing depression: qualitative study. *JMIR formative research* **5,** e16949 (2021).

26. Twomey, C., O'Reilly, G. & Meyer, B. Effectiveness of an individually-tailored computerised CBT programme (Deprexis) for depression: a meta-analysis. *Psychiatry research* **256,** 371–377 (2017).

27. Neerincx, M. A., Van Vught, W., Blanson Henkemans, O., Oleari, E., Broekens, J., Peters, R., Kaptein, F., Demiris, Y., Kiefer, B., Fumagalli, D., *et al.* Socio-cognitive engineering of a robotic partner for child's diabetes self-management. *Frontiers in Robotics and AI* **6,** 118 (2019).

28. Al Owayyed, M. *Motivating PhD candidates with depression symptoms to complete thoughts-strengthening exercises via a conversational agent.* Master's thesis, `https://repository.tudelft.nl/islandora/object/uuid%3Ac5ed5bde-e118-4b73-9f4b-85531920f19a`. 2020.

29. Huckins, J. F., DaSilva, A. W., Hedlund, E. L., Murphy, E. I., Rogers, C., Wang, W., Obuchi, M., Holtzheimer, P. E., Wagner, D. D. & Campbell, A. T. Causal factors of anxiety and depression in college students: longitudinal ecological momentary assessment and causal analysis using Peter and Clark momentary conditional independence. *JMIR mental health* **7,** e16684 (2020).

30. Cuijpers, P., van Straten, A., Warmerdam, L. & Andersson, G. Psychological treatment of depression: a meta-analytic database of randomized studies. *BMC psychiatry* **8,** 1–6 (2008).

31. Arntz, A. & Jacob, G. *Schema therapy in practice: An introductory guide to the schema mode approach* (John Wiley & Sons, 2017).

32. Kaltenthaler, E., Brazier, J., De Nigris, E., Tumur, I., Ferriter, M., Beverley, C., Parry, G., Rooney, G. & Sutcliffe, P. Computerised cognitive behaviour therapy for depression and anxiety update: a systematic review and economic evaluation. *Health Technology Assessment* **10,** iii–xi (2006).

33. Mohr, D., Cuijpers, P. & Lehman, K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. *Journal of medical Internet research* **13,** e30 (2011).

34. Feijt, M. A., de Kort, Y. A., Bongers, I. M. & IJsselsteijn, W. A. Perceived drivers and barriers to the adoption of eMental health by psychologists: the construction of the levels of adoption of eMental health model. *Journal of medical Internet research* **20,** e9485 (2018).

35. Vaart, R., Deursen, L. v., Standaar, L., Wouters, M., Suijkerbuijk, A., van Tuyl, L., Aardoom, J., Versluis, A. & Rompelberg, C. E-healthmonitor 2021: stand van zaken digitale zorg. (2022).

36. Darcy, A. M., Louie, A. K. & Roberts, L. W. Machine learning and the profession of medicine. *Jama* **315,** 551–552 (2016).

A

Search terms for the literature survey

A

A.1. SEARCH TERMS

Technology	Condition	Goal	Evaluation	Exclude
computer*/ *Computers*/ *Computer Systems*	depress*/ *depressive disorder*/ *depressive disorder, major*	prevent*	evaluat*	PTSD
internet/ *Internet*	dysthmi*/ *dysthymic disorder*	treat*	trial	postpartum
Software		train*	test	postnatal
online		monitor*	study	perinatal*
web*/ *Web Browser*		simulat*	pilot	adolescent*
mail*		interven*	rct	child*
mobile/ *Cell Phones*/ *Computers, Handheld*		support*	conduct*	telephone
phone			experiment*	tele$health
text message			significan*	review
app			allocat*	meta$analy*
techn*			control*	survey
digital			participant*	screening
automat*			random*	psychosis
virtual			recruit*	psychotic
remote				bipolar
self$help/ *Self-Help Devices*				
cCBT/ *Therapy, Computer-Assisted*				
iCBT				

Table A.1: The search terms that were used in retrieving primary articles from the databases Scopus, PubMed, and Web of Science are shown. Columns are combined with logical ANDs while cells within columns are combined with logical ORs. The first two columns were searched for within titles and keywords, the third and fourth column within abstracts. Terms from the *Exclude* column were not allowed to appear within titles only. The first and second column include in italics the Medical Subject Heading (MeSH) terms that were used in addition to the regular search terms in PubMed. Finally, where possible, wildcards were used to expand the search terms with * denoting any string including the empty one and $ denoting any string of length 1 or the empty string.

A.2. PUBMED

(((internet[Title] OR online[Title] OR web*[Title] OR mobile[Title] OR computer*[Title] OR techn*[Title] OR telemedic*[Title] OR tele-medic*[Title] OR virtual[Title] OR remote[Title] OR e-health[Title] OR ehealth[Title] OR automat*[Title] OR "text message"[Title] OR mail*[Title] OR mhealth[Title] OR m-health[Title] OR digital[Title] OR phone[Title] OR app[Title] OR cCBT[Title] OR distance[Title] OR self-help[Title]) OR (Internet[MeSH Term] OR "Web Browser"[MeSH Term] OR "Therapy, Computer-Assisted"[MeSH Term] OR Software[MeSH Term] OR "Computers, Handheld"[MeSH Term] OR Telemedicine[MeSH Term] OR "Cell Phones"[MeSH Term] OR "Self-Help Devices"[MeSH Term] OR "Computer Systems"[MeSH Term] OR Computers[MeSH Term]) OR (internet[Other Term] OR online[Other Term] OR web*[Other Term] OR mobile[Other Term] OR computer* [Other Term] OR techn*[Other Term] OR telemedic*[Other Term] OR tele-medic*[Other Term] OR virtual[Other Term] OR remote[Other Term] OR ehealth[Other Term] OR e-health[Other Term] OR automat*[Other Term] OR "text message"[Other Term] OR mail*[Other Term] OR mhealth[Other Term]

OR m-health[Other Term] OR digital[Other Term] OR phone[Other Term] OR app[Other Term] OR cCBT[Other Term] OR distance[Other Term] OR self-help[Other Term])) AND ((depress*[Title] OR dysthymi*[Title]) OR (depressive disorder[MeSH Terms] OR depressive disorder, major[MeSH Terms] OR dysthymic disorder[MeSH Terms]) OR (depress*[Other Term] OR dysthymi*[Other Term])) AND ((prevent*[Title/Abstract] OR treat*[Title/Abstract] OR monitor*[Title/Abstract] OR train*[Title/Abstract] OR simulat*[Title/Abstract] OR interven*[Title/Abstract] OR support*[Title/Abstract]) AND (evaluat*[Title/Abstract] OR trial[Title/Abstract] OR test[Title/Abstract] OR pilot[Title/Abstract] OR study[Title/Abstract] OR RCT[Title/Abstract] OR conduct*[Title/Abstract] OR experiment*[Title/Abstract] OR significan*[Title/Abstract] OR allocat*[Title/Abstract] OR control*[Title/Abstract] OR random*[Title/Abstract] OR recruit*[Title/Abstract] OR participant*[Title/Abstract])) NOT (PTSD[Title] OR postpartum[Title] OR postnatal[Title] OR adolescent*[Title] OR child*[Title] OR bipolar[Title]) AND ((Classical Article[ptyp] OR Clinical Conference[ptyp] OR Clinical Study[ptyp] OR Clinical Trial[ptyp] OR Comparative Study[ptyp] OR Controlled Clinical Trial[ptyp] OR Evaluation Studies[ptyp] OR Journal Article[ptyp] OR Observational Study[ptyp] OR Pragmatic Clinical Trial[ptyp] OR Randomized Controlled Trial[ptyp] OR Technical Report[ptyp] OR Validation Studies[ptyp]) AND hasabstract[text] AND ("2000/01/01"[PDat] : "2017/12/31"[PDat])))

A.3. SCOPUS

(((((((TITLE (internet OR online OR web* OR mobile OR computer* OR techn* OR virtual OR remote OR e?health OR "text message" OR mail* OR m?health OR digital OR phone OR app OR ccbt OR icbt) OR KEY (internet OR online OR web* OR mobile OR computer* OR techn* OR virtual OR remote OR e?health OR "text message" OR mail* OR m?health OR digital OR phone OR app OR ccbt OR icbt)) AND (TITLE (depress* OR dysthymi*) OR KEY (depress* OR dysthymi*))) AND ABS (prevent* OR treat* OR monitor* OR train* OR simulat* OR interven* OR support*)) AND ABS (evaluat* OR trial OR test OR pilot OR study OR rct OR conduct* OR experiment* OR significan* OR allocat* OR control* OR random* OR recruit* OR participant*)) AND (PUBYEAR > 2000)) AND NOT (TITLE (ptsd OR postpartum OR postnatal OR perinatal OR adolescent* OR child* OR bipolar OR review OR meta?analy* OR survey OR screen* OR assess* OR psychosis OR telephone OR tele?health OR tele?medicine OR addict* OR predict* OR stigma OR schizophreni* OR reproduct* OR maternal OR sleep OR anti$depress* OR "support group") OR KEY (ptsd OR postpartum OR postnatal OR perinatal OR adolescent* OR child* OR bipolar OR review OR meta?analy* OR survey OR screen* OR assess* OR psychosis OR telephone OR tele?health OR tele?medicine OR addict* OR predict* OR stigma OR schizophreni* OR reproduct* OR maternal OR sleep OR anti$depress* OR "support group")))) AND (LIMIT-TO (DOCTYPE , "ar ") OR LIMIT-TO (DOCTYPE , "cp ") OR LIMIT-TO (DOCTYPE , "ip ")) AND (LIMIT-TO (LANGUAGE , "English "))

A

A.4. WEB OF SCIENCE

TS=(internet OR online OR web* OR mobile OR computer* OR techn* OR virtual OR remote OR e?health OR "text message" OR mail* OR m?health OR digital OR phone OR app OR ccbt OR icbt) AND TS=(depress* OR dysthymi*) AND TS=(prevent* OR treat* OR monitor* OR train* OR simulat* OR interven* OR support*) AND TS=(evaluat* OR trial OR test OR pilot OR study OR rct OR conduct* OR experiment* OR significan* OR allocat* OR control* OR random* OR recruit* OR participant*) NOT (TI=(ptsd OR postpartum OR postnatal OR perinatal OR adolescent* OR child* OR bipolar OR review OR meta?analy* OR survey OR screen* OR assess* OR psychosis OR telephone OR tele?health OR tele?medicine OR addict* OR predict* OR stigma OR schizophreni* OR reproduct* OR maternal OR sleep OR anti$depress* OR "support group") OR TS=(ptsd OR postpartum OR postnatal OR perinatal OR adolescent* OR child* OR bipolar OR review OR meta?analy* OR survey OR screen* OR assess* OR psychosis OR telephone OR tele?health OR tele?medicine OR addict* OR predict* OR stigma OR schizophreni* OR reproduct* OR maternal OR sleep OR anti$depress* OR "support group")) AND LANGUAGE: (English)Refined by: DOCUMENT TYPES: (ARTICLE OR PROCEEDINGS PAPER) AND WEB OF SCIENCE CATEGORIES: (ONCOLOGY OR PSYCHIATRY OR GERONTOLOGY OR PSYCHOLOGY CLINICAL OR COMPUTER SCIENCE INFORMATION SYSTEMS OR PUBLIC ENVIRONMENTAL OCCUPATIONAL HEALTH OR HEALTH CARE SCIENCES SERVICES OR ENGINEERING ELECTRICAL ELECTRONIC OR INTEGRATIVE COMPLEMENTARY MEDICINE OR PSYCHOLOGY MULTIDISCIPLINARY OR COMPUTER SCIENCE THEORY METHODS OR SUBSTANCE ABUSE OR BEHAVIORAL SCIENCES OR MEDICINE GENERAL INTERNAL OR REMOTE SENSING OR PSYCHOLOGY EXPERIMENTAL OR PSYCHOLOGY OR PSYCHOLOGY APPLIED OR NURSING OR MEDICAL INFORMATICS OR SOCIAL SCIENCES INTERDISCIPLINARY OR MEDICINE RESEARCH EXPERIMENTAL OR COMPUTER SCIENCE ARTIFICIAL INTELLIGENCE OR COMPUTER SCIENCE CYBERNETICS OR REHABILITATION OR GERIATRICS GERONTOLOGY OR COMPUTER SCIENCE INTERDISCIPLINARY APPLICATIONS) Timespan: 2000-2017. Indexes: SCI-EXPANDED, SSCI, A-HCI, CPCI-S, CPCI-SSH, ESCI.

B

LIST OF ARTICLES INCLUDED IN THE LITERATURE REVIEW

B

1. Aguilera A, Muñoz RF. Text messaging as an adjunct to CBT in low-income populations: A usability and feasibility pilot study. Prof Psychol Res Pract. 2011;42(6):472-478. doi:10.1037/a0025499
2. Agyapong VIO, Milnes J, Farren CK. Supportive text messages for patients with alcohol use disorder and comorbid depression - how useful are they for the patients. Eur Psychiatry. 2013;28(March 2015):1. doi:10.1016/s0924-9338(13)75858-9
3. Ahmedani BK, Belville-Robertson T, Hirsch A, Jurayj A. An Online Mental Health and Wellness Intervention Supplementing Standard Care of Depression and Anxiety. Arch Psychiatr Nurs. 2016;30(6):666-670. doi:10.1016/j.apnu.2016.03.003
4. Ahmedani BK, Crotty N, Abdulhak MM, Ondersma SJ. Pilot feasibility study of a brief, tailored mobile health intervention for depression among patients with chronic pain. Behav Med. 2015;41(1):25-32. doi:10.1080/08964289.2013.867827
5. Aikens JE, Trivedi R, Heapy A, Pfeiffer PN, Piette JD. Potential Impact of Incorporating a Patient-Selected Support Person into mHealth for Depression. J Gen Intern Med. 2015;30(6):797-803. doi:10.1007/s11606-015-3208-7
6. Alavi N, Hirji A, Sutton C, Naeem F. Online CBT Is Effective in Overcoming Cultural and Language Barriers in Patients with Depression. J Psychiatr Pract. 2016;22(1):2-8. doi:10.1097/PRA.0000000000000119
7. Andersson G, Bergström J, Holländare F, Carlbring P, Kaldo V, Ekselius L. Internet-based self-help for depression: Randomised controlled trial. Br J Psychiatry. 2005;187(NOV.):456-461. doi:10.1192/bjp.187.5.456
8. Andersson G, Hesser H, Veilord A, et al. Randomised controlled non-inferiority trial with 3-year follow-up of internet-delivered versus face-to-face group cognitive behavioural therapy for depression. J Affect Disord. 2013;151(3):986-994. doi:10.1016/j.jad.2013.08.022
9. Anguera JA, Gunning FM, Areán PA. Improving late life depression and cognitive control through the use of therapeutic video game technology: A proof-of-concept randomized trial. Depress Anxiety. 2017;34(6):508-517. doi:10.1002/da.22588
10. Bae J-Y, Lee S-W, Yoon S-H, An K-E. Preliminary evaluation of a web site for depressive symptoms management. Stud Health Technol Inform. 2006.
11. Batterham PJ, Neil AL, Bennett K, Griffiths KM, Christensen H. Predictors of adherence among community users of a cognitive behavior therapy website. Patient Prefer Adherence. 2008;2(May):97-105. http://www.ncbi.nlm.nih.gov/pubmed/19920949%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2770409.
12. Beevers CG, Clasen PC, Enock PM, Schnyer DM. Attention bias modification for major depressive disorder: Effects on attention bias, resting state connectivity, and symptom change. J Abnorm Psychol. 2015. doi:10.1037/abn0000049
13. Berger T, Hämmerli K, Gubser N, Andersson G, Caspar F. Internet-Based Treatment of Depression: A Randomized Controlled Trial Comparing Guided with Unguided Self-Help. Cogn Behav Ther. 2011;40(4):251-266. doi:10.1080/16506073.2011.616531
14. Berman MI, Buckey JC, Hull JG, et al. Feasibility Study of an Interactive Multimedia Electronic Problem Solving Treatment Program for Depression: A Preliminary Uncontrolled Trial. Behav Ther. 2014;45(3):358-375. doi:10.1016/j.beth.2014.02.001
15. Birney AJ, Gunn R, Russell JK, Ary D V. MoodHacker Mobile Web App With Email for Adults to Self-Manage Mild-to-Moderate Depression: Randomized Controlled Trial. JMIR mHealth uHealth. 2016;4(1):e8. doi:10.2196/mhealth.4231
16. Blackwell SE, Browning M, Mathews A, et al. Positive imagery-based cognitive bias modification as a web-based treatment tool for depressed adults: A

randomized controlled trial. Clin Psychol Sci. 2015;3(1):91-111. doi:10.1177/2167702614560746

17. Blackwell SE, Holmes EA. Modifying interpretation and imagination in clinical depression: A single case series using cognitive bias modification. Appl Cogn Psychol. 2010;24(3):338-350. doi:10.1002/acp.1680
18. Blom K, Jernelöv S, Kraepelien M, et al. Internet Treatment Addressing either Insomnia or Depression, for Patients with both Diagnoses: A Randomized Trial. Sleep. 2015;38(2):267-277. doi:10.5665/sleep.4412
19. Blom MM, Zarit SH, Groot Zwaaftink RBM, Cuijpers P, Pot AM. Effectiveness of an internet intervention for family caregivers of people with dementia: Results of a randomized controlled trial. PLoS One. 2015;10(2):1-11. doi:10.1371/journal.pone.0116622
20. Boeschoten RE, Dekker J, Uitdehaag BMJ, et al. Internet-based treatment for depression in multiple sclerosis: A randomized controlled trial. Mult Scler. 2017;23(8):1112-1122. doi:10.1177/1352458516671820
21. Boeschoten RE, Nieuwenhuis MM, Van Oppen P, et al. Feasibility and outcome of a web-based self-help intervention for depressive symptoms in patients with multiple sclerosis: A pilot study. J Neurol Sci. 2012;315(1-2):104-109. doi:10.1016/j.jns.2011.11.016
22. Bolier L, Haverman M, Kramer J, et al. An internet-Based intervention to promote mental fitness for mildly depressed adults: Randomized controlled trial. J Med Internet Res. 2013;15(9):1-16. doi:10.2196/jmir.2603
23. Bond GE, Burr RL, Wolf FM, Feldt K. The effects of a web-based intervention on psychosocial well-being among adults aged 60 and older with diabetes: A randomized trial. Diabetes Educ. 2010;36(3):446-456. doi:10.1177/0145721710366758
24. Botella C, Etchemendy E, Castilla D, et al. An e-Health System for the Elderly (Butler Project): A Pilot Study on Acceptance and Satisfaction. CyberPsychology Behav. 2009;12(3):255-262. doi:10.1089/cpb.2008.0325
25. Botella C, Mira A, Moragrega I, et al. An internet-based program for depression using activity and physiological sensors: Efficacy, expectations, satisfaction, and ease of use. Neuropsychiatr Dis Treat. 2016;12:393-406. doi:10.2147/NDT.S93315
26. Both F, Cuijpers P, Hoogendoorn M, Klein M. TOWARDS FULLY AUTOMATED PSYCHOTHERAPY FOR ADULTS - BAS - Behavioral Activation Scheduling Via Web and Mobile Phone. 2011:375-380. doi:10.5220/0002688503750380
27. Bowie CR, Gupta M, Holshausen K, Jokic R, Best M, Milev R. Cognitive Remediation for Treatment-Resistant Depression: Effects on Cognition and Functioning and the Role of Online Homework. J Nerv Ment Dis. 2013;201(8):680-685. doi:10.1097/NMD.0b013e31829c5030
28. Braithwaite SR, Fincham FD. ePREP: Computer Based Prevention of Relationship Dysfunction, Depression and Anxiety. J Soc Clin Psychol. 2007;26(5):609-622. doi:10.1521/jscp.2007.26.5.609
29. Bresó A, Martínez-Miranda J, Fuster-García E, García-Gómez JM. A novel approach to improve the planning of adaptive and interactive sessions for the treatment of Major Depression. Int J Hum Comput Stud. 2016;87:80-91. doi:10.1016/j.ijhcs.2015.11.003
30. Browning M, Holmes EA, Charles M, Cowen PJ, Harmer CJ. Using attentional bias modification as a cognitive vaccine against depression. Biol Psychiatry. 2012;72(7):572-579. doi:10.1016/j.biopsych.2012.04.014
31. Buhrman M, Syk M, Burvall O, Hartig T, Gordh T, Andersson G. Individualized guided internet-delivered cognitive-behavior therapy for chronic pain patients with comorbid depression and anxiety: A randomized controlled trial. Clin J Pain. 2015;31(6):504-516. doi:10.1097/AJP.0000000000000176

B

32. Bunge EL, Williamson RE, Cano M, Leykin Y, Muñoz RF. Mood management effects of brief unsupported internet interventions. Internet Interv. 2016;5:36-43. doi:10.1016/j.invent.2016.06.001

33. Buntrock C, Ebert DD, Lehr D, et al. Effect of a web-based guided self-help intervention for prevention of major depression in adults with subthreshold depression a randomized clinical trial. JAMA - J Am Med Assoc. 2016;315(17):1854-1863. doi:10.1001/jama.2016.4326

34. Buntrock C, Ebert D, Lehr D, et al. Effectiveness of a Web-Based Cognitive Behavioural Intervention for Subthreshold Depression: Pragmatic Randomised Controlled Trial. Psychother Psychosom. 2015;84(6):348-358. doi:10.1159/000438673

35. Burns MN, Begale M, Duffecy J, et al. Harnessing context sensing to develop a mobile intervention for depression. J Med Internet Res. 2011;13(3):1-19. doi:10.2196/jmir.1838

36. Burton C, Szentagotai Tatar A, McKinstry B, et al. Pilot randomised controlled trial of Help4Mood, an embodied virtual agent-based system to support treatment of depression. J Telemed Telecare. 2016;22(6):348-355. doi:10.1177/1357633x15609793

37. Bush NE, Ouellette G, Kinn J. Utility of the T2 Mood Tracker Mobile Application Among Army Warrior Transition Unit Service Members. Mil Med. 2014;179(12):1453-1457. doi:10.7205/milmed-d-14-00271

38. Calkins AW, McMorran KE, Siegle GJ, Otto MW. The Effects of Computerized Cognitive Control Training on Community Adults with Depressed Mood. Behav Cogn Psychother. 2015;43(5):578-589. doi:10.1017/s1352465814000046

39. Carlbring P, Hägglund M, Luthström A, et al. Internet-based behavioral activation and acceptance-based treatment for depression: A randomized controlled trial. J Affect Disord. 2013;148(2-3):331-337. doi:10.1016/j.jad.2012.12.020

40. Cartreine JA, Locke SE, Buckey JC, Sandoval L, Hegel MT. Electronic problem-solving treatment: Description and pilot study of an interactive media treatment for depression. J Med Internet Res. 2012;14(5):1-14. doi:10.2196/resprot.1925

41. Cassimatis M, Kavanagh DJ, Hills AP, et al. Development of the OnTrack Diabetes Program. JMIR Res Protoc. 2015;4(2):e24. doi:10.2196/resprot.2823

42. Cassimatis M, Kavanagh DJ, Hills AP, et al. The OnTrack Diabetes Web-Based Program for Type 2 Diabetes and Dysphoria Self-Management: A Randomized Controlled Trial Protocol. JMIR Res Protoc. 2015;4(3):e97. doi:10.2196/resprot.2813

43. Cavanagh K, Shapiro DA, Van Den Berg S, Swain S, Barkham M, Proudfoot J. The effectiveness of computerized cognitive behavioural therapy in routine care. Br J Clin Psychol. 2006;45(4):499-514. doi:10.1348/014466505X84782

44. Cavanagh K, Shapiro DA, van Den Berg S, Swain S, Barkham M, Proudfoot J. The acceptability of computer-aided cognitive behavioural therapy: A pragmatic study. Cogn Behav Ther. 2009;38(4):235-246. doi:10.1080/16506070802561256

45. Cavanagh K, Strauss C, Cicconi F, Griffiths N, Wyper A, Jones F. A randomised controlled trial of a brief online mindfulness-based intervention. Behav Res Ther. 2013;51(9):573-578. doi:10.1016/j.brat.2013.06.003

46. Chan R, Dear BF, Titov N, Chow J, Suranyi M. Examining internet-delivered cognitive behaviour therapy for patients with chronic kidney disease on haemodialysis: A feasibility open trial. J Psychosom Res. 2016;89:78-84. doi:10.1016/j.jpsychores.2016.08.012

47. Choi I, Zou J, Titov N, et al. Culturally attuned internet treatment for depression amongst Chinese Australians: A randomised controlled trial. J Affect Disord. 2012;136(3):459-468. doi:10.1016/j.jad.2011.11.003

48. Christensen H, Batterham PJ, Gosling JA, et al. Effectiveness of an online insomnia program (SHUTi) for prevention of depressive episodes (the GoodNight Study): A randomised controlled trial. The Lancet Psychiatry. 2016;3(4):333-341. doi:10.1016/S2215-0366(15)00536-2

49. Christensen H, Griffiths KM, Mackinnon AJ, Brittliffe K. Online randomized controlled trial of brief and full cognitive behaviour therapy for depression. Psychol Med. 2006;36(12):1737-1746. doi:10.1017/S0033291706008695

50. Christensen H, Griffiths KM, Jorm AF. Delivering interventions for depression by using the internet: randomised controlled trial. BMJ. 2004;328(7434):265. doi:10.1136/bmj.37945.566632.EE

51. Christensen H, Griffiths KM, Korten AE, Brittliffe K, Groves C. A comparison of changes in anxiety and depression symptoms of spontaneous users and trial participants of a cognitive behavior therapy website. J Med Internet Res. 2004;6(4):e46. doi:10.2196/jmir.6.4.e46

52. Christensen H, Griffiths KM, Korten A. Web-based cognitive behavior therapy: Analysis of site usage and changes in depression and anxiety scores. J Med Internet Res. 2002;4(1):29-40. doi:10.2196/jmir.4.1.e3

53. Clarke G, Eubanks D, Reid E, et al. Overcoming depression on the internet (ODIN) (2): A randomized trial of a self-help depression skills program with reminders. J Med Internet Res. 2005;7(2):1-12. doi:10.2196/jmir.7.2.e16

54. Clarke G, Kelleher C, Hornbrook M, DeBar L, Dickerson J, Gullion C. Randomized effectiveness trial of an internet, pure self-help, cognitive behavioral intervention for depressive symptoms in young adults. Cogn Behav Ther. 2009;38(4):222-234. doi:10.1080/16506070802675353

55. Clarke G, Reid E, Eubanks D, et al. Overcoming Depression on the Internet (ODIN): A randomized controlledtrialof an Internet depression skills intervention program. J Med Internet Res. 2002;4(3):5-17. doi:10.2196/jmir.4.3.e14

56. Clarke J, Proudfoot J, Ma H. Mobile Phone and Web-based Cognitive Behavior Therapy for Depressive Symptoms and Mental Health Comorbidities in People Living With Diabetes: Results of a Feasibility Study. JMIR Ment Heal. 2016;3(2):e23. doi:10.2196/mental.5131

57. Cohn MA, Pietrucha ME, Saslow LR, Hult JR, Moskowitz JT. An online positive affect skills intervention reduces depression in adults with type 2 diabetes. J Posit Psychol. 2014. doi:10.1080/17439760.2014.920410

58. Cooper CL, Hind D, Parry GD, et al. Computerised cognitive behavioural therapy for the treatment of depression in people with multiple sclerosis: External pilot trial. Trials. 2011;12:1-14. doi:10.1186/1745-6215-12-259

59. Cukrowicz KC, Joiner TE. Computer-based intervention for anxious and depressive symptoms in a non-clinical population. Cognit Ther Res. 2007;31(5):677-693. doi:10.1007/s10608-006-9094-x

60. Currie SL, Mcgrath PJ, Day V. Development and usability of an online CBT program for symptoms of moderate depression, anxiety, and stress in post-secondary students. Comput Human Behav. 2010;26(6):1419-1426. doi:10.1016/j.chb.2010.04.020

61. Day V, McGrath PJ, Wojtowicz M. Internet-based guided self-help for university students with anxiety, depression and stress: A randomized controlled clinical trial. Behav Res Ther. 2013;51(7):344-351. doi:10.1016/j.brat.2013.03.003

62. De Graaf LE, Gerhards SAH, Arntz A, et al. Clinical effectiveness of online computerised cognitive-behavioural therapy without support for depression in primary care: Randomised trial. Br J Psychiatry. 2009;195(1):73-80. doi:10.1192/bjp.bp.108.054429

63. Deady M, Kay-Lambkin F, Teesson M, Mills K. Developing an integrated, Internet-based self-help programme for young people with depression and alcohol use problems. Internet Interv. 2014;1(3):118-131. doi:10.1016/j.invent.2014.06.004

64. Deady M, Mills KL, Teesson M, Kay-Lambkin F. An online intervention for co-occurring depression and problematic alcohol use in young people: Primary outcomes from a randomized controlled trial. J Med Internet Res. 2016;18(3):1-12. doi:10.2196/jmir.5178

65. Dear BF, Gandy M, Karin E, et al. The Pain Course. Pain. 2015;156(10):1920-1935. doi:10.1097/j.pain.0000000000000251

66. Dear BF, Titov N, Perry KN, et al. The Pain Course: A randomised controlled trial of a clinician-guided Internet-delivered cognitive behaviour therapy program for managing chronic pain and emotional well-being. Pain. 2013;154(6):942-950. doi:10.1016/j.pain.2013.03.005

67. Dear BF, Titov N, Schwencke G, et al. An open trial of a brief transdiagnostic internet treatment for anxiety and depression. Behav Res Ther. 2011;49(12):830-837. doi:10.1016/j.brat.2011.09.007

68. Dear BF, Zou J, Titov N, et al. Internet-delivered cognitive behavioural therapy for depression: A feasibility open trial for older adults. Aust N Z J Psychiatry. 2013;47(2):169-176. doi:10.1177/0004867412466154

69. Dimidjian S, Beck A, Felder JN, Boggs JM, Gallop R, Segal Z V. Web-based Mindfulness-based Cognitive Therapy for reducing residual depressive symptoms: An open trial and quasi-experimental comparison to propensity score matched controls. Behav Res Ther. 2014;63:83-89. doi:10.1016/j.brat.2014.09.004

70. Dobbin A, Maxwell M, Elton R. A benchmarked feasibility study of a self-hypnosis treatment for depression in primary care. Int J Clin Exp Hypn. 2009;57(3):293-318. doi:10.1080/00207140902881221

71. Doherty G, Coyle D, Sharry J. Engagement with online mental health interventions: An exploratory clinical study of a treatment for depression. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ; 2012:1421. doi:10.1145/2207676.2208602

72. Donker T, Bennett K, Bennett A, et al. Internet-delivered interpersonal psychotherapy versus internet-delivered cognitive behavioral therapy for adults with depressive symptoms: Randomized controlled noninferiority trial. J Med Internet Res. 2013;15(5):1-16. doi:10.2196/jmir.2307

73. Drake G, Csipke E, Wykes T. Assessing your mood online: Acceptability and use of Moodscope. Psychol Med. 2013;43(7):1455-1464. doi:10.1017/S0033291712002280

74. Drozd F, Skeie LG, Kraft P, Kvale D. A web-based intervention trial for depressive symptoms and subjective well-being in patients with chronic HIV infection. AIDS Care - Psychol Socio-Medical Asp AIDS/HIV. 2014;26(9):1080-1089. doi:10.1080/09540121.2013.869541

75. Ebert DD, Lehr D, Boß L, et al. Efficacy of an internet-based problem-solving training for teachers: Results of a randomized controlled trial. Scand J Work Environ Heal. 2014;40(6):582-596. doi:10.5271/sjweh.3449

76. Eisma MC, Boelen PA, van den Bout J, et al. Internet-Based Exposure and Behavioral Activation for Complicated Grief and Rumination: A Randomized Controlled Trial. Behav Ther. 2015;46(6):729-748. doi:10.1016/j.beth.2015.05.007

77. Espinosa HD, Carrasco Á, Moessner M, et al. Acceptability Study of "Ascenso": An Online Program for Monitoring and Supporting Patients with Depression in Chile. Telemed e-Health. 2016;22(7):577-583. doi:10.1089/tmj.2015.0124

78. Falconer CJ, Rovira A, King JA, et al. Embodying self-compassion within virtual reality and its effects on patients with depression. BJPsych Open. 2016;2(1):74-80. doi:10.1192/bjpo.bp.115.002147

79. Farrer L, Christensen H, Griffiths KM, Mackinnon A. Internet-based CBT for depression with and without telephone tracking in a national helpline: Randomised controlled trial. PLoS One. 2011;6(11). doi:10.1371/journal.pone.0028099

80. Fischer A, Schröder J, Vettorazzi E, et al. An online programme to reduce depression in patients with multiple sclerosis: A randomised controlled trial. The Lancet Psychiatry. 2015;2(3):217-223. doi:10.1016/S2215-0366(14)00049-2

81. Fledderus M, Bohlmeijer ET, Fox JP, Schreurs KMG, Spinhoven P. The role of psychological flexibility in a self-help acceptance and commitment therapy intervention for psychological distress in a randomized controlled trial. Behav Res Ther. 2013;51(3):142-151. doi:10.1016/j.brat.2012.11.007

82. Gandy M, Karin E, Fogliati VJ, McDonald S, Titov N, Dear BF. A feasibility trial of an Internet-delivered and transdiagnostic cognitive behavioral therapy treatment program for anxiety, depression, and disability among adults with epilepsy. Epilepsia. 2016;57(11):1887-1896. doi:10.1111/epi.13569

83. Geisner IM, Varvil-Weld L, Mittmann AJ, Mallett K, Turrisi R. Brief web-based intervention for college students with comorbid risky alcohol use and depressed mood: Does it work and for whom? Addict Behav. 2015;42:36-43. doi:10.1016/j.addbeh.2014.10.030

84. Geraedts AS, Kleiboer AM, Twisk J, Wiezer NM, Van Mechelen W, Cuijpers P. Long-term results of a web-based guided self-help intervention for employees with depressive symptoms: Randomized controlled trial. J Med Internet Res. 2014;16(7):1-17. doi:10.2196/jmir.3539

85. Gilbody S, Littlewood E, Hewitt C, et al. Computerised cognitive behaviour therapy (cCBT) as treatment for depression in primary care (REEACT trial): Large scale pragmatic randomised controlled trial. BMJ. 2015;351:1-13. doi:10.1136/bmj.h5627

86. Glozier N, Christensen H, Naismith S, et al. Internet-Delivered Cognitive Behavioural Therapy for Adults with Mild to Moderate Depression and High Cardiovascular Disease Risks: A Randomised Attention-Controlled Trial. PLoS One. 2013;8(3):1-8. doi:10.1371/journal.pone.0059139

87. González-Robles A, Mira A, Díaz A, et al. Opinion of the Patients About an Internet-Based Psychological Treatment Protocol. In: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). ; 2015:460-466. doi:10.1007/978-3-319-26401-1_43

88. Griffiths KM, Bennett K, Walker J, Goldsmid S, Bennett A. Effectiveness of MH-Guru, a brief online mental health program for the workplace: A randomised controlled trial. Internet Interv. 2016;6:29-39. doi:10.1016/j.invent.2016.09.004

89. Griffiths KM, Mackinnon AJ, Crisp DA, Christensen H, Bennett K, Farrer L. The Effectiveness of an Online Support Group for Members of the Community with Depression: A Randomised Controlled Trial. PLoS One. 2012;7(12). doi:10.1371/journal.pone.0053244

90. Grime PR. Computerized cognitive behavioural therapy at work: A randomized controlled trial in employees with recent stress-related absenteeism. Occup Med (Chic Ill). 2004;54(5):353-359. doi:10.1093/occmed/kqh077

91. Hadjistavropoulos HD, Nugent MM, Alberts NM, Staples L, Dear BF, Titov N. Transdiagnostic Internet-delivered cognitive behaviour therapy in Canada: An open trial comparing results of a specialized online clinic and nonspecialized community clinics. J Anxiety Disord. 2016;42:19-29. doi:10.1016/j.janxdis.2016.05.006

92. Hadjistavropoulos HD, Pugh NE, Nugent MM, et al. Therapist-assisted Internet-delivered cognitive behavior therapy for depression and anxiety: Translating evidence into clinical practice. J Anxiety Disord. 2014;28(8):884-893. doi:10.1016/j.janxdis.2014.09.018

93. Hallgren M, Helgadóttir B, Herring MP, et al. Exercise and internet-based cognitive-behavioural therapy for depression: Multicentre randomized controlled trial with 12-month follow-up. Br J Psychiatry. 2016;209(5):414-420. doi:10.1192/bjp.bp.115.177576

94. Hetrick SE, Goodall J, Yuen HP, et al. Comprehensive online self- monitoring to support clinicians manage risk of suicide in youth depression: A pilot study. Crisis. 2017;38(3):147-157. doi:10.1027/0227-5910/a000422

95. Hinkle JF. The Stress Gym: An Online Intervention to Improve Stress and Depressive Symptoms in Adults. Issues Ment Health Nurs. 2015;36(11):870-876. doi:10.3109/01612840.2015.1074768

96. Høifødt RS, Lillevoll KR, Griffiths KM, et al. The clinical effectiveness of web-based cognitive behavioral therapy with face-to-face therapist support for depressed primary care patients: Randomized controlled trial. J Med Internet Res. 2013;15(8). doi:10.2196/jmir.2714

97. Høifødt RS, Mittner M, Lillevoll K, et al. Predictors of response to web-based cognitive behavioral therapy with high-intensity face-to-face therapist guidance for depression: A Bayesian analysis. J Med Internet Res. 2015;17(9):e197. doi:10.2196/jmir.4351

98. Holländare F, Johnsson S, Randestad M, et al. Randomized trial of Internet-based relapse prevention for partially remitted depression. Acta Psychiatr Scand. 2011;124(4):285-294. doi:10.1111/j.1600-0447.2011.01698.x

99. Hoorelbeke K, Faelens L, Behiels J, Koster EHW. Internet-delivered cognitive control training as a preventive intervention for remitted depressed patients: Protocol for a randomized controlled trial. BMC Psychiatry. 2015;15(1):125. doi:10.1186/s12888-015-0511-0

100. Hunkeler EM, Hargreaves WA, Fireman B, et al. A Web-Delivered Care Management and Patient Self-Management Program for Recurrent Depression: A Randomized Trial. Psychiatr Serv. 2013;63(11):1063-1071. doi:10.1176/appi.ps.005332011

101. Imamura K, Kawakami N, Furukawa TA, et al. Effects of an internet-based cognitive behavioral therapy (iCBT) program in manga format on improving subthreshold depressive symptoms among healthy workers: A randomized controlled trial. PLoS One. 2014;9(5):e97167. doi:10.1371/journal.pone.0097167

102. Imamura K, Kawakami N, Tsuno K, Tsuchiya M, Shimada K, Namba K. Effects of web-based stress and depression literacy intervention on improving symptoms and knowledge of depression among workers: A randomized controlled trial. J Affect Disord. 2016;203:30-37. doi:10.1016/j.jad.2016.05.045

103. Jakobsen H, Andersson G, Havik OE, Nordgreen T. Guided Internet-based cognitive behavioral therapy for mild and moderate depression: A benchmarking study. Internet Interv. 2017;7:1-8. doi:10.1016/j.invent.2016.11.002

104. Janevic MR, Aruquipa Yujra AC, Marinec N, et al. Feasibility of an interactive voice response system for monitoring depressive symptoms in a lower-middle income Latin American country. Int J Ment Health Syst. 2016;10(1):1-11. doi:10.1186/s13033-016-0093-3

105. Johansson R, Björklund M, Hornborg C, et al. Affect-focused psychodynamic psychotherapy for depression and anxiety through the Internet: a randomized controlled trial. PeerJ. 2013;1:e102. doi:10.7717/peerj.102

106. Johansson R, Ekbladh S, Hebert A, et al. Psychodynamic guided self-help for adult depression through the internet: A randomised controlled trial. PLoS One. 2012;7(5). doi:10.1371/journal.pone.0038021

107. Johansson R, Nyblom A, Carlbring P, Cuijpers P, Andersson G. Choosing between Internet-based psychodynamic versus cognitive behavioral therapy for depression: A pilot preference study. BMC Psychiatry. 2013. doi:10.1186/1471-244X-13-268

108. Johansson R, Sjöberg E, Sjögren M, et al. Tailored vs. Standardized internet-based cognitive behavior therapy for depression and comorbid symptoms: A randomized controlled trial. PLoS One. 2012;7(5):1-9. doi:10.1371/journal.pone.0036905

109. Johnston L, Dear BF, Gandy M, et al. Exploring the efficacy and acceptability of Internet-delivered cognitive behavioural therapy for young adults with anxiety and depression: An open trial. Aust N Z J Psychiatry. 2014;48(9):819-827. doi:10.1177/0004867414527524

110. Joutsenniemi K. E-mail-based Exercises in Happiness, Physical Activity and Readings: A Randomized Trial on 3274 Finns. J Psychiatry. 2015;17(5). doi:10.4172/psychiatry.1000140

111. Kawai K, Yamazaki Y, Nakayama K. Process Evaluation of a Web-based Stress Management Program to Promote Psychological Well-being in a Sample of White-collar Workers in Japan. Ind Health. 2010. doi:10.2486/indhealth.48.265

112. Kay-Lambkin FJ, Baker AL, Kelly B, Lewin TJ. Clinician-assisted computerised versus therapist-delivered treatment for depressive and addictive disorders: A

randomised controlled trial. Med J Aust. 2011. doi:10.5694/j.1326-5377.2011.tb03265.x

113. Kay-Lambkin FJ, Baker AL, Lewin TJ, Carr VJ. Computer-based psychological treatment for comorbid depression and problematic alcohol and/or cannabis use: A randomized controlled trial of clinical efficacy. Addiction. 2009;104(3):378-388. doi:10.1111/j.1360-0443.2008.02444.x

114. Kayrouz R, Dear BF, Karin E, Fogliati VJ, Titov N. A pilot study of a clinician-guided internet-delivered cognitive behavioural therapy for anxiety and depression among Arabs in Australia, presented in both English and Arabic languages. Internet Interv. 2016;5:5-11. doi:10.1016/j.invent.2016.06.002

115. Kayrouz R, Dear BF, Karin E, et al. A pilot study of self-guided internet-delivered cognitive behavioural therapy for anxiety and depression among Arabs. Internet Interv. 2016;3:18-24. doi:10.1016/j.invent.2015.10.005

116. Kelders SM, Bohlmeijer ET, Pots WTM, van Gemert-Pijnen JEWC. Comparing human and automated support for depression: Fractional factorial randomized controlled trial. Behav Res Ther. 2015;72:72-80. doi:10.1016/j.brat.2015.06.014

117. Kenter RMF, Cuijpers P, Beekman A, van Straten A. Effectiveness of a Web-Based Guided Self-help Intervention for Outpatients With a Depressive Disorder: Short-term Results From a Randomized Controlled Trial. J Med Internet Res. 2016;18(3):e80. doi:10.2196/jmir.4861

118. Kenter R, Warmerdam L, Brouwer-Dudokdewit C, Cuijpers P, van Straten A. Guided online treatment in routine mental health care: An observational study on uptake, drop-out and effects. BMC Psychiatry. 2013;13(1):43. doi:10.1186/1471-244X-13-43

119. Killen A, Macaskill A. Using a Gratitude Intervention to Enhance Well-Being in Older Adults. J Happiness Stud. 2015;16(4):947-964. doi:10.1007/s10902-014-9542-3

120. Kirkpatrick T, Manoukian L, Dear BF, Johnston L, Titov N. A feasibility open trial of internet-delivered cognitive-behavioural therapy (iCBT) among consumers of a non-governmental mental health organisation with anxiety. PeerJ. 2013. doi:10.7717/peerj.210

121. Kleiboer A, Donker T, Seekles W, van Straten A, Riper H, Cuijpers P. A randomized controlled trial on the role of support in Internet-based problem solving therapy for depression and anxiety. Behav Res Ther. 2015;72(July):63-71. doi:10.1016/j.brat.2015.06.013

122. Klein JP, Berger T, Schr’der J, et al. Effects of a psychological internet intervention in the treatment of mild to moderate depressive symptoms: Results of the evident study, a randomized controlled trial. Psychother Psychosom. 2016;85(4):218-228. doi:10.1159/000445355

123. Kok G, Burger H, Riper H, et al. The three-month effect of mobile internet-based cognitive therapy on the course of depressive symptoms in remitted recurrently depressed patients: Results of a randomized controlled trial. Psychother Psychosom. 2015;84(2):90-99. doi:10.1159/000369469

124. Kooistra LC, Ruwaard J, Wiersma JE, et al. Development and initial evaluation of blended cognitive behavioural treatment for major depression in routine specialized mental health care. Internet Interv. 2016;4:61-71. doi:10.1016/j.invent.2016.01.003

125. Kordy H, Wolf M, Aulich K, et al. Internet-Delivered Disease Management for Recurrent Depression: A Multicenter Randomized Controlled Trial. Psychother Psychosom. 2016;85(2):91-98. doi:10.1159/000441951

126. Kraaij V, van Emmerik A, Garnefski N, et al. Effects of a cognitive behavioral self-help program and a computerized structured writing intervention on depressed mood for HIV-infected people: A pilot randomized controlled trial. Patient Educ Couns. 2010;80(2):200-204. doi:10.1016/j.pec.2009.08.014

127. Kraft S, Wolf M, Klein T, Becker T, Bauer S, Puschner B. Text Message Feedback to Support Mindfulness Practice in People With Depressive Symptoms: A Pilot

B

Randomized Controlled Trial. JMIR mHealth uHealth. 2017;5(5):e59. doi:10.2196/mhealth.7095

128. Kroenke K, Theobald D, Wu J, et al. Effect of Telecare Management on Pain and Depression in Patients With Cancer. JAMA. 2010;304(2):163. doi:10.1001/jama.2010.944

129. Krusche A, Cyhlarova E, Williams JMG. Mindfulness online: An evaluation of the feasibility of a web-based mindfulness course for stress, anxiety and depression. BMJ Open. 2013;3(11). doi:10.1136/bmjopen-2013-003498

130. Lai T-Y, Larson EL, Rockoff ML, Bakken S. User Acceptance of HIV TIDES--Tailored Interventions for Management of Depressive Symptoms in Persons Living with HIV/AIDS. J Am Med Informatics Assoc. 2008;15(2):217-226. doi:10.1197/jamia.M2481

131. Lai T-Y. Iterative refinement of a tailored system for self-care management of depressive symptoms in people living with HIV/AIDS through heuristic evaluation and end user testing. Int J Med Inform. 2007;76:S317-S324. doi:10.1016/j.ijmedinf.2007.05.007

132. Lai T-Y, Bakken S. Heuristic evaluation of HIV-TIDES - Tailored Interventions for management of DEpressive Symptoms in HIV-infected individuals. AMIA . Annu Symp proceedings AMIA Symp. 2006:996.

133. Lang TJ, Blackwell SE, Harmer CJ, Davison P, Holmes EA. Cognitive Bias Modification Using Mental Imagery for Depression: Developing a Novel Computerized Intervention to Change Negative Thinking Styles. Eur J Pers. 2012;26(2):145-157. doi:10.1002/per.855

134. Lappalainen P, Granlund A, Siltanen S, et al. ACT Internet-based vs face-to-face? A randomized controlled trial of two ways to deliver Acceptance and Commitment Therapy for depressive symptoms: An 18-month follow-up. Behav Res Ther. 2014;61(July 2015):43-54. doi:10.1016/j.brat.2014.07.006

135. Lappalainen P, Langrial S, Oinas-Kukkonen H, Tolvanen A, Lappalainen R. Web-Based Acceptance and Commitment Therapy for Depressive Symptoms With Minimal Support: A Randomized Controlled Trial. Behav Modif. 2015;39(6):805-834. doi:10.1177/0145445515598142

136. Lara MA, Tiburcio M, Aguilar Abrego A, Sánchez-Solís A. A four-year experience with a Web-based self-help intervention for depressive symptoms in Mexico. Rev Panam Salud Publica. 2014;35(5-6):399-406. http://www.ncbi.nlm.nih.gov/pubmed/25211568.

137. Learmonth D, Rai S. Taking computerized CBT beyond primary care. Br J Clin Psychol. 2008;47(1):111-118. doi:10.1348/014466507X248599

138. Lee JS, Mathews A, Shergill S, Yiu Chan DK, Majeed N, Yiend J. How can we enhance cognitive bias modification techniques? The effects of prospective cognition. J Behav Ther Exp Psychiatry. 2015;49(March):120-127. doi:10.1016/j.jbtep.2015.03.007

139. Lee SW, Kim I, Yoo J, Park S, Jeong B, Cha M. Insights from an expressive writing intervention on Facebook to help alleviate depressive symptoms. Comput Human Behav. 2016;62:613-619. doi:10.1016/j.chb.2016.04.034

140. Lemma A, Fonagy P. Feasibility study of a psychodynamic online group intervention for depression. Psychoanal Psychol. 2013;30(3):367-380. doi:10.1037/a0033239

141. Levin W, Campbell DR, McGovern KB, et al. A computer-assisted depression intervention in primary care. Psychol Med. 2011;41(7):1373-1383. doi:10.1017/S0033291710001935

142. Leykin Y, Muñoz RF, Contreras O, Latham MD. Results from a trial of an unsupported internet intervention for depressive symptoms. Internet Interv. 2014;1(4):175-181. doi:10.1016/j.invent.2014.09.002

143. Lillevoll KR, Vangberg HCB, Griffiths KM, Waterloo K, Eisemann MR. Uptake and adherence of a self-directed internet-based mental health intervention with

tailored e-mail reminders in senior high schools in Norway. BMC Psychiatry. 2014. doi:10.1186/1471-244X-14-14

144. Lillevoll KR, Wilhelmsen M, Kolstrup N, et al. Patients' experiences of helpfulness in guided internet-based treatment for depression: Qualitative study of integrated therapeutic dimensions. J Med Internet Res. 2013;15(6):1-12. doi:10.2196/jmir.2531

145. Lin MF, Moyle W, Chang HJ, Chou MH, Hsu MC. Effect of an interactive computerized psycho-education system on patients suffering from depression. J Clin Nurs. 2008;17(5):667-676. doi:10.1111/j.1365-2702.2007.02085.x

146. Lintvedt OK, Griffiths KM, Sørensen K, et al. Evaluating the effectiveness and efficacy of unguided internet-based self-help intervention for the prevention of depression: A randomized controlled trial. Clin Psychol Psychother. 2013;20(1):10-27. doi:10.1002/cpp.770

147. Littlewood E, Duarte A, Hewitt C, et al. A randomised controlled trial of computerised cognitive behaviour therapy for the treatment of depression in primary care: The Randomised Evaluation of the Effectiveness and Acceptability of Computerised Therapy (REEACT) trial. Health Technol Assess (Rockv). 2015;19(101). doi:10.3310/hta191010

148. Løventoft PK, Nørregaard LB, Frøkjær E. Designing daybuilder: An Experimental App to Support People with Depression. Proc 12th Particip Des Conf Explor Pap Work Descr Ind Cases - Vol 2 - PDC '12. 2012;(May 2014):1. doi:10.1145/2348144.2348146

149. Lucassen MFG, Merry SN, Hatcher S, Frampton CMA. Rainbow SPARX: A novel approach to addressing depression in sexual minority youth. Cogn Behav Pract. 2015;22(2):203-216. doi:10.1016/j.cbpra.2013.12.008

150. Lundgren JG, Dahlström Ö, Andersson G, Jaarsma T, Kärner Köhler A, Johansson P. The Effect of Guided Web-Based Cognitive Behavioral Therapy on Patients With Depressive Symptoms and Heart Failure: A Pilot Randomized Controlled Trial. J Med Internet Res. 2016;18(8):e194. doi:10.2196/jmir.5556

151. Lundgren J, Andersson G, Dahlström Ö, Jaarsma T, Köhler AK, Johansson P. Internet-based cognitive behavior therapy for patients with heart failure and depressive symptoms: A proof of concept study. Patient Educ Couns. 2015;98(8):935-942. doi:10.1016/j.pec.2015.04.013

152. Ly KH, Topooco N, Cederlund H, et al. Smartphone-supported versus full behavioural activation for depression: A randomised controlled trial. PLoS One. 2015;10(5):1-16. doi:10.1371/journal.pone.0126559

153. Ly KH, Trüschel A, Jarl L, et al. Behavioural activation versus mindfulness-based guided self-help treatment administered through a smartphone application: a randomised controlled trial. BMJ Open. 2014;4(1):e003440. doi:10.1136/bmjopen-2013-003440

154. Månsson KNT, Ruiz ES, Gervind E, Dahlin M, Andersson G. Development and initial evaluation of an internet-based support system for face-to-face cognitive behavior therapy: A proof of concept study. J Med Internet Res. 2013;15(12). doi:10.2196/jmir.3031

155. McCombie A, Gearry R, Andrews J, Mulder R, Mikocka-Walus A. Does Computerized Cognitive Behavioral Therapy Help People with Inflammatory Bowel Disease? A Randomized Controlled Trial. Inflamm Bowel Dis. 2016;22(1):171-181. doi:10.1097/MIB.0000000000000567

156. McMurchie W, Macleod F, Power K, Laidlaw K, Prentice N. Computerised cognitive behavioural therapy for depression and anxiety with older people: a pilot study to examine patient acceptability and treatment outcome. Int J Geriatr Psychiatry. 2013:n/a-n/a. doi:10.1002/gps.3935

157. Meglic M, Furlan M, Kuzmanic M, et al. Feasibility of an eHealth service to support collaborative depression care: Results of a pilot study. J Med Internet Res. 2010;12(5):1-11. doi:10.2196/jmir.1510

158. Meiser B, Peate M, Levitan C, et al. A Psycho-Educational Intervention for People with a Family History of Depression: Pilot Results. J Genet Couns. 2017;26(2):312-321. doi:10.1007/s10897-016-0011-5

159. Melnyk BM, Amaya M, Szalacha LA, Hoying J, Taylor T, Bowersox K. Feasibility, Acceptability, and Preliminary Effects of the COPE Online Cognitive-Behavioral Skill-Building Program on Mental Health Outcomes and Academic Performance in Freshmen College Students: A Randomized Controlled Pilot Study. J Child Adolesc Psychiatr Nurs. 2015;28(3):147-154. doi:10.1111/jcap.12119

160. Merry SN, Stasiak K, Shepherd M, Frampton C, Fleming T, Lucassen MFG. The effectiveness of SPARX, a computerised self help intervention for adolescents seeking help for depression: Randomised controlled non-inferiority trial. BMJ. 2012;344(7857):1-16. doi:10.1136/bmj.e2598

161. Mewton L, Hobbs MJ, Sunderland M, Newby J, Andrews G. Reductions in the internalising construct following internet-delivered treatment for anxiety and depression in primary care. Behav Res Ther. 2014;63(March 2016):132-138. doi:10.1016/j.brat.2014.10.001

162. Mewton L, Andrews G. Cognitive behaviour therapy via the internet for depression: A useful strategy to reduce suicidal ideation. J Affect Disord. 2015;170:78-84. doi:10.1016/j.jad.2014.08.038

163. Meyer B, Berger T, Caspar F, Beevers CG, Andersson G, Weiss. Effectiveness of a novel integrative online treatment for depression (Deprexis): Randomized field trial. 2009:1-8.

164. Mira A, Bretón-López J, García-Palacios A, Baños RM, Botella C. Pervasive Computing Paradigms for Mental Health. Vol 604. (Serino S, Matic A, Giakoumis D, Lopez G, Cipresso P, eds.). Cham: Springer International Publishing; 2016. doi:10.1007/978-3-319-32270-4

165. Möbius M, Tendolkar I, Lohner V, Baltussen M, Becker ES. Refilling the half-empty glass - Investigating the potential role of the Interpretation Modification Paradigm for Depression (IMP-D). J Behav Ther Exp Psychiatry. 2015;49:37-43. doi:10.1016/j.jbtep.2015.03.002

166. Mogoaşe C, Brăilean A, David D. Can concreteness training alone reduce depressive symptoms? A randomized pilot study using an internet-delivered protocol. Cognit Ther Res. 2013;37(4):704-712. doi:10.1007/s10608-012-9514-z

167. Mohr DC, Duffecy J, Ho J, et al. A Randomized Controlled Trial Evaluating a Manualized TeleCoaching Protocol for Improving Adherence to a Web-Based Intervention for the Treatment of Depression. PLoS One. 2013;8(8). doi:10.1371/journal.pone.0070086

168. Mohr DC, Duffecy J, Jin L, et al. Multimodal e-mental health treatment for depression: A feasibility trial. J Med Internet Res. 2010;12(5):1-12. doi:10.2196/jmir.1370

169. Montero-Marín J, Araya R, Pérez-Yus MC, et al. An internet-based intervention for depression in primary care in Spain: A randomized controlled trial. J Med Internet Res. 2016;18(8):1-12. doi:10.2196/jmir.5695

170. Morgan AJ, Jorm AF, Mackinnon AJ. Email-based promotion of self-help for subthreshold depression: Mood Memos randomised controlled trial. Br J Psychiatry. 2012;200(5):412-418. doi:10.1192/bjp.bp.111.101394

171. Morgan AJ, Jorm AF, Mackinnon AJ. Self-Help for Depression via E-mail: A Randomised Controlled Trial of Effects on Depression and Self-Help Behaviour. PLoS One. 2013;8(6):1-10. doi:10.1371/journal.pone.0066537

172. Moritz S, Quan H, Rickhi B, et al. A home study-based spirituality education program decreases emotional distress and increases quality of life - A randomized, controlled trial. Altern Ther Health Med. 2006.

173. Moritz S, Schilling L, Hauschildt M, Schröder J, Treszl A. A randomized controlled trial of internet-based therapy in depression. Behav Res Ther. 2012;50(7-8):513-521. doi:10.1016/j.brat.2012.04.006

174. Morris RR, Schueller SM, Picard RW. Efficacy of a web-based, crowdsourced peer-to-peer cognitive reappraisal platform for depression: Randomized controlled trial. J Med Internet Res. 2015. doi:10.2196/jmir.4167

175. Mullin A, Dear BF, Karin E, et al. The UniWellbeing course: A randomised controlled trial of a transdiagnostic internet-delivered cognitive behavioural therapy (CBT) programme for university students with symptoms of anxiety and depression. Internet Interv. 2015;2(2):128-136. doi:10.1016/j.invent.2015.02.002

176. Nelson CB, Abraham KM, Walters H, Pfeiffer PN, Valenstein M. Integration of peer support and computer-based CBT for veterans with depression. Comput Human Behav. 2014;31(1):57-64. doi:10.1016/j.chb.2013.10.012

177. Newby JM, Mackenzie A, Williams AD, et al. Internet cognitive behavioural therapy for mixed anxiety and depression: A randomized controlled trial and evidence of effectiveness in primary care. Psychol Med. 2013;43(12):2635-2648. doi:10.1017/S0033291713000111

178. Newby JM, Lang T, Werner-Seidler A, Holmes E, Moulds ML. Alleviating distressing intrusive memories in depression: A comparison between computerised cognitive bias modification and cognitive behavioural education. Behav Res Ther. 2014;56(1):60-67. doi:10.1016/j.brat.2014.03.001

179. Newby JM, Mewton L, Andrews G. Transdiagnostic versus disorder-specific internet-delivered cognitive behaviour therapy for anxiety and depression in primary care. J Anxiety Disord. 2017;46:25-34. doi:10.1016/j.janxdis.2016.06.002

180. Newby JM, Mewton L, Williams AD, Andrews G. Effectiveness of transdiagnostic internet cognitive behavioural treatment for mixed anxiety and depression in primary care. J Affect Disord. 2014;165(May):45-52. doi:10.1016/j.jad.2014.04.037

181. Newby J, Robins L, Wilhelm K, et al. Web-based cognitive behavior therapy for depression in people with diabetes mellitus:A randomized controlled trial. J Med Internet Res. 2017;19(5). doi:10.2196/jmir.7274

182. Nicky M. Computerised CBT self-help for depression in higher education: Reflections on a pilot. Couns Psychother Res. 2009;9(4):280-286. doi:10.1080/14733140902993343

183. Nobis S, Lehr D, Ebert DD, et al. Efficacy of a web-based intervention with mobile phone support in treating depressive symptoms in adults with type 1 and type 2 diabetes: A randomized controlled trial. Diabetes Care. 2015;38(5):776-783. doi:10.2337/dc14-1728

184. Nørregaard LB, Løventoft PK, Frøkjær E, et al. Patient expectations and experiences from a clinical study in psychiatric care using a self-monitoring system. In: Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational. ; 2014:991-994. doi:10.1145/2639189.2670258

185. O' Leary K, Dockray S. The Effects of Two Novel Gratitude and Mindfulness Interventions on Well-Being. J Altern Complement Med. 2015;21(4):243-245. doi:10.1089/acm.2014.0119

186. Paredes P, Giald-Bachrach R, Czerwinski M, Roseway A, Rowan K, Hernandez J. PopTherapy: Coping with Stress through Pop-Culture. 2014. doi:10.4108/icst.pervasivehealth.2014.255070

187. Patten SB. Prevention of Depressive Symptoms Through the Use of Distance Technologies. Psychiatr Serv. 2003;54(3):396-398. doi:10.1176/appi.ps.54.3.396

188. Perini S, Titov N, Andrews G. Clinician-assisted Internet-based treatment is effective for depression: Randomized controlled trial. Aust N Z J Psychiatry. 2009;43:571-578.

189. Perini S, Titov N, Andrews G. The Climate Sadness program: an open trial of Internet-based treatment for depression. E-Journal Appl Psychol. 2013;4(2):18-24. doi:10.7790/ejap.v4i2.135

B

B

190. Peters KD, Constans JI, Mathews A. Experimental modification of attribution processes. J Abnorm Psychol. 2011;120(1):168-173. doi:10.1037/a0021899

191. Pfeiffer PN, Henry J, Ganoczy D, Piette JD. Pilot study of psychotherapeutic text messaging for depression. J Telemed Telecare. 2017;23(7):665-672. doi:10.1177/1357633X16659955

192. Phillips R, Schneider J, Molosankwe I, et al. Randomized controlled trial of computerized cognitive behavioural therapy for depressive symptoms: Effectiveness and costs of a workplace intervention. Psychol Med. 2014;44(4):741-752. doi:10.1017/S0033291713001323

193. Pictet A, Jermann F, Ceschi G. When less could be more: Investigating the effects of a brief internet-based imagery cognitive bias modification intervention in depression. Behav Res Ther. 2016;84(August):45-51. doi:10.1016/j.brat.2016.07.008

194. Piette JD, Aikens JE, Trivedi R, et al. Depression self-management assistance using automated telephonic assessments and social support. Am J Manag Care. 2013;19(11):892-900. http://www.ncbi.nlm.nih.gov/pubmed/24511987.

195. Pittaway PGCE S, Arowobusoye MFPH N, Milne RB, et al. Comparative, clinical feasibility study of three tools for delivery of cognitive behavioural therapy for mild to moderate depression and anxiety provided on a self-help basis. Ment Health Fam Med. 2009;6(September 2009):145-154. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2838647/pdf/MHFM-06-145.pdf.

196. Pots WTM, Fledderus M, Meulenbeek PAM, ten Klooster PM, Schreurs KMG, Bohlmeijer ET. Acceptance and commitment therapy as a web-based intervention for depressive symptoms: randomised controlled trial. Br J Psychiatry. 2016;208(1):69-77. doi:10.1192/bjp.bp.114.146068

197. Preschl B, Maercker A, Wagner B. The working alliance in a randomized controlled trial comparing online. BMC Psychiatry. 2011.

198. Proudfoot J, Swain S, Widmer S, et al. The development and beta-test of a computer-therapy program for anxiety and depression: Hurdles and lessons. Comput Human Behav. 2003;19(3):277-289. doi:10.1016/S0747-5632(02)00062-6

199. Proudfoot J, Clarke J, Birch M-R, et al. Impact of a mobile phone and web program on symptom and functional outcomes for people with mild-to-moderate depression, anxiety and stress: a randomised controlled trial. BMC Psychiatry. 2013;13(1):312. doi:10.1186/1471-244X-13-312

200. Proudfoot J, Goldberg D, Mann A, Everitt B, Marks I, Gray JA. Computerized, interactive, multimedia cognitive-behavioural program for anxiety and depression in general practice. Psychol Med. 2003;33(2):217-227. doi:10.1017/S0033291702007225

201. Proudfoot J, Ryden C, Everitt B, et al. Clinical efficacy of computerised cognitive-behavioural therapy for anxiety and depression in primary care: randomised controlled trial. Br J Psychiatry. 2004;185:46-54. http://www.ncbi.nlm.nih.gov/pubmed/15231555.

202. Proyer RT, Gander F, Wellenzohn S, Ruch W. Nine beautiful things: A self-administered online positive psychology intervention on the beauty in nature, arts, and behaviors increases happiness and ameliorates depressive symptoms. Pers Individ Dif. 2016;94:189-193. doi:10.1016/j.paid.2016.01.028

203. Proyer RT, Gander F, Wellenzohn S, Ruch W. Positive psychology interventions in people aged 50-79 years: Long-term effects of placebo-controlled online interventions on well-being and depression. Aging Ment Heal. 2014;18(8):997-1005. doi:10.1080/13607863.2014.899978

204. Purves DG, Bennett M, Wellman N. An Open Trial in the NHS of Blues Begone®: A New Home Based Computerized CBT Program. Behav Cogn Psychother. 2009;37(5):541-551. doi:10.1017/s1352465809990282

205. Räsänen P, Lappalainen P, Muotka J, Tolvanen A, Lappalainen R. An online guided ACT intervention for enhancing the psychological wellbeing of university students: A randomized controlled clinical trial. Behav Res Ther. 2016;78:30-42. doi:10.1016/j.brat.2016.01.001

206. Rebar AL, Boles C, Burton N, et al. Healthy mind, healthy body: A randomized trial testing the efficacy of a computer-tailored vs. interactive web-based intervention for increasing physical activity and reducing depressive symptoms. Ment Health Phys Act. 2016;11:29-37. doi:10.1016/j.mhpa.2016.08.001

207. Rice S, Gleeson J, Davey C, et al. Moderated online social therapy for depression relapse prevention in young people: pilot study of a 'next generation' online intervention. Early Interv Psychiatry. 2018;12(4):613-625. doi:10.1111/eip.12354

208. Richards D, Timulak L, O'Brien E, et al. A randomized controlled trial of an internet-delivered treatment: Its potential as a low-intensity community intervention for adults with symptoms of depression. Behav Res Ther. 2015;75(October):20-31. doi:10.1016/j.brat.2015.10.005

209. Richards D, Murphy T, Viganó N, et al. Acceptability, satisfaction and perceived efficacy of "Space from Depression" an internet-delivered treatment for depression. Internet Interv. 2016;5:12-22. doi:10.1016/j.invent.2016.06.007

210. Richards D, Timulak L, Hevey D. A comparison of two online cognitive-behavioural interventions for symptoms of depression in a student population: The role of therapist responsiveness. Couns Psychother Res. 2013;13(3):184-193. doi:10.1080/14733145.2012.733715

211. Robertson L, Smith M, Tannenbaum D. Case management and adherence to an online disease management system. J Telemed Telecare. 2005;11(2_suppl):73-75. doi:10.1258/135763305775124885

212. Robinson J, Hetrick S, Cox G, et al. Can an Internet-based intervention reduce suicidal ideation, depression and hopelessness among secondary school students: Results from a pilot study. Early Interv Psychiatry. 2016;10(1):28-35. doi:10.1111/eip.12137

213. Roepke AM, Jaffee SR, Riffle OM, McGonigal J, Broome R, Maxwell B. Randomized Controlled Trial of SuperBetter, a Smartphone-Based/Internet-Based Self-Help Tool to Reduce Depressive Symptoms. Games Health J. 2015;4(3):235-246. doi:10.1089/g4h.2014.0046

214. Rosso IM, Killgore WDS, Olson EA, et al. Internet-based cognitive behavior therapy for major depressive disorder: A randomized controlled trial. Depress Anxiety. 2017;34(3):236-245. doi:10.1002/da.22590

215. Ruwaard J, Schrieken B, Schrijver M, et al. Standardized Web-Based Cognitive Behavioural Therapy of Mild to Moderate Depression: A Randomized Controlled Trial with a Long-Term Follow-Up. Cogn Behav Ther. 2009;38(4):206-221. doi:10.1080/16506070802408086

216. Salisbury C, O'Cathain A, Edwards L, et al. Effectiveness of an integrated telehealth service for patients with depression: A pragmatic randomised controlled trial of a complex intervention. The Lancet Psychiatry. 2016;3(6):515-525. doi:10.1016/S2215-0366(16)00083-3

217. Salisbury C, Thomas C, O'Cathain A, et al. TElehealth in CHronic disease: Mixed-methods study to develop the TECH conceptual model for intervention design and evaluation. BMJ Open. 2015;5(2):1-12. doi:10.1136/bmjopen-2014-006448

218. Santucci LC, McHugh RK, Elkins RM, et al. Pilot implementation of computerized cognitive behavioral therapy in a university health setting. Adm Policy Ment Heal Ment Heal Serv Res. 2014;41(4):514-521. doi:10.1007/s10488-013-0488-2

219. Schneider J, Foroushani PS, Grime P, Thornicroft G. Acceptability of online self-help to people with depression: Users' views of moodgym versus informational websites. J Med Internet Res. 2014;16(3):1-13. doi:10.2196/jmir.2871

220. Schröder J, Brückner K, Fischer A, et al. Efficacy of a psychological online intervention for depression in people with epilepsy: A randomized controlled trial. Epilepsia. 2014;55(12):2069-2076. doi:10.1111/epi.12833

B

221. Schueller S, Mohr D. Initial Field Trial of a Coach-Supported Web-Based Depression Treatment. In: Proceedings of the 9th International Conference on Pervasive Computing Technologies for Healthcare. ICST; 2015. doi:10.4108/icst.pervasivehealth.2015.260115

222. Seligman MEP, Steen TA, Park N, Peterson C. Positive Psychology Progress: Empirical Validation of Interventions. Am Psychol. 2005;60(5):410-421. doi:10.1037/0003-066X.60.5.410

223. Seligman MEP, Schulman P, Tryon AM. Group prevention of depression and anxiety symptoms. Behav Res Ther. 2007;45(6):1111-1126. doi:10.1016/j.brat.2006.09.010

224. Sergeant S, Mongrain M. An online optimism intervention reduces depression in pessimistic individuals. J Consult Clin Psychol. 2014;82(2):263-274. doi:10.1037/a0035536

225. Sethi S. Treating youth depression and anxiety: A randomised controlled trial examining the efficacy of computerised versus face-to-face cognitive behaviour therapy. Aust Psychol. 2013;48(4):249-257. doi:10.1111/ap.12006

226. Shamekhi A, Bickmore T, Lestoquoy A, Gardiner P. Augmenting Group Medical Visits with Conversational Agents for Stress Management Behavior Change. 2017:55-67. doi:10.1007/978-3-319-55134-0_5

227. Shamekhi A, Bickmore T, Lestoquoy A, Negash L, Gardiner P. Blissful Agents: Adjuncts to Group Medical Visits for Chronic Pain and Depression. In: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). ; 2016:433-437. doi:10.1007/978-3-319-47665-0_49

228. Shapira LB, Mongrain M. The benefits of self-compassion and optimism exercises for individuals vulnerable to depression. J Posit Psychol. 2010;5(5):377-389. doi:10.1080/17439760.2010.516763

229. Sharry J, Davidson R, McLoughlin O, Doherty G. A service-based evaluation of a therapist-supported online cognitive behavioral therapy program for depression. J Med Internet Res. 2013;15(6):1-14. doi:10.2196/jmir.2248

230. Skolarus LE, Piette JD, Pfeiffer PN, et al. Interactive Voice Response—An Innovative Approach to Post-Stroke Depression Self-Management Support. Transl Stroke Res. 2017;8(1):77-82. doi:10.1007/s12975-016-0481-7

231. Spates CR, Kalata AH, Ozeki S, Stanton CE, Peters S. Initial Open Trial of a Computerized Behavioral Activation Treatment for Depression. Behav Modif. 2013;37(3):259-297. doi:10.1177/0145445512455051

232. Spek V, Nyklicek I, Smits N, et al. Internet-based cognitive behavioural therapy for subthreshold depression in people over 50 years old: a randomized controlled clinical trial. Psychol Med. 2007;37(12):1797-1806. doi:10.1017/S0033291707000542

233. Staples LG, Fogliati VJ, Dear BF, Nielssen O, Titov N. Internet-delivered treatment for older adults with anxiety and depression: implementation of the Wellbeing Plus Course in routine clinical care and comparison with research trial outcomes. BJPsych Open. 2016;2(5):307-313. doi:10.1192/bjpo.bp.116.003400

234. Ström M, Uckelstam C-J, Andersson G, Hassmén P, Umefjord G, Carlbring P. Internet-delivered therapist-guided physical activity for mild to moderate depression: a randomized controlled trial. PeerJ. 2013;1(2013):e178. doi:10.7717/peerj.178

235. Thompson NJ, Walker ER, Obolensky N, et al. Distance delivery of mindfulness-based cognitive therapy for depression: Project UPLIFT. Epilepsy Behav. 2010;19(3):247-254. doi:10.1016/j.yebeh.2010.07.031

236. Tiburcio M, Lara MA, Aguilar Abrego A, Fernández M, Martínez Vélez N, Sánchez A. Web-Based Intervention to Reduce Substance Abuse and Depressive Symptoms in Mexico: Development and Usability Test. JMIR Ment Heal. 2016;3(3):e47. doi:10.2196/mental.6001

B

237. Titov N, Dear BF, Staples LG, et al. Disorder-specific versus transdiagnostic and clinician-guided versus self-guided treatment for major depressive disorder and comorbid anxiety disorders: A randomized controlled trial. J Anxiety Disord. 2015;35:88-102. doi:10.1016/j.janxdis.2015.08.002

238. Titov N, Andrews G, Davies M, McIntyre K, Robinson E, Solley K. Internet Treatment for Depression: A Randomized Controlled Trial Comparing Clinician vs. Technician Assistance. PLoS One. 2013;5(6):1-9. doi:10.1371/journal.pone.0010939

239. Titov N, Dear BF, Johnston L, et al. Improving adherence and clinical outcomes in self-guided internet treatment for anxiety and depression: a 12-month follow-up of a randomised controlled trial. PLoS One. 2014;9(2):e89591. doi:10.1371/journal.pone.0089591

240. Titov N, Dear BF, Johnston L, et al. Improving Adherence and Clinical Outcomes in Self-Guided Internet Treatment for Anxiety and Depression: Randomised Controlled Trial. PLoS One. 2013;8(7). doi:10.1371/journal.pone.0062873

241. Titov N, Dear BF, Schwencke G, et al. Transdiagnostic internet treatment for anxiety and depression: A randomised controlled trial. Behav Res Ther. 2011;49(8):441-452. doi:10.1016/j.brat.2011.03.007

242. Titov N, Fogliati VJ, Staples LG, et al. Treating anxiety and depression in older adults: randomised controlled trial comparing guided V. self-guided internet-delivered cognitive–behavioural therapy. BJPsych Open. 2016;2(1):50-58. doi:10.1192/bjpo.bp.115.002139

243. Titov, N, Dear, et al. Clinical and cost-effectiveness of therapist-guided internet-delivered cognitive behavior therapy for older adults with symptoms of depression: a randomized controlled trial (Provisional abstract). Behav Ther. 2014;46(2):206-217.

244. Topolovec-Vranic J, Cullen N, Michalak A, et al. Evaluation of an online cognitive behavioural therapy program by patients with traumatic brain injury and depression. Brain Inj. 2010;24(5):762-772. doi:10.3109/02699051003709599

245. Torkan H, Blackwell SE, Holmes EA, et al. Positive imagery cognitive bias modification in treatment-seeking patients with major depression in Iran: A pilot study. Cognit Ther Res. 2014;38(2):132-145. doi:10.1007/s10608-014-9598-8

246. Twomey C, O'Reilly G, Byrne M, et al. A randomized controlled trial of the computerized CBT programme, MoodGYM, for public mental health service users waiting for interventions. Br J Clin Psychol. 2014;53(4):433-450. doi:10.1111/bjc.12055

247. Ünlü Ince B, Cuijpers P, Van'T Hof E, Van Ballegooijen W, Christensen H, Riper H. Internet-based, culturally sensitive, problem-solving therapy for Turkish migrants with depression: Randomized controlled trial. J Med Internet Res. 2013;15(10):1-16. doi:10.2196/jmir.2853

248. Van Bastelaar KMP, Pouwer F, Cuijpers P, Riper H, Snoek FJ. Web-based depression treatment for type 1 and type 2 diabetic patients: A randomized, controlled trial. Diabetes Care. 2011;34(2):320-325. doi:10.2337/dc10-1248

249. Van Den Berg S, Shapiro DA, Bickerstaffe D, Cavanagh K. Computerized cognitive-behaviour therapy for anxiety and depression: A practical solution to the shortage of trained therapists. J Psychiatr Ment Health Nurs. 2004;11(5):508-513. doi:10.1111/j.1365-2850.2004.00745.x

250. Van Der Zanden R, Kramer J, Gerrits R, Cuijpers P. Effectiveness of an online group course for depression in adolescents and young adults: A randomized trial. J Med Internet Res. 2012;14(3):1-15. doi:10.2196/jmir.2033

251. Van Dongen AJCM, Nelen WLDM, Inthout J, Kremer JAM, Verhaak CM. e-Therapy to reduce emotional distress in women undergoing assisted reproductive technology (ART): A feasibility randomized controlled trial. Hum Reprod. 2016;31(5):1046-1057. doi:10.1093/humrep/dew040

B

252. Van Spijker BAJ, Van Straten A, Kerkhof AJFM. Effectiveness of online self-help for suicidal thoughts: Results of a randomised controlled trial. PLoS One. 2014;9(2). doi:10.1371/journal.pone.0090118

253. Van Straten A, Cuijpers P, Smits N. Effectiveness of a web-based self-help intervention for symptoms of depression, anxiety, and stress: Randomized controlled trial. J Med Internet Res. 2008;10(1). doi:10.2196/jmir.954

254. Van Voorhees BW, Ellis JM, Gollan JK, et al. Development and process evaluation of a primary care internet-based intervention to prevent depression in emerging adults. Prim Care Companion J Clin Psychiatry. 2007;9(5):346-355. http://www.ncbi.nlm.nih.gov/pubmed/17998953%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2040278.

255. Vernmark K, Lenndin J, Bjärehed J, et al. Internet administered guided self-help versus individualized e-mail therapy: A randomized trial of two versions of CBT for major depression. Behav Res Ther. 2010;48(5):368-376. doi:10.1016/j.brat.2010.01.005

256. Wagner B, Horn AB, Maercker A. Internet-based versus face-to-face cognitive-behavioral intervention for depression: A randomized controlled non-inferiority trial. J Affect Disord. 2014;152-154(1):113-121. doi:10.1016/j.jad.2013.06.032

257. Wahle F, Kowatsch T, Fleisch E, Rufer M, Weidt S. Mobile Sensing and Support for People With Depression: A Pilot Trial in the Wild. JMIR mHealth uHealth. 2016;4(3):e111. doi:10.2196/mhealth.5960

258. Warmerdam L, Van Straten A, Twisk J, Riper H, Cuijpers P. Internet-based treatment for adults with depressive symptoms: Randomized controlled trial. J Med Internet Res. 2008;10(4):1-12. doi:10.2196/jmir.1094

259. Watts S, Mackenzie A, Thomas C, et al. CBT for depression: A pilot RCT comparing mobile phone vs. computer. BMC Psychiatry. 2013. doi:10.1186/1471-244X-13-49

260. Wellenzohn S, Proyer RT, Ruch W. Humor-based online positive psychology interventions: A randomized placebo-controlled long-term trial. J Posit Psychol. 2016;11(6):584-594. doi:10.1080/17439760.2015.1137624

261. Whiteside U. Online Cognitive Behavioral Therapy for Depressed Primary Care Patients: A Pilot Feasibility Project. Perm J. 2014;18(2):21-27. doi:10.7812/tpp/13-155

262. Whitfield G, Hinshelwood R, Pashely A, Campsie L, Williams C. The Impact of a Novel Computerized CBT CD Rom (Overcoming Depression) Offered to Patients Referred to Clinical Psychology. Behav Cogn Psychother. 2006;34(1):1-11. doi:10.1017/s135246580500250x

263. Williams AD, Andrews G. The Effectiveness of Internet Cognitive Behavioural Therapy (iCBT) for Depression in Primary Care: A Quality Assurance Study. PLoS One. 2013;8(2). doi:10.1371/journal.pone.0057447

264. Williams AD, Blackwell SE, Mackenzie A, Holmes EA, Andrews G. Combining imagination and reason in the treatment of depression: A randomized controlled trial of internet-based cognitive-bias modification and internet-CBT for depression. J Consult Clin Psychol. 2013;81(5):793-799. doi:10.1037/a0033247

265. Williams AD, O'Moore K, Blackwell SE, Smith J, Holmes EA, Andrews G. Positive imagery cognitive bias modification (CBM) and internet-based cognitive behavioral therapy (iCBT): A randomized controlled trial. J Affect Disord. 2015;178:131-141. doi:10.1016/j.jad.2015.02.026

266. Williams AD, Thompson J, Andrews G. The impact of psychological distress tolerance in the treatment ofdepression. Behav Res Ther. 2013;51(8):469-475. doi:10.1016/j.brat.2013.05.005

267. Wright AS, Salmon P, Beck AT, Kuykendall J, Goldsmith LJ, Zickel MB. Development and Initial Testing of a Multimedia Program for Computer-assisted Cognitive Therapy. Am J Psychother. 2002;56(1):76-86.

268. Wright JH, Wright AS, Albano AM, et al. Computer-assisted cognitive therapy for depression: Maintaining efficacy while reducing therapist time. Am J Psychiatry. 2005;162(6):1158-1164. doi:10.1176/appi.ajp.162.6.1158

269. Wu S, Vidyanti I, Liu P, et al. Patient-centered technological assessment and monitoring of depression for low-income patients. J Ambul Care Manage. 2014. doi:10.1097/JAC.0000000000000027

270. Yiend J, Lee JS, Tekes S, et al. Modifying interpretation in a clinically depressed sample using “cognitive bias modification-errors”: A double blind randomised controlled trial. Cognit Ther Res. 2014;38(2):146-159. doi:10.1007/s10608-013-9571-y

B

C

TAXONOMY OF SOFTWARE SYSTEMS FOR DEPRESSION

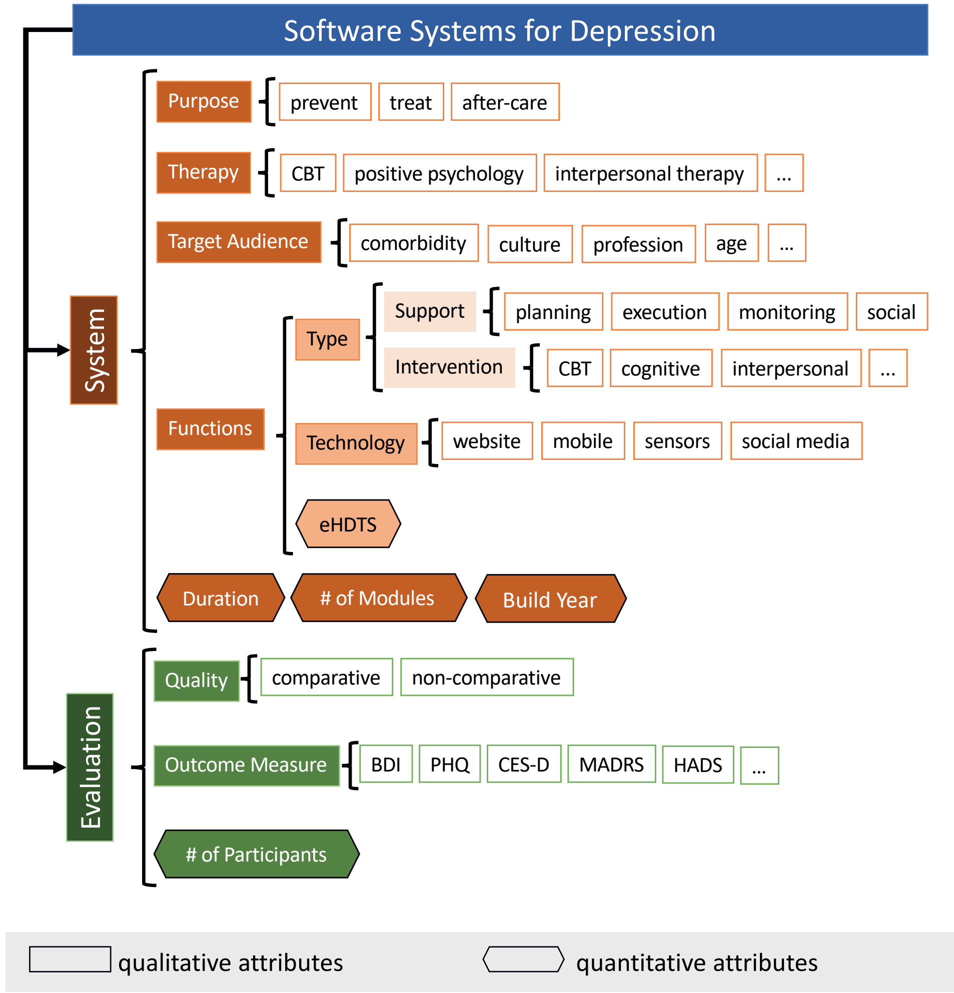

Figure C.1: Taxonomy of software systems for depression. This was inspired by the one for conversational agents, presented in Montenegro JL, da Costa CA, and da Rosa Righi R. "Survey of conversational agents in health." Expert Systems with Applications (2019). It is important to note that this is not an exact graphical representation of all concepts in the database but is intended as an illustration of the most descriptive attributes of software systems for depression. For readers interested in an exact graphical representation of the database, we refer to the SQL schema diagram on the EHealth4MDD (database) website.

D

Categorization of support functions

Table D.1: The categorization of support functions at level 0 (L0), level 1 (L1), and level 2 (L2). Customization as an Execution function type means that the system or intervention could be altered throughout the usage period according to the user's preferences, while as a Planning function type, customization was only possible at the start of the intervention. The difference between Management and Organization within the Execution type is that Organization pertains to management of aspects of the specific system and intervention, whereas Management pertains to dealing with higher-level problems or aspects.

L0	**L1**	**L2**
User preferences (intervention/system)	Customization	Execution
Tunneling, Intervention reasoning, Risk management, Troubleshooting	Management	Execution
Scheduling, Goal setting, Self-screening	Organization	Execution
Rewarding, Encouraging, Empathy, Similarity, Authority, Consideration	Social Role	Execution
Signal, Spark, Facilitator	Trigger	Execution
Automatic/Manual activity monitoring	Activity	Monitoring
Automatic/Manual context monitoring	Context	Monitoring
Automatic/Manual progress monitoring	Progress	Monitoring
Automatic/Manual symptom monitoring	Symptoms	Monitoring
User preferences (intervention/system)	Customization	Planning
Goal setting, Tutorial	Organization	Planning
Peer support, Administrative support, Lay-person support	Direct: NonProfessional	Social
Nurse support, Professional support, Therapist support	Direct: Professional	Social
Social learning/comparison/facilitation, Recognition, Cooperation	Indirect	Social

E

Categorization of intervention functions

Table E.1: The categorization of therapeutic frameworks into therapies (L1). An example function at level 0 (L0) is provided for each therapeutic framework. Some of the therapies were mentioned by authors as having influenced the design of the system, but using our classification, no intervention functions pertaining to the therapy could be found. Therefore, no example can be given. This does not mean that no functionality reflecting the therapy was implemented, for example, a symptom monitoring approach might well result in functionality to aid in the monitoring of symptoms. However, with our classification, this would be classified as a monitoring support function rather than as an intervention function. Similarly, influences from Social Cognitive Theory may have found their way into the system in the form of vignettes, which we classify as indirect social support functions rather than intervention functions.

L0 Example	Therapeutic Framework	L1
Mood monitoring training	Cognitive Behavioral Therapy	Cognitive-Behavioral
Increase self-esteem and acceptance	Acceptance & Commitment Therapy	Cognitive-Behavioral
Learning to be mindful of moods and emotions	Mindfulness(-Based Stress Reduction)	Cognitive-Behavioral
Behavior pattern identification	Behavioral Activation	Behavioral
Breaking a task down into manageable parts	Problem Solving	Behavioral
Graded exposure	Exposure Therapy	Behavioral
-	Social Cognitive Theory	Behavioral
Increase physical activity	Physical Activity	Behavioral
Notice the effects of automatic thoughts	Cognitive Therapy	Cognitive
Concreteness training	Cognitive Bias Modification Training	Cognitive
Cognitive control training	Cognitive Control Training	Cognitive
Learning to recognize threat-monitoring	Meta-Cognitive Therapy	Cognitive
-	Cognitive Remediation	Cognitive
Childhood and early schemata reprocessing	Schema Therapy	Cognitive
Assertiveness and communication skills	Interpersonal Therapy	Interpersonal
Reflecting on past successes	Positive Psychology	Positive Psychology
Breaking unhelpful unconscious patterns	Psychodynamic	Psychodynamic
Building the motivation for change	Motivational Interviewing	Independent
Psychoeducation on depression	Bibliotherapy	Independent
-	Symptom Monitoring	Independent
Expressive Writing	Expressive Writing	Independent
Preventing relapses	Transdiagnostic	Independent
Mind/Body Programming	Hypnosis	Other
Life Review	Life Review	Other
Spiritual and religious resources	Spirituality	Other

F

eHDTS SCALES AND DEGREES

Website of the ehealth4mdd Database

e-Health degree of technological sophistication scales

This table lists the five different eHDTS scales, defines each level, provides an example function for each level and each scale as well as the inter-rater agreement that was obtained when calculating Cohen's kappa between two raters that rated a sample of approximately 25 function descriptions (similar to the examples) using the scale level definitions.

Scale Name	Abbreviation	Level 0	Level 1	Level 2	Level 3	Level 4
Intervention	eHDTS_i	functionality is provided by a human (e.g. face-to-face, by email or telephone)	functionality is not interactive but purely informational (e.g. downloadable pdf, information website)	functionality is provided in an interactive manner (e.g. the user can fill in information on a website) but the information is not processed by the system in any way	functionality is provided in an interactive manner and input from the user is responded to but without processing content of input (e.g. more information is presented after the user has typed a certain the number of characters)	functionality is provided in an interactive manner; content of user input is processed and feedback is provided
Example	*eHDTS_i*	*[Function: Activity planning] [Platform: FACE-TO-FACE] [Description: create an activity plan] [Implementation Details:]*	*[Function: Activity planning][Platform: ONLINE] [Description: Instructions about controlling physical symptoms using strategies such as controlled breathing and scheduling pleasant activities.] [Implementation Details:]*	*[Function: Activity planning] [Platform: ONLINE] [Description: schedule positive activities] [Implementation Details: Each lesson was paired with an interactive tool to provide participants with opportunities to apply the treatment concepts discussed in the lesson. Examples include an activity calendar with which participants could monitor and schedule their activities. These tools were designed to be completed in just a few minutes.]*	*[Function: Activity planning][Platform: APP] [Description: Daybuilder includes a calendar, with one novel facility called "event types". Event types are used for modeling events that require some extra actions. For example, when going to a wedding one has to be nicely dressed and maybe buy a gift. The Daybuilder calendar lets the user create event types and use them to be reminded of the extra actions required for the event.] [Implementation Details:]*	*[Function: Activity planning] [Platform: ONLINE] [Description: a website tutorial guides users through identification of eight to 12 pleasant activities that matter most to their mood (Figure 2). The website then assists with creating a personalized self-contract to make small but consistent increases in the frequency of these activities. Users are prompted to return to the website every few days to record total daily activities and a daily mood rating (on a 7-point Likert scale) for each intervening day.] [Implementation Details: Website algorithms process these tracking data to generate personalized feedback regarding the*

						association (or lack of) between each user's daily mood and his or her pleasant activity level. The website identifies those activities that users report occurring infrequently or not at all or those that are not associated with increased mood, and suggests that these might be replaced with other, more mood-lifting activities.]
Planning Support	eHDTS_p	functionality provided by a human either through the system or entirely externally	functionality provided in a non-interactive/informational manner (e.g. the user is only informed of possible paths through the intervention but the system does not adapt in any manner)	functionality is provided in an interactive manner but no support, this includes also user-driven tailoring if all users are provided with the same options to choose from (e.g. the user can set certain parameters to personalize the system)	functionality is provided with tailored support, tailoring on the basis of interactions with the system	functionality is provided with personalized, context-aware support (e.g. by integrating with Google calendar for scheduling)
Example	*eHDTS_p*	*[Function: Goal setting] [Platform: TELEPHONE] [Description: After an initial assessment and goal-setting telephone call] [Implementation Details:]*	*[Function: Goal setting] [Platform: EMAIL] [Description: Introduction, goals] [Implementation Details: During each online session, participants were sent approximately 20 to 30 PowerPoint slides, including general information on a particular topic, an overview of helpful skills related to the issue being covered, and homework sheets in Farsi (participants' first language).]*	*[Function: Goal setting] [Platform: ONLINE] [Description: SB users interacted with a game-like platform and were invited to describe a goal (an epic win; here, overcoming depression)] [Implementation Details:]*	*[Function: User preferences (intervention)] [Platform: COMPUTER] [Description: Item content covered seven categories: academic, family, mood and health, relationships, social activities, hobbies and work. In order to maximise the personal relevance of the intervention all participants were asked to indicate at screening which one of these categories was least relevant to them. For those subsequently assigned to the active condition, randomisation included automatic allocation to a personalised version of the intervention programme, that omitted all items in a participant's least*	*[Function: User preference (system)] [Platform: INTERACTIVE VOICE RESPONSE] [Description: Outbound automated calls are scheduled based on a dynamic protocol of type of call (screening or monitoring), calendar dates, clinical events, call history, and patient preferences (such as call time), using data and the DCAT algorithms in the DMR. Patients can opt for password-protected access to enhance*

F

					relevant category.] [Implementation Details:]	privacy and can use the system to reschedule the call or request human follow-up.] [Implementation Details:]
Execution Support	eHDTS_e	functionality provided by a human either through the system or externally	functionality provided in an informational manner or functionality is provided in a rule-based fashion (if-then), condition is user-independent (e.g. if Monday, release module)	functionality provided in a rule-based fashion (if-then), consequence is dependent on user interaction data (e.g. if user has not interacted with system for 7 days, send reminder)	functionality provided in a rule-based fashion (if-then), consequence is dependent on a static user model (e.g. if user is severely depressed, release new modules twice a week instead of once for mildly and moderately depressed users)	functionality is provided in a rule-based fashion (if-then), but consequence is dependent on a dynamically changing user model (e.g. system checks whether user has acquired new skills and if so those modules that require the application of these skills become accessible)
Example	*eHDTS_e*	*[Function: Tunneling] [Platform: ONLINE] [Description: Participants were given gradual access to the self-help modules. The therapists gave feedback on the clients' experiences and administered the gradual access to the modules.] [Implementation Details:]*	*[Function: Tunneling] [Platform: ONLINE] [Description: The participant logged into the program and was instructed to finish the lessons in a predetermined order.] [Implementation Details:]*	*[Function: Tunneling] [Platform: ONLINE] [Description: in the latest version of the MoodGYM program (Mark III) core assessments are compulsory not allowing the user to skip or alternate between the different modules.] [Implementation Details:]*	*[Function: Tunneling][Platform: ONLINE] [Description: The 20-item CES-D was the first level assessment in HIV TIDES. CES-D scores were used for the intervention pathways because the instrument is multi-factorial thus facilitating tailoring of specific self-care interventions. The cut-off points of 8 and 16 (total range from 0 to 60) were used to divide system users into three major groups which determined the tailoring path the users would experience.] [Implementation Details:]*	*[Function: Tunneling] [Platform: COMPUTER] [Description: Session 5 included a comprehensive assessment to evaluate whether participants had acquired sufficient knowledge to progress to the next five sessions or if remediation was warranted. The program utilized an algorithm that branched to earlier sessions when remedial learning objectives needed to be met based on scores obtained in Session 5.] [Implementation Details: Mini-lessons were prioritized based on the comprehensive assessment of participant needs during Session 5. Those content areas showing greatest need based on item score sampling*

F

						across all mini-lesson options (16 in total) were given highest priority and presented earlier and those with lower scores presented later.]
Monitoring Support	eHDTS_m	data collected but not used in intervention (e.g. data is collected for study purposes)	data collected and forwarded to human third party for use in intervention (e.g. a therapist receives data and makes choices accordingly)	data collected and presented to the user for self-monitoring (data is not interpreted by the system)	data collected and used to provide feedback to users on this specific data	data collected, analyzed, and used by system in the intervention (e.g. for personalization, intervention improvement, making prognoses)
Example	*eHDTS_m*	*[Function: Manual monitoring of symptoms] [Platform: SMS] [Description: The following mood monitoring message was sent once daily at random times: What is your mood right now on a scale of 1 to 10?] [Implementation Details:]*	*[Function: Manual monitoring of symptoms][Platform: ONLINE] [Description: Prior to each lesson, the patient completed the Kessler Psychological Distress Scale(K10) (Kessler et al.,2003) as a means of tracking patient progress during the iCBT course.] [Implementation Details: Clinicians were automatically notified if their patient's K10 score increased by half a standard deviation or more.]*	*[Function: Manual monitoring of symptoms] [Platform: ONLINE] [Description: Activity report for self-monitoring, which provides feedback to users, showing that their mood is related to the activities performed, and the benefits of being active.] [Implementation Details:]*	*[Function: Manual monitoring of symptoms][Platform: ONLINE] [Description: rating mood by selecting the extent to which each of 20 interactive mood-adjective playing cards describes current mood. The rating is based on the extent that users currently experience the 20 (positive and negative) emotions (e.g. proud, nervous, determined). Cards appear one at a time and the user selects either the flip or rotate button until their selection (very slightly or not at all, a little, quite a bit, extremely) is highlighted. They then click their selection and the next card appears. Test completion takes approximately 5 min and generates a daily score to enable mood tracking over time. Historical scores can be retrieved as a graph to which the user can add annotations. Users receive two forms of feedback on mood: automated feedback from the website, which comprises a short, two-paragraph supportive summation based on comparisons between the most*	*[Function: Manual monitoring of symptoms] [Platform: COMPUTER] [Description: The aim of this application is to detect symptoms of anxiety and depression in users. It includes a decision algorithm that allows Butler to react in real time to the user's clinical needs. To do this, an evaluation using widely used and validated psychological scales is performed when the user accesses the system. If the clinical status of the person is within the normal range, Butler offers its full potential. In the case that any change in mood (anxiety or depression) is detected, the system performs a more detailed exploration; depending on the result, it offers the most appropriate options for the user's*

F

					recent and previous scores (e.g. things aren't as good as they looked the last time you took the test and got 35%).] [Implementation Details:]	*emotional state and sends the appropriate warning to the professional user platform. This application also summarizes the information so the professional can process it efficiently through period analyses, data tables, bar graphics, and more.] [Implementation Details:]*
Social Support	eHDTS_s	human support is part of intervention but not integrated into system (e.g. system is an adjunct to face-to-face therapy) or the support is initiated by the support-provider at pre-specified times (e.g. therapists always sends feedback email to user after homework submission)	system provides support in a non-interactive/informational manner (e.g. it provides information on how to get in touch with human support or it includes videos of role models or summary statistics of how other users of the same system are doing)	system has interactive, dynamic, integrated social support (e.g. integrated chat functionality, integrated video conferencing, a peer support forum)	system is proactive about contacting human support on the basis of user activity in the system (e.g. it automatically contact support in case of risk detection, or it suggest to the user to cooperate with others on the basis of detected low adherence)	system intelligently supports the provision of human support by going beyond the specific user (e.g. by integrating with other social platforms and connecting a user with other people)
Example	*eHDTS_s*	*[Function: Nurse support] [Platform: TELEPHONE] [Description: Both groups continued nurse-supervised treatment for 12 weeks, which involved regular phone calls checking for problems] [Implementation Details:]*	*[Function: Social learning][Platform: ONLINE][Description: The Course comprises material both in didactic form, that is, text based instructions and information, and case-enhanced learning examples. Case-enhanced learning is informed by principles of Social Cognitive Theory (SCT) [33] and uses educational stories that identify a problem and a solution which an example (i.e., a case) resolves for the learner.] [Implementation Details:]*	*[Function: Social facilitation] [Platform: ONLINE] [Description: As a first step, the user can see anonymous indications of other people in the system. The intention is to reassure users that they are not alone in experiencing difficulties and that many other people have experienced similar problems and overcome them.] [Implementation Details: Users can respond to content by indicating that they "like" it, and can see how many other people liked it, helping to reduce the*	*[Function: Lay person support] [Platform: INTERACTIVE VOICE RESPONSE/EMAIL] [Description: based on subject's weekly IVR assessments, the CarePartner received structured emails or IVR calls with feedback about the subject's status with suggestions for how the CarePartner could support depressive symptom self-management. CarePartners were notified immediately via an IVR call if the subject reported (1) suicidal thoughts or a suicidal plan, (2) rarely or never taking the depression medication as prescribed, and (3) side effects that are making the subject consider taking less medication. Structured fax alerts*	*[Function: Cooperation] [Platform: SOCIAL MEDIA] [Description: Respondents in our study were a mixture of other Panoply users and paid workers from MTurk. Responses on Panoply fall into three categories: support, debug, and reframe. Support responses offer emotional support and active listening. Debug responses help users identify and dispute cognitive distortions (bugs). Reframe responses offer alternative, more positive ways of thinking*

				sense of isolation.]	*were sent automatically to the subject's PCP if the subject reported (1) medication non-adherence or(2) PHQ-9 >15 twice in the last month] [Implementation Details:]*	*about the stressful situation. Respondents are not asked to use any one particular reappraisal strategy but instead are given a bulleted list of tactics to consider in case they need inspiration.] [Implementation Details: All of the aforementioned interactions are coordinated entirely through Panoply's automation. The user needs to only submit their post to start this sequence of crowd work.]*

F

G

Ranking of systems according to evidence base

The following two tables show the ranking of the 133 systems contained in the database by evidence base, as far as this has been recorded in the database. The first table shows the evidence base of systems evaluated in comparative trials (such that are randomized, controlled, or both), whereas the second table shows the evidence base of systems evaluated in noncomparative trials (single-group trials). Both tables are sorted first on the number of evaluations, on the number of participants recruited to take part in the study (sum over all study arms), and the number of participants who completed the study (not including follow-up). For readers interested in more information, the system key, as denoting systems in the database, is provided. This should allow for easy querying of associated versions, authors, and articles, to name only a few things.

Table 1. Systems evaluated in comparative trials ranked by evidence base.

System Key	*Name of System*	*Number of Versions*	*Evaluation Quality*	*Number of Evaluations*	*Number of Participants Pre*	*Number of Participants Post*
13	MoodGYM	15	comparative	11	7294	2636
16	Sadness Program	13	comparative	9	2738	1495
32	Alles onder controle	8	comparative	9	1960	1498
10	The Wellbeing Course	18	comparative	9	1879	1771
15	Beating the Blues	5	comparative	9	1422	1036
24	Andersson Unnamed	9	comparative	7	1422	1214
2	Deprexis	3	comparative	6	1863	1377
20	Colour your life	4	comparative	6	1574	1189
36	CBM Blackwell	6	comparative	5	376	347
14	e-couch	4	comparative	3	2703	1256
40	Get.On Mood Enhancer	3	comparative	3	1072	906
51	Living to the full	2	comparative	3	851	676
11	Overcoming Depression on the Internet	4	comparative	3	714	161
3	SHADE	2	comparative	3	475	56
54	Good Life Compass	2	comparative	3	145	139
34	Mood Memos	1	comparative	2	3062	1143
1	Living Life to the Full	2	comparative	2	659	575
115	Staerkentraining	5	comparative	2	623	462
18	Smiling is Fun	5	comparative	2	356	294
82	Cukrowicz Unnamed	2	comparative	2	242	119
129	CBM Moebius	1	comparative	2	191	191
78	Ly Unnamed	2	comparative	2	174	163
37	CBM-errors	2	comparative	2	120	114

109	Joutsenniemi Unnamed	1	comparative	1	3274	1288
98	UTSMed	1	comparative	1	1236	1020
133	Wellenzohn Unnamed	5	comparative	1	1162	994
53	SHUTi	1	comparative	1	1149	581
132	Shapira Unnamed	2	comparative	1	1002	188
46	Patten Unnamed	1	comparative	1	786	NA
65	Internet CBT program; Useful mental health solutions series for business.	1	comparative	1	762	606
104	Bunge Unnamed	4	comparative	1	728	464
61	myCompass	1	comparative	1	720	519
92	Seligman PPI	1	comparative	1	577	411
116	MyPAA	1	comparative	1	514	97
117	Walk 2.0	1	comparative	1	514	97
108	MH-Guru	1	comparative	1	507	386
93	Sergeant PPI	1	comparative	1	466	166
76	Geisner Unnamed	1	comparative	1	349	311
7	INCPAD	1	comparative	1	309	246
30	MoodHacker	1	comparative	1	300	286
74	Psyfit	1	comparative	1	284	214
39	SuperBetter	2	comparative	1	283	74
63	Depression Free	1	comparative	1	239	193
119	van Spijker Unnamed	1	comparative	1	236	215
47	SUMMIT	2	comparative	1	236	205
23	Seligman Unnamed	1	comparative	1	227	212
87	Panoply	1	comparative	1	217	166
31	Mindful Mood Balance	1	comparative	1	200	153
83	Tame Your Gut	1	comparative	1	199	143

G

85	Wellness Workshop	1	comparative	1	191	190
27	Space from Depression	1	comparative	1	188	152
90	Spirituality teaching program	1	comparative	1	165	147
118	Shamekhi Unnamed	1	comparative	1	160	154
72	Recovery Road	3	comparative	1	140	93
21	Creating Opportunities for Personal Empowerment	2	comparative	1	121	93
33	Digicoach	1	comparative	1	109	80
124	Learning Mindfulness Online	1	comparative	1	104	58
42	eCare for Moods	1	comparative	1	103	100
45	moodManager	2	comparative	1	102	89
56	APT Johansson	1	comparative	1	100	100
114	PopTherapy	4	comparative	1	95	33
75	Alavi Unnamed	1	comparative	1	93	84
99	SUBGAP	1	comparative	1	92	88
19	iCBT-MDD	1	comparative	1	91	72
126	Killen Unnamed	1	comparative	1	88	87
28	Hollandare Unnamed	1	comparative	1	84	77
101	Ly Unnamed Mindfulness	1	comparative	1	81	72
26	Depressionshjalpen	1	comparative	1	80	78
50	Hoorelbeke Unnamed	1	comparative	1	68	67
128	Emotion Diary	1	comparative	1	68	44
69	Feeling Better	3	comparative	1	66	53
5	Avanti	1	comparative	1	63	52
17	Wagner Unnamed	1	comparative	1	62	53
103	Bond Unnamed	1	comparative	1	62	NA
131	OLeary Unnamed	2	comparative	1	61	35

123	CBM Browning	2	comparative	1	61	NA
127	Dobbin Unnamed	1	comparative	1	58	49
86	Calkins Unnamed	1	comparative	1	56	48
38	CBM Peters	1	comparative	1	54	52
57	Agyapong Unnamed	1	comparative	1	54	50
25	Ruwaard Unnamed	1	comparative	1	54	49
100	UPLIFT	1	comparative	1	53	NA
121	CBM Beevers	1	comparative	1	52	44
68	Buhrman Unnamed	1	comparative	1	52	43
9	DAHLIA	1	comparative	1	49	42
48	Strom Unnamed	1	comparative	1	48	48
125	Eisma Unnamed	2	comparative	1	47	36
43	improvehealth.eu	1	comparative	1	46	22
88	Cognitive Therapy: A Multimedia Learning Program	2	comparative	1	45	40
130	CBM Mogoase	1	comparative	1	42	41
120	Kraft Unnamed	1	comparative	1	41	35
89	Interactive Computer-Assisted Psycho-Education System	1	comparative	1	32	32
91	Cognitive Remediation	1	comparative	1	28	21
95	Help4Mood	2	comparative	1	28	21
112	OLDIT	2	comparative	1	24	NA
102	EVO	1	comparative	1	22	22
80	ePST	2	comparative	1	14	13
35	STREAM	1	comparative	1	NA	NA

G

Table 2. Systems evaluated in non-comparative trials ranked by evidence base.

System Key	*Name of System*	*Number of Versions*	*Evaluation Quality*	*Number of Evaluations*	*Number of Participants Pre*	*Number of Participants Post*
10	The Wellbeing Course	18	non-comp.	9	1131	979
15	Beating the Blues	5	non-comp.	7	584	361
16	Sadness Program	13	non-comp.	5	2198	1090
13	MoodGYM	15	non-comp.	4	80837	22496
66	Janevic Unnamed	4	non-comp.	4	653	375
24	Andersson Unnamed	9	non-comp.	3	72	57
64	ADep	2	non-comp.	2	16584	646
27	Space from Depression	1	non-comp.	2	721	334
18	Smiling is Fun	5	non-comp.	2	101	85
113	Daybuilder	2	non-comp.	2	38	4
94	DCAT ATA	1	non-comp.	1	442	NA
73	Leykin Unnamed	1	non-comp.	1	309	109
59	Pfeiffer Unnamed	1	non-comp.	1	190	66
110	Kawai Unnamed	1	non-comp.	1	168	126
60	MOSS App	1	non-comp.	1	126	12
49	Stress Gym	1	non-comp.	1	110	95
137	Hetrick Unnamed	1	non-comp.	1	101	55
84	Blues Begone	1	non-comp.	1	100	58
88	Cognitive Therapy: A Multimedia Learning Program	2	non-comp.	1	96	75
61	myCompass	1	non-comp.	1	90	49

System Key	Name of System	Number of Versions	Evaluation Quality	Number of Evaluations	Number of Participants Pre	Number of Participants Post
77	Hadjistavropoulos Unnamed	1	non-comp.	1	80	41
67	Ahmedani Unnamed	1	non-comp.	1	75	64
52	LINKS	1	non-comp.	1	52	24
99	SUBGAP	1	non-comp.	1	44	33
32	Alles onder controle	8	non-comp.	1	44	23
12	Rebound	1	non-comp.	1	42	39
21	Creating Opportunities for Personal Empowerment	2	non-comp.	1	39	31
136	ASCENSO	1	non-comp.	1	35	23
8	HIV Tides	1	non-comp.	1	32	NA
80	ePST	2	non-comp.	1	29	23
55	Reframe IT	1	non-comp.	1	27	21
105	MindBalance	1	non-comp.	1	25	18
71	MyStrength	1	non-comp.	1	24	20
134	Overcoming Depression	1	non-comp.	1	22	15
45	moodManager	2	non-comp.	1	21	19
107	Moodscope	1	non-comp.	1	20	16
118	Shamekhi Unnamed	1	non-comp.	1	20	13
97	Falconer Unnamed	1	non-comp.	1	18	15
22	Mansson Unnamed	1	non-comp.	1	15	15
79	Building a Meaningful Life through Behavioral Activation	1	non-comp.	1	15	8
29	van Voorhees Unnamed	1	non-comp.	1	14	8
58	Aguilera Unnamed	1	non-comp.	1	12	10
41	Bae Unnamed	1	non-comp.	1	10	10

System Key	Name of System	Number of Versions	Evaluation Quality	Number of Evaluations	Number of Participants Pre	Number of Participants Post
69	Feeling Better	3	non-comp.	1	10	10
122	Both Unnamed	1	non-comp.	1	9	9
44	Schueller Unnamed	1	non-comp.	1	9	9
96	Health Buddy	1	non-comp.	1	9	9
111	Minddistrict	1	non-comp.	1	9	7
106	T2 Mood Tracker	1	non-comp.	1	8	8
36	CBM Blackwell	6	non-comp.	1	8	7
62	Mobilyze	1	non-comp.	1	8	7
135	Butler system	1	non-comp.	1	4	4
95	Help4Mood	2	non-comp.	1	2	2
70	Mindfulness Online	1	non-comp.	1	NA	273

H

FICTIONAL FUNCTIONS FOR EACH eHDTS LEVEL

Table H.1: The table provides a description of some possible functions that a system at each of the e-mental Health Degree of Technological Sophistication (eHDTS) score levels might have. These descriptions are fictional, and eHDTS scores at the system level are averages.

eHDTS score	General system description
0	System has no functions delivered by means of technology. Technology might, however, serve as a medium between counsellor and patient, as is the case in telehealth. Additionally, technology might be used to gather data about the user, not for use in the intervention but for the purposes of conducting the study, e.g. e-mails being sent to collect pre- and post-measurements.
1	System delivers functions in an informational fashion. For example, a typical CBT intervention might include eight modules of psychoeducation, exercise instructions, and links to additional resources. This is presented on a website and each week of the intervention a new module is made available. The user can click through the different modules but is not prompted to interact with the intervention in any other way. Applying the insights from the educational texts in real life is left to the user. Human guidance is integrated in the same informational manner, e.g. by showing (fictional) video vignettes of other users or similar patients or using prototypical characters. Data might be collected concerning the progress of the user, such as weekly questionnaires or which module they are on and this is then forwarded to a counsellor who might get in touch with the user.
2	System functions are interactive but neither responsive nor intelligent. In the case of the core modules, this might mean that the user is prompted to enter information into the system, such as to describe their automatic thoughts or to create an activity schedule. The system, however, does not process this information to any extent beyond simply saving it for the user to see back. When the system collects monitoring data, it presents this back to the user without offering an interpretation, e.g. it might automatically create a graph of daily depression questionnaire scores. The user might be able to tailor the systems to their own needs or preferences by adapting settings, such as what gets displayed on their landing page. New modules might be triggered according to how frequently the user has been interacting with the program. Human guidance may be integrated into the system in an interactive and responsive way, though the responsiveness stems from other humans and not from the system, as is the case in a forum, for example.
3	System functions are interactive and responsive to meta-data and might show some intelligence. For example, the core modules might deliver information in accordance with the amount of user activity on the interactive exercises. Some tailoring may be provided in the execution support functionality, but this is based on a static user model created at the outset of the intervention. For example, users might be determined to be male or female initially and the intervention is adapted accordingly. Monitoring data that is collected is processed to draw conclusions for the user and provide feedback. For example, users might see a qualitative interpretation of their depression scale score. The system actively attempts to integrate human support when it detects that the user may need it (e.g. in the case of suicidal ideation).

H

I

Cumulative density deciles of eHDTS scores

Table I.1: Deciles and their corresponding scale values for the weighted and unweighted scale, e.g. 50 % of systems have an average technological sophistication of 1.5 or lower. The weighted column takes into account the number of functions that a system implements.

Deciles	eHDTS	eHDTS$_w$
10	1.00	0.10
20	1.09	0.22
30	1.25	0.32
40	1.33	0.39
50	1.50	0.45
60	1.67	0.55
70	1.89	0.69
80	2.06	0.82
90	2.53	1.11
100	3.70	2.31

J
Heatmap of the maximum eHDTS degree

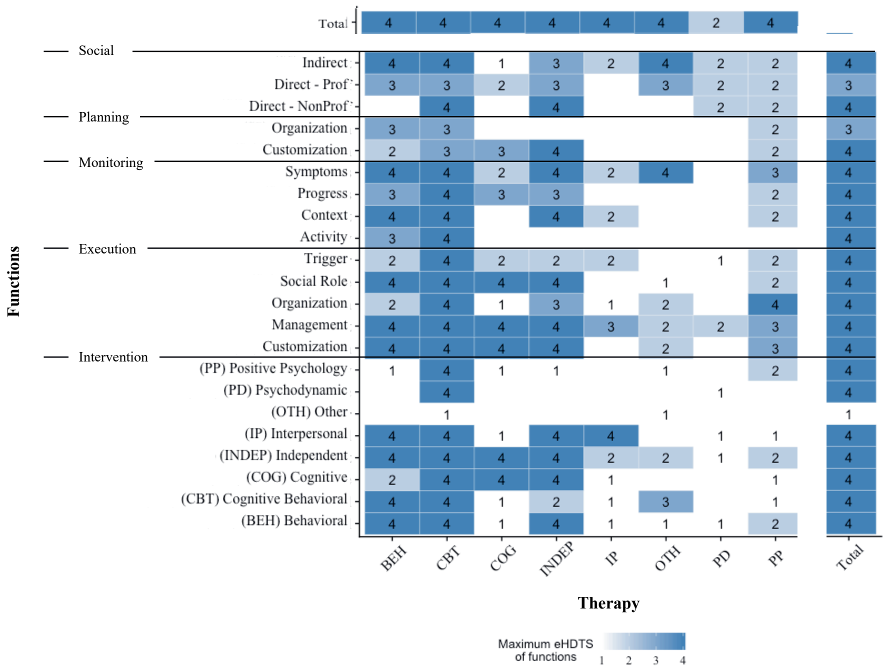

Figure J.1: Heatmap of the maximum degree of technological sophistication per function type and therapy. This gives insight into the technological state of the art of each function type and therapy.

J

K

Experimental flow of the thought recording study

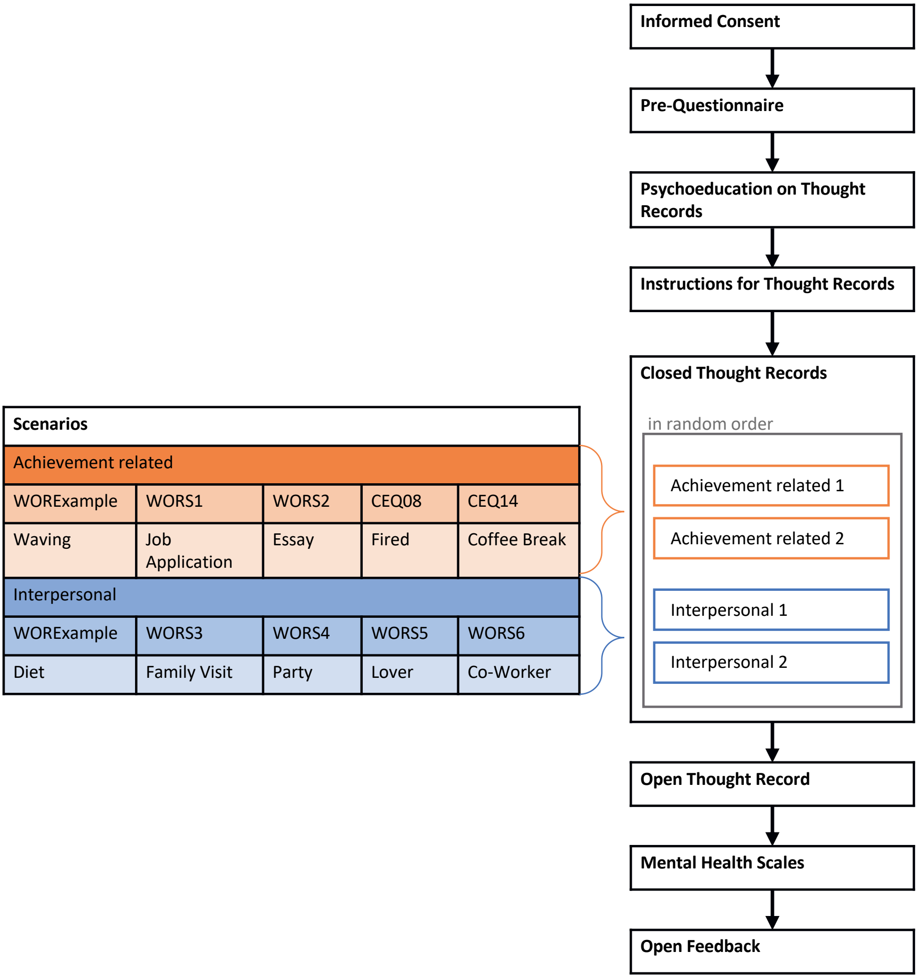

Figure K.1: Figure displays the different stages of the experiment as traversed by the participants.

L

Computation of predictor and outcome variables for H3

Open Thought Record, Participant 13224

Utterance	Attachment	Competence	Global self-evaluation	Health	Power and control	Meta-cognition	Other people	Hopelessness	Other people's views on self
I am miserable because I have such a hard time saying no.	0	0	0	0	3	2	0	0	0
I will continue to get stuck in situations I don't feel good about.	0	0	0	0	3	0	0	1	0
I won't be very happy with many outcomes in my life.	0	0	0	0	0	0	0	2	0
I will feel helpless in controlling how my life turns out.	0	0	0	0	3	0	0	1	0
Maximum value per schema	0	0	0	0	3	2	0	2	0

9 outcome values

Open thought record is *achievement-related*

Identify the two closed thought records of same participant that are also *achievement-related*

Closed Thought Record 1, Participant 13224

Utterance	Attachment	Competence	Global self-evaluation	Health	Power and control	Meta-cognition	Other people	Hopelessness	Other people's views on self
I must not be doing this diet thing right.	0	1	0	1	0	0	0	0	0
I won't be successful at losing weight.	0	2	0	3	0	0	0	0	0
I will feel bad about my appearance.	3	0	0	0	0	0	0	0	0
It will affect my attitude negatively.	0	0	0	0	0	1	0	0	0
Maximum value per schema	3	2	0	3	0	1	0	0	0

Closed Thought Record 2, Participant 13224

Utterance	Attachment	Competence	Global self-evaluation	Health	Power and control	Meta-cognition	Other people	Hopelessness	Other people's views on self
I must not be good at persuading people.	0	1	0	0	1	0	2	0	0
I will get pushed around in life.	0	0	0	0	3	0	0	1	0
I won't get the relationships I want with others.	3	0	0	0	1	0	0	0	0
I'll be constantly disappointed with life.	0	0	0	0	0	0	0	2	0
Maximum value per schema	3	1	0	0	3	0	2	2	0

Average the two sets of maximum values together per schema

Average value per schema	3	1.5	0	1.5	1.5	0.5	1	1	0

9 predictor values

Figure L.1: Graphical illustration of how we determined the predictor and outcome variables for the nine models of hypothesis 3.

L

M

Results of the five linear models to test H4

Table M.1: Table that summarizes the main outcomes of the five linear models that were fit to assess whether there is a link between schemas and outcomes on various mental health questionnaires.

Schema	**HDAS-D**				**HDAS-A**				**BDI-IA**				**CD-R**				**CD-A**			
	b	β	t	p	b	β	t	p	b	β	t	p	b	β	t	p	b	β	t	p
Attachment	0.06	0.01	0.24	0.81	-0.13	-0.03	-0.51	0.61	0.29	0.03	0.49	0.63	-0.13	-0.01	-0.22	0.83	-0.39	-0.03	-0.58	0.56
Competence	-0.24	-0.05	-0.90	0.37	-0.24	-0.05	-0.85	0.40	-0.85	-0.07	-1.25	0.21	0.11	0.01	0.16	0.87	0.82	0.06	1.07	0.28
Global self-evaluation	0.31	0.09	1.56	0.12	0.63	0.18	2.98	**<0.01**	1.25	0.15	2.49	0.013	2.11	0.24	4.08	**<0.001**	2.06	0.21	3.64	**<0.001**
Health	0.69	0.10	1.59	0.11	0.57	0.07	1.26	0.21	1.90	0.10	1.75	0.08	1.00	0.05	0.89	0.37	1.52	0.07	1.25	0.21
Power and Control	0.92	0.11	1.86	0.06	0.94	0.10	1.80	0.07	1.39	0.06	1.12	0.27	3.44	0.15	2.69	**<0.01**	2.31	0.09	1.66	0.10
Meta-cognition	0.90	0.04	0.77	0.44	1.36	0.06	1.12	0.27	1.80	0.03	0.62	0.54	-2.22	-0.04	-0.74	0.46	-0.65	-0.01	-0.20	0.84
Other people	-0.24	-0.01	-0.21	0.83	-0.80	-0.04	-0.67	0.50	0.56	0.01	0.19	0.85	-2.24	-0.04	-0.76	0.45	-2.96	-0.05	-0.92	0.36
Hopeless	-0.06	-0.01	-0.11	0.92	0.15	0.01	0.24	0.81	0.68	0.03	0.44	0.66	1.37	0.05	0.87	0.39	1.88	0.06	1.09	0.28
Other's views on self	-0.25	-0.05	-0.84	0.40	0.05	0.01	0.16	0.87	-1.36	-0.10	-1.82	0.07	0.54	0.04	0.71	0.48	-0.01	-0.00	-0.01	0.99

Abbreviations: HDAS-D – Hospital Depression and Anxiety Scale - Depression, HDAS-A – Hospital Depression and Anxiety Scale - Anxiety, BDI-IA – Beck Depression Inventory - I amended, CD-R – Cognitive Distortions - Relatedness, CD-A – Cognitive Distortions - Achievement, b – estimates for the regression coefficient, β – standardized estimates for the regression coefficient

M

ACKNOWLEDGEMENTS

I've heard many a PhD candidate refer to the project as their baby. Well, they say it takes a village to raise a child and no less is true for a PhD. Please meet all the helpful, if not even instrumental, people who have been temporary or permanent residents of my PhD village.

The two people without whom the research would never have been possible are my promotors, Willem-Paul Brinkman and Mark Neerincx. Willem-Paul has been a true mentor. He made a point of seeing me (and all his PhD students) weekly and for a full hour. While I started out in 2016 with a hierarchical supervision idea, this completely dissolved over the years because he made made the progress meetings not about my progress but about exchanging ideas and collaborating together on a research project that we both cared about. Willem-Paul, I very sincerely want to thank you for your availability, approachability, down-to-earth coaching style, your efforts to always think along and think ahead, and, of course, all the wonderful anecdotes and worldly wisdom. Mark, although our meetings were sparse, you always made your presence count by really tuning in. Many times I have found your ability to filter the important from the unimportant and to contribute in meaningful ways to projects that you only receive occasional updates about to be astounding. Thank you for your unwavering trust in my ability to finish the PhD, even when the going got tough, for your sincere interest in the well-being of your PhD students, and for all the helpful guidance along the way.

I would also like to extend my gratitude to the independent members of my doctoral committee: Tony Belpaeme, Geert-Jan Houben, Catholijn Jonker, Emiel Krahmer, and Marcus Specht. I appreciate the time and effort that you have spent on reading this thesis and I look forward to the discussions we will have during the defense ceremony.

My first encounter with the Interactive Intelligence group was when I needed a desk to work at as a Master student under the supervision of Joost Broekens. My boyfriend and I were living in a 30 sqm apartment with sloping walls on all sides; the only spot where one could sit upright at a desk was occupied for 50 % of the time by said boyfriend. Needless to say that I'm grateful to Frank, Rifca, Pietro, and Thomas for not only sharing their desks with me for a few months but giving me such a warm welcome that I didn't want to leave anymore; and here we are six years later... Over the years, many of my II colleagues became friends, enriching the PhD experience with lavish coffee breaks, basement ping-pong tournaments, much moral support, the occasional dinner or game night, and as of recently even play dates. Thank you, Ding, Rifca, Ursula, Myrthe, Bernd, Thomas, Pietro, Frank, Vincent, Ilir, Mike, Elie, Miguel, Merijn, Elena, Rolf, Roel, Alex, Aleks, Rijk, and Catha for being my willing victims when it came to social activities most of the time. Research is not always fun (especially not when you are doing manual data extraction from papers for months or scoring somewhat depressing thought records), but you all ensured that before the pandemic, I loved coming to the office every day regardless. Also thanks to Catholijn for being a fantastic head of the group, and for Anita, Bart, Ruud, and Wouter for always having our backs!

My PhD project was funded by the 4TU Center for Humans & Technology. This framework allowed me to meet another set of amazing people that have influenced my PhD research in various ways. First of all, I want to thank Mark, Wijnand, and Dirk for being the masterminds behind the center and for bringing us all together. I have many fond memories of the write-athons, the winter school, various symposia and workshops together with the other PhD students and postdocs in the center. Minha, Jan, Michel, Sima, Merijn, Gijs, and Bernd, thank you all for being awesome and always so happy to help. We talked much initially about how forced the collaboration seemed, but looking back now, we really bonded, had joined publications, and at least some of us will probably continue to collaborate well into the future.

I served for one year on the PhD council, where I met a number of awesome EEMCS PhD candidates. Thank you Yuki, Rishabh, Jacopo, Chirag, Sobhan, Christine, and Nakisa for a great collaboration and some good times. A very warm thanks is also due for my friends outside of the academic context. I here want to especially mention all the members of the Fun-/Spreekclub for letting me practice my Dutch in an entirely non-judgemental setting and for lending an open ear at all times. I would always look forward to our Monday meetings the entire week, so thank you Gosia, Reza, Renée, Sascha, Teresa, Carina, Jair, and the various more fleeting members. All the friends that I have made over the many years prior to becoming a PhD student and that have stuck around (you know who you are) also deserve a big thanks for just being great people and great friends. And one person who definitely kept me going when I got pretty close to despairing was Nina, our babysitter and friend. Nina, you managed to give me some breathing space en daarvoor nen hele dikke merci.

As is true for anyone, I would never be what I am today if it wasn't for my parents. However, I have been so fortunate as to have parents that have always supported me in the choices that I made or threatened to make. This not only saved me from ending up completely broke in Australia and various other (almost) disasters during my life but also enabled me to pursue many years of higher education and ultimately this PhD degree. Thank you, Mama and Papa, for your unconditional love and support. In the same breath, I also want to thank my brothers, Benedikt and Lukas for being the most faithful visitors from abroad over the years. I know I can always count on you two when I need you.

And finally, to the queen and king of my PhD village. Amalia, I marvel at you every day. Thank you for always reminding me of the true purpose of life and for finally teaching me to say "no." And Volker, you really are my better half. You challenge me to become a better human being all the time but you are also the shoulder that I can lean on when life feels overwhelming. Thank you for always being there for me, for staying by my side through thick and thin, for everything.

I embarked on this PhD journey with the aim to build a behavior change support system that would help PhD candidates with symptoms of depression. Over the years, I got to know some of the ups and downs of the process first and second hand. However, I think that, just like when raising a child, the support network that one has can make all the difference. I consider myself truly blessed in this regard.

CURRICULUM VITÆ

Franziska Viktoria BURGER

14-11-1991	Born in Waiblingen, Germany.

EDUCATION

1996–2010	Lindenschule Hohenacker, Waiblingen (1996–2001) Staufer Gymnasium, Waiblingen (2001–2002) Tuscaloosa Academy, Tuscaloosa (AL) (2002–2006) Königin-Olga-Stift-Gymnasium, Stuttgart (2006–2010)
2010–2013	Bachelor of Science in Cognitive Science Eberhard Karls Universität Tübingen
2014–2015	Master of Science in Artificial Intelligence Radboud Universiteit Nijmegen

WORK EXPERIENCE

2011–2012	Research Assistant Knowledge Media Research Center, Tübingen
2012–2013	Research Assistant Max Planck Institute for Biological Cybernetics, Tübingen
2013–2014	Intern Mercedes Benz Research and Development North America, Sunnyvale (CA)
2016–2021	PhD Candidate Delft University of Technology, Delft *Thesis:* Supporting Electronic Mental Health with Artificial Intelligence — Thought Record Analysis and Guidance *Promotors:* Prof. dr. M.A. Neerincx and Dr. ir. W.-P. Brinkman
2022–	Post-doctoral Researcher Delft University of Technology, Delft
2022–	Post-doctoral Researcher Eindhoven University of Technology, Eindhoven

List of Relevant Publications

Peer-Reviewed Articles

4. Burger, F., Neerincx, M. A. & Brinkman, W. P. Ehealth4MDD: A database of e-health systems for the prevention and treatment of depressive disorders. *Annual Review of CyberTherapy and Telemedicine* **16,** 18–24 (2018)
3. Burger, F., Neerincx, M. A. & Brinkman, W.-P. Technological State of the Art of Electronic Mental Health Interventions for Major Depressive Disorder: Systematic Literature Review. *Journal of Medical Internet Research* **22,** e12599 (2020)
2. Burger, F., Neerincx, M. A. & Brinkman, W. P. Natural language processing for cognitive therapy: Extracting schemas from thought records. *PLoS ONE* **16** (2021)
1. Burger, F., Neerincx, M. A. & Brinkman, W.-P. Using a conversational agent for thought recording as a cognitive therapy task: Feasibility, content, and feedback. *Frontiers in Digital Health* **4** (2022)

Data Sets

4. Burger, F., Neerincx, M. A. & Brinkman, W.-P. *Reliability and validity analyses for the coding of information entered into the Ehealth4MDD database* 4TU.ResearchData `https://doi.org/10.4121/uuid:7e7e91ab-7afc-4b48-8915-e2bc80b23c99`. (2018)
3. Burger, F. *Analysis of technological state of the art in e-mental health for the prevention and treatment of depression.* 4TU.ResearchData `https://doi.org/10.4121/uuid:623b37bc-3ea7-48e3-ab09-575aa3b1b9ac`. (2019)
2. Burger, F. *Dataset and Analyses for Extracting Schemas from Thought Records using Natural Language Processing.* 4TU.ResearchData `https://doi.org/10.4121/16685347`. (2021)
1. Burger, F., Neerincx, M. A. & Brinkman, W.-P. *Dataset and Analyses for Using a conversational agent for thought recording as a cognitive therapy task: feasibility, content, and feedback* 4TU Research Data Repository `https://doi.org/10.4121/20137736`. 2022