The social life of mental health chatbots

THE SOCIAL LIFE OF MENTAL HEALTH CHATBOTS

Eoin Fullam

Submitted for the degree of Doctor of Philosophy
Birkbeck, University of London,
February 2024

The work presented in the thesis is the candidate's own.

Abstract

This PhD project is an analysis of automated mental health therapy, so called 'mental health chatbots'. There are many kinds of mental health apps, but most claim to involve Cognitive Behavioural Therapy (CBT) or Mindfulness. My project is focused on CBT apps in general and CBT chatbots in particular. My research primarily concerns a smartphone-based mental health chatbot called ReMind, made by a company of the same name. The focus of my research is on the production of this technology. I am interested in what is going into this technology in terms of mental health/mental illness concepts, treatment styles, and the technical and economic conditions. The project covers a range of grounds: the history of computation, theories of subjectivity, analysis of therapeutic methods, and economic contextualisation feature in this work alongside ethnographic analysis. The aim of this project is to consider automated therapy from the perspective of ReMind - the software application and the company of the same name. It will rely on analysis of the app and ethnographic data gathered during fieldwork with ReMind. This is alongside analysis of other similar mental health chatbots, as well as looking at theoretical and journalistic material concerning chatbots, artificial intelligence, contemporary mental health treatment, and political economy. Drawing on historical and emerging scholarship on critical theory, science and technology studies, psychoanalytic theory and philosophy of computation, this project investigates the social life of mental health therapy applications and seeks to determine the underlying assumptions about subjectivity, consciousness and mental health that underpin them. I explore how these chatbots are on one hand produced in response to contemporary social, clinical, technical and economic conditions; and on the other hand, how they are conceptualised and put to work by their developers who have their own biases, assumptions and social conditions.

Word count: 97260

Table of Contents

List of Figures

Acknowledgments

Endless thank you to my supervisors. Thank you, Dr. Silvia Posocco for helping me navigate a course through difficult and oftentimes uncertain terrain. Thank you, Prof. Stephen Frosh for rigorous feedback, good humour and helping me stay the course. Both of your resounding patience and encouragement has been immensely appreciated.

So much gratitude to my parents, Briege and Dave, you have helped me in ways that you may not even know.

Maike Hitzeroth, for everything.
Donal Fullam, for the feedback and the chats.
Bushra Connors, for one last idea to tie things up.

I wish to acknowledge Sci-Hub and Library Genesis for providing me with an immense wealth of academic research - two beleaguered online resources that resist the enclosure of the commons that is scientific knowledge.

Introduction

This PhD project is an analysis of automated mental health therapy, so called "mental health chatbots".[1] There are many different kinds of mental health apps, but most claim to involve Cognitive Behavioural Therapy (CBT) or Mindfulness. My project is focused on CBT apps in general and CBT chatbots in particular. My research primarily concerns a smartphone-based mental health chatbot called ReMind, made by a company of the same name. ReMind is a pseudonym due to a nondisclosure agreement I made with the company when I did my fieldwork with them. A chatbot is a computerised conversational computer program; therapy chatbots tend to operate through text, but some are now introducing natural language processing allowing users to take part in a spoken conversation with the chatbot. The focus of my research is on the production of this technology. I am interested in what is going into this technology in terms of mental health/mental illness concepts, treatment styles, technical and economic conditions. The research is focused on the development and not the users of computerised treatment because I want to approach this technology in terms of the social, economic, psychological and technological tendencies which are historically situated and have culminated in this novel form of mental health intervention. The reason for this focus is because my methodological approach contends that mental health chatbots are the product of social concepts of mental health but also, due to being expressions of these concepts in terms of causes and treatment, *produce* concepts of mental health. This includes the term 'mental health' itself, which, as will be discussed, encapsulates specific ways of thinking about how we experience and express mental suffering.

The research is interdisciplinary and engaged in a theoretical debate while supported by an ethnographic component. This means that the project is not an ethnography as such, but covers a range of grounds: the history of computation, theories of subjectivity, analysis of therapeutic methods, and economic contextualisation feature in this work alongside ethnographic analysis. I focus primarily on ReMind, and secondarily on competing apps, primarily 'Wysa'[2] and 'Woebot',[3] but others[4] are touched on also. These applications all offer a form of mental health therapy delivered by an AI chatbot operating through online text messaging services. They are functionally very similar to ReMind in that they offer a form of automated 'talking therapy' which purports to use AI techniques,[5] and as such, will often be used as examples and for analysis. The aim of this project is to consider automated therapy from the perspective of ReMind - the software application and the company of the same name. It will rely on analysis of the app and ethnographic data gathered during fieldwork with ReMind. This is alongside analysis of other similar mental health chatbots, as well as looking at theoretical and journalistic material concerning chatbots, artificial intelligence, contemporary mental health treatment, and political economy. Drawing on historical and emerging scholarship on critical theory, science and technology studies, psychoanalytic

[1] http://www.x2ai.com (Last accessed on 22/11/2023)
[2] https://www.wysa.com (last accessed on 03/12/2023)
[3] https://woebothealth.com (Last accessed on 01/11/2023)
[4] Cass (https://www.cass.ai/), Elomia (https://elomia.com/), Nuna (https://www.nuna.ai/), Youper (https://www.youper.ai/) (All last accessed 10/01/24)
[5] "Our proprietary technology combines decades of research in psychology with advanced AI to assess symptoms of anxiety, depression, and other mental health needs and respond with empathy" (https://woebothealth.com/what-powers-woebot/). The definition and usage of AI in these chatbots will be discussed in further chapters.

theory and philosophy of computation, this project investigates the social life of mental health therapy applications and seeks to determine the underlying assumptions about subjectivity, consciousness and mental health that underpin them. I will explore how these chatbots are on one hand produced in response to contemporary social, clinical, technical and economic conditions; and on the other hand, how they are conceptualised and put to work by their developers. My project is guided by an overarching double-sided question:

What are the conditions of possibility for automated mental health treatment, and how does automated treatment alter subjectivity?

The question is two-sided because each aspect of the question implies the other: the conditions of possibility for this technology are social, historical, technical and economic, the human subject is bound up in these same conditions. This question will draw upon psychosocial theories of 'the subject'. 'The subject' equates to the individual insofar as one is produced by one's social environment. This means that a concept of the subject, and of mental health, can be reconstructed through analysis of the practices and objects involved in the production and deployment of this technology.

We are, at an accelerating rate, required to adapt to new technologies without an understanding of the consequences of these technologies. The automation of mental health therapy is already underway, but critical research is alarmingly lacking. On one hand, there is very little consensus in regard to the causes which give rise to 'mental illness', 'poor mental health', or any other framing of the ways in which we suffer in our minds. On the other hand, technology and the automation of human labour are similarly contested in terms of social consequences. Automated therapy poses as revolutionising the cost and accessibility of mental health treatment, but raises concerns over how therapy will be delivered and received in digital form, and over who benefits, who is excluded, and who might be exploited. 32 million people access the mental health section of NHS Choices website every year,[6] but mental health service in the UK is chronically underfunded.[7] Automation is often seen as a solution to these problems, where the scale of the crisis is matched by the possibilities of scale promised by technological automation. However, handing over of human labour onto technology frequently has unforeseen consequences,[8] and the effects of automation are often only understood in hindsight. However, by the time a technology has become ubiquitous, "tacit normative consensus"[9] is achieved, and it becomes extremely difficult to put the genie back in the bottle.

This research will provide important insight and analysis to: the makers of these technologies, who's concern for practical implementation might overshadow contemplation;

[6] The AHSN Network (2017) 'Disruptive and Collaborative Innovations in Mental Health'. Online: https://thehealthinnovationnetwork.co.uk/wp-content/uploads/2018/12/Mental_Health_Brochure.pdf (Last accessed 18/05/22)
[7] British Medical Report (2018) 'Lost in transit? Funding for mental health services in England'. Online: https://www.bma.org.uk/collective-voice/policy-and-research/public-and-population-health/mental-health/funding-mental-health-services. (Last accessed 04/02/21)
[8] Bordenkircher, B.A. (2020) 'The Unintended Consequences of Automation and Artificial Intelligence: Are Pilots Losing their Edge?' *Issues in Aviation Law and Policy*, Vol. 19 no. 2
[9] Feenberg, A. (1994) 'The Technocracy Thesis Revisited: On The Critique of Power'. *Inquiry*, 37. pp.85-102

to mental health therapists who may be affected by the loss of demand for their services and the changing nature of treatment due to automation; to policy makers whose priority it is to legislate over public health and who are concerned with commercialisation of industries involved in the care of mentally suffering and otherwise vulnerable people. Finally, to the users must be informed of the impact that this new technology will have on their own mental health, and how mental health itself will be transformed.

ReMind

ReMind is the name of both the app and company. ReMind began as a company around 2015 under the name LifeTag, their aim was to produce a physical health app, but they moved on to a therapy chatbot soon after. LifeTag started as a small tech start-up, comprising just Jeff and Reese, the two founders, who built a chatbot to deliver simple mental health assistance. This bot evolved into ReMind as we know it today. They currently employ over 100 employees. While I conducted my fieldwork in summer 2022, ReMind was expanding in terms of both employees, field offices, investor funding and partnerships. The ReMind app is described by the company as a "relational AI-powered agent, for emotional coaching, self-help and mental wellness".[10] The app was launched in 2017 and is promoted as providing intervention in high-risk groups[11] using two primary methods:

- A chatbot
- A library of tools for self-help

Contemporary mental health chatbots usually do two things: 1. offer a text-based chat through which the user can hold a conversation with a chatbot, and 2. offer various self-help activities; these activities tend to be influenced by Cognitive Behavioural Therapy (CBT) or Mindfulness style treatment methods. There are currently thousands of apps which offer mental health support and a small number of chatbots which do so. Woebot and Wysa are currently the most popular and sophisticated mental health chatbots. Popular CBT chatbots like ReMind, Woebot and Wysa are not really chatbots as properly defined, they are more accurately (although less catchily) defined as responsive multimedia CBT apps that employ a personable character to give the impression that the user is having a conversation. The bot is a 'rules-based chatbot', meaning that ReMind is not capable of generating its own responses and must rely on a bank of pre-written responses. One review of Woebot stated: "The conversations don't veer too far off course…It is, as the creators call it, a 'choose-your-own-adventure' self-help book".[12] 'Genuine' artificial intelligence, i.e. generating spontaneous responses, is currently not compatible with therapy chatbots. A 'talking cure' style chatbot is not currently commercially available, however, beyond the scope of this project, contemporary generative chatbots such as the GPT series can be tasked with 'speaking like' a psychotherapist.[13]

[10] Source withheld to maintain anonymity
[11] Source withheld to maintain anonymity. The claim that ReMind is useful for "high risk groups" is made in an article by one of ReMind's psychologists hosted on ReMind's website
[12] Jope, J. (2017) 'I Talked to Woebot for a Month: Here's How It Went'. *Depression Defined.* Online: https://www.depressiondefined.com/self/woebot-part-one (Last accessed 21/06/22)
[13] Pirnay, E. (2023) 'We Spoke to People Who Started Using ChatGPT As Their Therapist'. *Vice Magazine.* Online: https://www.vice.com/en/article/z3mnve/we-spoke-to-people-who-started-using-chatgpt-as-their-therapist (Last accessed 10/10/23)

Structure

The structure of the thesis is divided into two parts, part one comprising the preliminary work and part two comprising the analysis chapters:

Part 1

1. Introduction. This present section introduces the project and provides a general overview of smartphone-based mental health apps to provide a sense of the field. This summarises the broad scope of the various mental health interventions that are available in smartphone app form. A number of different apps will be described to illustrate the range.

2. Literature Review. This section discusses the range of research so far done on computerised mental health interventions. It covers quantitative and qualitative research but also media articles about these apps. The reason for this is that, as will be shown, 'research' and 'promotion' in this technological field are often intertwined.

3. History Section. This will undertake a tour through conceptual lineages which thread through the historical development of behavioural psychology, electronic computation, cognitive science, cognitive & behavioural therapy and Mindfulness - a range of historical tendencies which have become intertwined in automated mental health treatment.

4. Methodology and Methods. An overview of the ethnography and a discussion about what was planned for the fieldwork and what ended up happening. This section will cover the work I did in planning and executing the fieldwork with ReMind which took place between January and July of 2022. It will also serve to contextualise ReMind in terms of how the company formed and how it currently operates.

Part 2

Chapters five, six and seven deal with a range of technical aspects of ReMind: how it is built, decisions that its makers have made throughout this process, and the consequences of these decisions. They focus on, respectively, the mental health activities that the app provides, how the bot is designed to converse with users, and how ReMind's intervention operates at scale. Chapters eight and nine pursue findings from the previous chapters but concern wider contexts: respectively, how relationships with chatbots are formed and maintained, and the economic context in which ReMind operates and which it responds to.

5. Digitisation. This chapter continues from the questions raised in the history section and focuses on the specific treatment methods which the ReMind bot provides. CBT and Mindfulness techniques provide the core methods from which ReMind draws in its automated intervention. It asks the question 'why CBT'? Why does chatbot therapy and almost all computerised automated mental health treatment claim CBT to a greater extent, and Mindfulness to a lesser extent as the basis for their treatment forms? The conceptual lineage of CBT and its connections to the invention of the computer will be explored to show that 'computerised therapy' is possible due to CBT comprising an already technological and algorithmic form of treatment.

6. Conversation Design. This chapter is about how ReMind observes, tracks, surveys, or otherwise situates the users of their chatbot. The mediating effect of the chatbot will be of primary concern: ReMind designs and controls all aspects of the chatbot in terms of coordinating its conversation, adding and removing features, and developing its therapeutic style, but the technical aspects of the chatbot also determines how the ReMind team understands who the 'user' is, influencing decisions about app design. The aim of this chapter will be an attempt to illustrate this circular dynamic, to discuss what this means for the type of treatment offered by the bot.

7. Macro-Treatment. This chapter deals with how ReMind responds to the users of the bot as produced and mediated by the bot. ReMind adjusts the mental health of the 'user', who has been generated and aggregated through the ReMind bot, in mass-form through mediation of the bot in terms of feedback mechanisms and adaptive techniques. While it is possible to view individual user conversations and to adjust conversational content in response, decisions about conversation design are also made in response to users being aggregated into large groups in order to be treated as classes or clusters.

8. Suspension of Disbelief. This chapter involves a discussion about how it is possible to engage with computerised agents or avatars on an interpersonal level - subjective engagement with relational artefacts.[14] While the development of a 'virtual therapist' which performs the same function as a human may not be feasible, the designers of mental health apps are concerned with understanding just how a user interacts with their apps on an interpersonal level.

9. Technocracy. This chapter explores a final paradox which takes into account the issues raised in the previous chapters. This paradox is related to how individuals, when presented with an automated mental health treatment method must both assume a sense of personal responsibility and at the same time, forgo responsibility. The paradox is approached through a discussion about the economic context in which ReMind is situated.

10. Conclusion. While an overview of the work done on this project comprises this chapter, the various logics which have been identified in the analysis chapters are discussed in terms of an alternate perspective. I use the distinction: *instantiated*, as opposed to *instrumental*, to draw out, on one hand, the approach which guides ReMind, and on the other hand, to open up alternative possibilities for design. I use 'instantiated' to mean that the app will be considered in terms of an intervention which is integral to its workings and the interaction between it and the user rather than as an effect of those workings.

[14] Turkle, S. et al (2006) 'Relational artifacts with children and elders: the complexities of cybercompanionship.' *Connection Science*, 18:4. pp.347-361

Mental Health Applications Overview

Modes, Methods, Activities, and Features

While this thesis is focused on a single mental health chatbot app - ReMind - it also looks at other chatbots and non-chatbot mental health apps to make comparisons, distinctions, and to discuss the field of automated treatment in general. This section provides an overview of the various ways that mental health assistance is provided through smartphone applications that are available to download and use. What follows is a brief overview of the two distinct modes of delivery (automated and mediated) through which apps operate, the methods of treatment provided (CBT, Mindfulness, etc.) and the features (journaling, mood-tracking, etc) through which these are delivered. Searching 'mental health' on the Google Play Store brings up hundreds of options, but they can be broadly put into a number of categories depending on the mode of treatment that the apps employ and the mental health treatment methods that inform them. There is also a large range of different features, some of which will be mentioned, but more attention will be given to these in the analysis chapters. Features are distinct from methods, which refer to the underlying mental health techniques which inform and are delivered through the features. Treatment methods vary among the many different mental health apps, but most offer a few choices from within a narrow range, usually informed by CBT-based and meditation-based methods. Apps provide a range of these different features in varying ratios and in different combinations. Throughout this thesis, 'features' refers to the various suggested activities (such as breathing exercises), app-actions (such as push-notifications) and tools (such as mood-trackers) that the apps offer for the user to engage with, comprising, on different levels, the interface between the user and the treatment methods that the apps are informed by. Most apps, when downloaded, begin with a series of questions asking the user about their treatment needs (help with sleeping, social anxiety, depression, etc.) and goals (be more confident, dwell less on negative thoughts, etc.). These choices then determine the methods which will be provided within the range of features programmed into the app, the depth and range of which depend on the sophistication of the app. The distinctions between 'modes', 'methods' and 'features' are my own and do not conform to an established convention, they are used here to help differentiate between the available range of smartphone applications. These terms are used generally but not exclusively throughout the thesis, if other terms are used (such as 'technique', or 'style'), reasons for this will either be self-evident or provided.

Modes

Mental health apps use two distinct modes of treatment delivery - automated and mediated. Some of the more popular apps provide both modes of treatment, but most apps provide automated treatment in some form, being cheaper to develop, with mediated treatment being confined to a small range of apps.

Automated

Automated treatment involves no immediate human intervention, with the app providing a discrete, standalone service. The less sophisticated apps aim to provide a fully automated service, with assistant or CBT apps simply providing simple features such as list making, reminders, and text-logging. More sophisticated automated apps aim to simulate the clinical experience – a 'therapist' chatbot which the user communicates with.

Mediated

Mediated treatment involves access to a human in some way, either as an extra feature - an app might predominantly operate through automated measures but could offer access to a human therapist for a fee - or an app's sole feature might be access to a human therapist. Two forms of mediated treatment predominate: access to a clinical professional through text or video, and access to other users, which is usually in the form of private forums where users can share personal insights, offer each other support and share mental health strategies. Some apps offer treatment crossing both modes, for example an app might predominantly act as a mental health assistant while also providing access to a private forum.

Methods

Meditation

The majority of the most popular apps provide meditation or Mindfulness treatment. These treatments often take the form of encouraging the user to make time during the day to survey their mental state and to try to sort through their immediate sensations or recent experiences. Meditation can be guided or non-guided; both offer the same techniques, but guided meditation offers a video, audio, or text explainer to provide context and motivation whereas non-guided provides, usually in list-form, the steps which the user must follow to perform the meditation task. Automated (i.e. not providing access to a therapist) meditation apps seem to be the most popular form of mental health apps, likely due to being much cheaper to produce and access than 'mediated' therapy apps. It is difficult to clearly assess the popularity of apps, as they are available on different formats (Android and iOS), but Headspace, and Calm, both automated meditation-based apps, have received the most media attention, downloads, and user reviews, indicating that they are the frontrunners in smartphone-based treatment. Headspace states: "Meditation has been shown to help people stress less, focus more and even sleep better. Headspace is meditation made simple. We'll teach you the life-changing skills of meditation and mindfulness in just a few minutes a

day".[15] Calm is promoted as: "a leading app for meditation and sleep. Join the millions experiencing lower stress, less anxiety, and more restful sleep with our guided meditations, Sleep Stories, breathing programs, masterclasses, and relaxing music. Recommended by top psychologists, therapists, and mental health experts".[16]

Therapy/Counselling/Coaching

Some apps operate as a means to accessing a human therapist or counsellor, the app being a mediating tool between user and therapist. Therapy apps are similar to accessing a therapist by phone or through video link in that the user is communicating with a human therapist, most apps offer this service through text while some therapy apps like BetterHelp[17] offer video communication. BetterHelp is one of the few apps that strictly conforms to a mediated therapy mode in which the app acts as a communication device between user and therapist. BetterHelp offers "…over 10,000 counselors in BetterHelp, each with at least 3 years and 2,000 hours of hands-on experience. They are licensed, trained, experienced and accredited psychologists (PhD/PsyD), marriage and family therapists (MFT), clinical social workers (LCSW), licensed professional counselors (LPC), or similar credentials".[18] Text conversation with a therapist is often done through a message thread style format, where the user and therapist can scroll back through the previous messages.

Some apps simulate a talking therapist or companion ('chatbot'). Simulated therapist apps often claim to use 'artificial intelligence' to create natural feeling conversations, it is unclear what this means, as it seems that most 'AI' chatbot therapists rely on tightly scripted conversation structures rather than generating their own responses. Wysa offers both an AI chatbot companion and the opportunity to speak to a "well-being coach". Wysa describes itself as a "stress, depression & anxiety therapy chatbot" which offers "a mood tracker, mindfulness coach, anxiety helper, and mood-boosting buddy, all rolled into one".[19] Along with Wysa, Woebot is one of the most popular, chatbot-based therapy apps, or "AI-powered, personalised emotional support platform that detects users' symptoms and delivers clinically-validated psychological interventions to achieve better outcomes".[20] Intervention takes the form of a message thread between the user and Woebot, with conversation often being tightly scripted - users tend to only have the ability to respond to Woebot's prompts by choosing from a range of pre-provided responses. Woebot also provides a 'mood tracker' in which the user can track their mood over time and a 'gratitude journal' in which the user can write free-form pieces and save them.

Self-Help

The vast majority of mental health apps are simple (and sometimes crude) CBT apps which

[15] https://www.headspace.com/headspace-meditation-app (Last accessed 16/04/23)
[16] https://play.google.com/store/apps/details?id=com.calm.android&hl=en_GB&gl=US (Last accessed 16/04/23)
[17] https://www.betterhelp.com (Last accessed 18/04/23)
[18] https://play.google.com/store/apps/details?id=com.betterhelp&hl=en&gl=US (Last accessed 18/04/23)
[19] https://play.google.com/store/apps/details?id=bot.touchkin&hl=en_US&gl=US (Last accessed 18/04/23)
[20] https://woebothealth.com (Last accessed 20/12/23)

allow the user to record their daily moods, take standardised assessments, and to correlate moods with behaviours over time. While many mental health apps claim to provide CBT-based treatment, this is usually subsumed into the overall 'wellness' treatment offered by the app; these apps rarely offer strictly defined CBT courses. This is due, as with counselling apps, to the necessity of providing extra features available to the smartphone format. CBT Companion along with other apps like Mindshift[21] offer traditional CBT self-treatment in digital form. Its developers claim their app "is the most comprehensive CBT app that exists today with easy to follow visual tools".[22] CBT Companion, like its contemporaries, comprises a number of different courses of treatment which can be selected by the user. Like many other mental health apps, CBT Companion treads a fine line between claiming to offer genuine therapy and being simply a 'wellness' app.

Companion

While most apps, especially apps which include a chatbot, include aspects which make the bot more companionable, such as providing friendly encouragement and check-ins, there are also many (non-therapy) chatbots available that are promoted as 'virtual companions'. These are sometimes then used in therapeutic ways by users.[23] Replika is the most popular companion style chatbot which uses artificial intelligence systems to shape its responses to individual users, offering a long-form style conversation. Replika is an 'AI companion' which encourages the user to form an emotional connection with a chatbot character. While not a therapy app, Replika is still promoted by its developers as offering emotional support and providing mental wellbeing. "Replika is an AI that you can form an actual emotional connection with - and decide whether you want your Replika to be your friend, romantic partner or mentor".[24]

Activities and Features

Activities simply refers to the various self-help procedures that ReMind provides to the user. Sometimes mental health app makers refer to these as 'exercises', but I use the term activity to distinguish from physical exercise, which many apps also include. To avoid confusion, I use the term to refer to all procedures, from cognitive restructuring, to breathing techniques, to yoga activities. Features are all of the means through which the activities are delivered and are tied to the technical affordances of smartphones: touchscreen displays, microphones, speakers, web-access, etc. For example, a mental health app might offer video tutorials of different CBT or Mindfulness activities, a chatbot might offer its service via text or through a spoken conversation. Often mental health apps provide journaling features which allow the user to log and track their moods. Other features include assistant-style support; this depends on the ability to provide timed reminders. This is not an exhaustive list but indicates how 'feature' is used throughout the thesis.

[21] https://apps.apple.com/ca/app/mindshift-cbt-anxiety-relief/id634684825 (Last accessed 20/01/24)
[22] https://play.google.com/store/apps/details?id=co.swasth.cbtcompanion (Last accessed 20/01/24)
[23] Pirnay, E. (2023) 'We Spoke to People Who Started Using ChatGPT As Their Therapist' *Vice Magazine.* Online: https://www.vice.com/en/article/z3mnve/we-spoke-to-people-who-started-using-chatgpt-as-their-therapist (Last accessed 10/10/23)
[24] https://play.google.com/store/apps/details?id=ai.replika.app&hl=en_US&gl=US (Last accessed 02/02/24)

Chapter Two: Literature Review

2.1 Introduction

This literature review comprises an overview of research articles which focus on therapy chatbots. The purpose is twofold: to gather and discuss qualitative and quantitative research so far conducted on mental health chatbot apps; and to establish a sense of the academic-commercial field of research which is involved in documenting and analysing this emerging software. The reason for this approach is because, as companies making commercial products, the makers of these apps conduct research not just to assess their effectiveness but also to establish 'effectiveness' as a defining measurement and to promote their apps as commercial products. This literature review covers quantitative research and some qualitative research: quantitative research comprises the majority of research of mental health apps. This balance will be discussed in terms of the demand for standardisation and measurement precision. Effectiveness tends to be assessed using the results of patient outcome forms and through randomised controlled trials. Comparisons are usually also made using statistical measures and often make comparisons between the automated computerised version of treatment to its closest non-automated version. An example of this would be comparing a chatbot which provides CBT to a non-chatbot CBT computer program.

This literature review is divided into two sections: academic research and commercial research. Non-commercial academic literature is broadly divided between overviews of already conducted research and analysis of individual therapy apps. Most academic research relies on patient outcome forms, often using Likert-scale, questionnaire-type measurements to measure the treatment. The use of more advanced statistical tools, which are often used by the makers of mental health software, is not usually possible because these require access to the data which are internal to the bots. Some research involves a bot that has been created specifically for the study, but most research uses commercially available software. Much of the academic literature is limited to small participant numbers and short timescales, meaning that research tends to fall into 'preliminary' investigations, from which further research avenues can be mapped out. The reasons why these investigations are limited is that large scale internal app-data such as user demographics, app-usage, etc are exclusively available to the makers of commercial apps. While some research does involve creating bespoke apps. These apps suffer from limitations such as small development teams, and narrow participant recruitment scope. On the other end of the spectrum, companies that make mental health chatbot apps conduct their own research in order to assess the effectiveness of their interventions. This research takes on two forms: one involves measuring the results of patient outcome forms, and the other is through assessing data gathered from the apps themselves. The latter assessment will be focused on as this type of measurement distinguishes commercial research from its non-commercial academic counterpart in that it comprises 'big-data' gathering and assessment techniques.

The aim of this literature review is to scrutinise the data-gathering and measurement techniques in order to survey the available knowledge and to show that this knowledge very often depends on a non-critical basis in which the measurement techniques and outcomes are taken at face-value, leading to incomplete, inconsistent, or questionable outcomes.

2.2 Academic Research

Overviews

Some research of mental health chatbots involves conducting overviews in the form of a descriptive survey of available mental health apps, a survey of published research, or a meta-analysis or scoping review of published research. Overviews tend to paint in broad strokes, usually explaining how chatbots work and their relevance in the context of the broad mental health treatment landscape. Meta-analyses tend to try and form systematic integrations of previously published research. As we will see, both forms of overview have their benefits and drawbacks. Aditya Vaidyam et al. conducted a review of research which covers mental health chatbots, to explore their role in screening, diagnosis and treatment of mental illnesses. They conclude “there is no consensus on the definition of psychiatric chatbots or their role in the clinic".[25] The study found ten texts that were deemed eligible for review, noting that the majority (75%) of the studies discovered using their initial search parameters were devoted to engineering problems: “we found the academic psychiatric literature to be surprisingly sparse".[26] The review concludes that “Preliminary evidence for psychiatric use of chatbots is favourable. However, given the heterogeneity of the reviewed studies, further research with standardised outcomes reporting is required to more thoroughly examine the effectiveness of conversational agents".[27] Vaidyam et al. note that while chatbot therapy is largely unstudied and untested, their usage by patients is becoming more common. They also point out that data-safety is often unaccounted for, with personal information being stored and transmitted using potentially insecure means. Vaidyam et al. also note that “it is also important to consider the potential relationships that may be formed with chatbots".[28] They go on to very briefly point out that users might be negatively affected if access to a bot with which a bond has been formed becomes limited or revoked. Ethical issues such as this are sometimes mentioned in the literature, but treatments of issues surrounding user-bot relationships and personal data-management are sparse.

Meta-analytic scoping reviews in which statistical analysis is used to integrate published research are fraught with challenges. Ahmad Jabir et al. conducted a scoping review to identify the “types of outcomes, outcome measurement instruments, and assessment methods”[29] used in studies which assess chatbot-based mental health treatment. 32 studies were included in the review comprising “experimental primary studies, such as RCTs, cluster randomized trials, quasirandomized trials, controlled before-and-after studies, uncontrolled before-and-after studies, interrupted time series, pilot studies, and feasibility studies".[30] The review concludes that:

---

[25] Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., & Torous, J. B. (2019) ‘Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape’. *Canadian journal of psychiatry, 64*(7) pp.456-464. p.457
[26] Ibid. p.458
[27] Ibid. p.459
[28] Ibid. p.463
[29] Jabir, A. I., Martinengo, L., Lin, X., Torous, J., Subramaniam, M., & Tudor Car, L. (2023) ‘Evaluating Conversational Agents for Mental Health: Scoping Review of Outcomes and Outcome Measurement Instruments’. *Journal of medical Internet research*, *25*. p.1
[30] Ibid. p.2

> The diversity of outcomes and the choice of outcome measurement instruments employed in studies on CAs[31] for mental health point to the need for an established minimum core outcome set and greater use of validated instruments. Future studies should also capitalize on the affordances made available by CAs and smartphones to streamline the evaluation and reduce participants' input burden inherent to self-reporting.[32]

The two points made in the conclusion will be discussed further below, but what they mean is that of 32 studies reviewed, the range of different outcome measures used is too wide and must be standardised in order to establish some kind of objective measure to assess chatbot therapy. The review found that of the 32 studies, 150 of the 203 outcome measurement instruments were unique instruments which were invented or modified for each particular study, and that 83.7% of the outcome measurements were self-reported questionnaires. Essentially, most of the studies created their own evaluation methods, leading to problems with replication and external evaluation. In their 'Recommendations for Future Research' section, Jabir et al. note that user attitudes and perceptions towards conversation agents must be evaluated in some way, as this is mostly absent from the literature. While some qualitative research does touch on user-experience, there is very little in the way of analysis of how and why people use mental health apps.

Descriptive overviews usually discuss chatbots in terms of explanation as to their rise in popularity due to a recent increase in unmet mental health needs and of their contextual status among other computerised mental health interventions. Elaine M. Boucher et al.'s overview is indicative of this approach in which they tie the two aspects together by discussing how the introduction of a chatbot to a computerised mental health intervention is often done to reduce user-attrition which is a common problem for non-chatbot computerised interventions.[33] Boucher et al.'s study comprises an overview of the various types of mental health chatbot, their most common functions ("diagnosis, content delivery, and symptom management".[34]) and a case study of a mental health chatbot called Anna. Half of Boucher et al.'s study involves discussing Anna, and a pilot test which surveyed 203 users of the app is analysed. Unlike most other (non-commercial) studies of this type, Boucher et al. appear to have access to the bot's conversation logs. Their analysis is based on interpretation of these logs. It is worth quoting their conclusion in full as this is highly representative of other studies:

> The availability of effective AI-supported interventions is an important avenue to reduce the longstanding burden on practitioners and improve the increasing shortage of mental health professionals. Although preliminary research suggests chatbots are perceived favorably and may help to improve engagement and mental health

[31] Conversational agents
[32] Jabir, A. I., Martinengo, L., Lin, X., Torous, J., Subramaniam, M., & Tudor Car, L. (2023) 'Evaluating Conversational Agents for Mental Health: Scoping Review of Outcomes and Outcome Measurement Instruments'. *Journal of medical Internet research*, *25*. p.1
[33] Boucher, E.M. Harake, N.R. Ward, H.E. Stoeckl, S.E. Vargas, J. Minkel, J. Parks, A.C. & Zilca, R. (2021) 'Artificially intelligent chatbots in digital mental health interventions: a review'. *Expert Review of Medical Devices*, 18:sup1. pp.37-49. p.38
[34] Ibid. p.39

> outcomes, more rigorous tests of chatbots within DMHIs[35] are needed. In particular, more research on how chatbots may help to improve mental health outcomes compared to other digital interventions without chatbots is an important next step, as is considering how individual and contextual factors might influence the impact of mental health chatbots.[36]

Some overview papers either make questionable claims or provide vague information. In a book chapter offering an overview of mental health chatbots Kerstin Denecke et al. state that there are two different types of chatbots: "unintelligent (rule-based) chatbots which generate their dialogue based on some predefined rules or decision trees, and intelligent chatbots which use Artificial Intelligence (AI) to understand the context and intent of a user utterance and respond to it".[37] This is not accurate: on one hand, there is a large range of different methods for designing conversational software and implementing this through chatbots beyond these two examples. On the other hand, rules-based chatbots may indeed have the ability to "understand" (through various automated interpretive measures) context and intent but are bound to respond through their predefined rules; an 'intelligent' chatbot is more associated with being able to generate its own responses, not just to understand context. These errors are common in research which discuss complex technical systems when attention is not paid to defining those systems in terms of their underlying mechanisms. Researchers must be vigilant to ensure that they are not relying on preformed assumptions and to question their own definitions. This is especially true in the case of mental health software which is usually produced in a commercial context: definitions, claims and references may be inherited from the companies which make the apps in question and as such require extra scrutiny. This lack of scrutiny often mars the available research. The chapter also mentions that "Two popular chatbot platforms used today are Wysa and SERMO".[38] While Wysa is clearly a popular chatbot in terms of the number of downloads on the various app stores, SERMO does not even appear to be available to download. In fact, it seems that at least one of the researchers (Denecke) involved in writing the overview is also involved in the development of SERMO.[39] We are not informed about this in a conflict-of-interest section of the chapter. The combination of error and omission does not inspire confidence in the findings. In this respect, findings in these types of overviews tend to direct their critiques in terms of the technical capabilities of chatbots, user-data storage ethics, user-safety and accountability.

Quantitative Research

Most current research into mental health treatment focuses on survey-based or statistical analysis of treatment effectiveness, whether this research is on face-to-face or computerised treatment. This might be because evaluation of computerised treatment lends itself much more to the analysis of data for statistical analysis. For instance, a CBT app can include in its software participation and attrition rates, user-satisfaction ratings, etc. As mentioned

[35] Digital mental health interventions
[36] Ibid. p.44
[37] Denecke, K. Abd-Alrazaq, A. Househ, M. (2021) 'Artificial Intelligence for Chatbots in Mental Health: Opportunities and Challenges'. In: Househ, M. Borycki, E. Kushniruk, A. (eds.) *Multiple Perspectives on Artificial Intelligence in Healthcare.* New York: Springer. p.1
[38] Ibid. p.2
[39] Ibid.

above, mental health chatbot research either involves creating a bespoke app or depends on already available apps. Shinichiro Suganuma et al. conducted a study using a chatbot (SABORI) developed by the laboratory of one of the authors. SABORI is described as “a Web-based unguided ICBT application available for use on a smartphone, tablet, or computer browser for company employees, university students, and housewives".[40] The study measured the mental health of 191 participants who completed a 15-day course using the chatbot compared to a control group. The measurements comprised the World Health Organization-Five Well-Being Index, Kessler 10 and Behavioral Activation for Depression Scale (BADS). These are all outcome measure-based Likert scales. The study claims that “The addition of the agent-based dialog feature potentially affected the strengthening of the therapeutic alliance between the system and the user". It is unclear how this claim was reached within the study however, and it is interesting to note “there is a need for further detailed research into the factors underlying the effect, as well as the component factors of the therapeutic alliance when utilizing agent".[41] It seems that only one such study has been conducted so far, by Woebot labs.[42] Woebot’s study has not been included in this literature review as it is analysed in chapter eight of this thesis. Suganuma et al. concluded that “This research can be seen to represent a certain level of evidence for the mental health application developed herein, indicating empirically that internet-based cognitive behavioral therapy with the embodied conversational agent can be used in mental health care".[43] Note the use of the term ‘mental health care’; as will be discussed, the various alternate terms for the dynamic between the user and the bot like ‘therapy’, ‘intervention’, ‘coaching’, and ‘care’ are often contextually dependent.

Research which involves purpose built chatbots tend to lack scope and depth. This lack is observable in both the data that is generated and the subsequent analysis. Research projects in which a chatbot is built as part of the project incur benefits and drawbacks which are reversed when using a bot that is already available. Researchers have full control of not just the design and implementation of the bot but also in terms of data collection; they are not confined to ‘external’ measures such as user satisfaction ratings, but they also have access to internal analytics, measures that can be derived from the workings of the app. The drawbacks involve lack of relative software development expertise and production time. On the other hand, using an already available app means that measures are confined to polling users in an ‘external’ manner, without access to internal analytics drawn from the software. Apart from Boucher et al.’s study which involves access to the bot’s conversation logs, academic researchers are confined to data which they can generate using their own methods, and not data which are generated from within the app. Commercial apps tend to be more technically sophisticated due to larger development teams with a range of expertise, from software developers to clinical psychologists. This type of research tends to

[40] Suganuma, S., Sakamoto, D., & Shimoyama, H. (2018) ‘An Embodied Conversational Agent for Unguided Internet-Based Cognitive Behavior Therapy in Preventative Mental Health: Feasibility and Acceptability Pilot Trial’. *JMIR mental health*, *5*(3), e10454. p.3
[41] Ibid. p.8
[42] Darcy, A. Daniels, J. Salinger, D. Wicks, P. & Robinson, A. (2021) ‘Evidence of Human-Level Bonds Established With a Digital Conversational Agent: Cross-sectional, Retrospective Observational Study’. *JMIR Form Res*, 5(5):e27868
[43] Suganuma, S., Sakamoto, D., & Shimoyama, H. (2018) ‘An Embodied Conversational Agent for Unguided Internet-Based Cognitive Behavior Therapy in Preventative Mental Health: Feasibility and Acceptability Pilot Trial’. *JMIR mental health*, *5*(3), e10454. p.1

involve very low participant numbers (often students), short timescales, and depend on patient self-reporting type measurements. One such study by Johan Nieva et al. using Woebot uses, “psychological distress assessment (PDA), pre-test and post-test stress level assessment (SLA), daily conversation assessment (DCA) and evaluation form".[44] These types of measurements usually involve self-assessment using scale-based evaluations using Likert scales. They tend to involve post-hoc user-satisfaction rating type scales, or checkbox tables where users can mark things like their preferred features, or their reasons for using the bot. Nieva et al.’s study also includes assessments of users’ conversations, in which users submit their conversations to the researchers, which “were selected, transcribed and forwarded to a psychologist for analysis".[45] Findings from analysis of conversations involve assessment of whether users willingly performed the bot’s suggested cognitive therapy activities, whether the users tended to enthusiastically converse with the bot, and whether the bot’s responses seemed appropriate or not.[46] This research involved 25 participants over a two-week period and is indicative of the participant levels and timescales involved in these research projects. Because of this, conclusions tend to be broad-based, speculative and indicate that “Further work will entail a deeper analysis…”[47] These research projects tend to present themselves as preliminary investigations to determine user assessments with an eye for possible further research.

In order to conduct research on mental health chatbots researchers usually depend on quantitative measurements. The striving for objectivity in mental health research has obvious benefits - results can be (potentially) externally validated, dangerous or risky methods can be screened out for having negative results, successful methods can be replicated, refined and further adapted. The demand for quantification is due to a requirement for standards across research projects, in order to claim scientific credentials: an objective assessment. This scientific objectivity was the aim of Aaron Beck and his cohort; however, standards tend to dip when research involves measures that are unique to individual studies, leading to difficulties in comparing and contrasting different studies. The inconsistent nature of research undermines the claim to objectivity and scientificity that the makers of mental health chatbots depend on. This paradox will be discussed in the second section. Throughout the quantitative process, the individual patients seem to get lost as they are gathered into aggregate models, and mental health itself becomes a numbers game. This reductionism inevitably leaves things out in order to provide a sound framework within which to operate. What gets left out is discussion of the patient’s expectations, mental associations, hopes, fears and attitudes towards the app - their relationship with it. Qualitative research ostensibly addresses this omission.

[44] Johan, N. Jose, J. Chaste, T. Ruzel, T. & Ethel, O. (2020) ‘Investigating Students’ Use of a Mental Health Chatbot to Alleviate Academic Stress’. *CHIuXiD '20: 6th International ACM In-Cooperation HCI and UX Conference.* p.4
[45] Ibid. p.7
[46] Ibid. p.7
[47] Ibid. p.9

Qualitative Research

The problems brought up when reducing subjective experience of mental health to quantifiable metrics are ostensibly addressed by the use of qualitative research in which user experience can be considered in terms of a wider range beyond that which can be captured by outcome surveys. We can understand the need to expand research beyond quantitative methods, not just in terms of assessing user experience, but also in terms of the historical and conceptual lineages that these types of mental health interventions draw on. One study by Kien Hoa Ly, while mostly focusing on quantitative measures, features a qualitative component. This involves categorising user-feedback of the bot into three themes with some underlying subthemes. The three themes are "Content", "Medium", and "Functionalities".[48] Analysis of these themes involves brief speculative commentaries on selected feedback. Small sample sizes are not exploited for more sophisticated qualitative methods such as interviews or case-studies, and no studies on chatbot-based interventions thus far published focus entirely on qualitative approaches. Qualitative studies of non-chatbot computerised CBT (cCBT) interventions have been conducted however, which are useful to look at for the potential that this type of research might have for chatbot-based interventions. Eight qualitative studies were reviewed and synthesised in a meta-analysis by Knowles et al.[49] The studies looked at user-experience of cCBT apps and include 'thick' description, open comments, and reflections by patients. The meta-analysis discussed such topics as how the patient situates themself in relation to the app, how they feel, and attitudes towards whether they feel cared for or not. Knowles et al. point out that while cCBT might pose as a suitable alternative to costly and time-intensive face-to-face treatment, the question as to whether patients deem this kind of treatment to be an acceptable alternative is rarely broached in research, they also point out that patients suffering from chronic depression may view computerised treatment differently at different times due to undergoing "identity shifts".[50] By focusing on user-experience rather than outcome surveys, cCBT can be considered from the perspective of the individuals undergoing treatment, rather than as an accumulation of survey responses. Knowles et al. identify:

> ...two key overarching concepts, regarding the need for treatments to be sensitive to the individual, and the dialectal nature of user experience, with different degrees of support and anonymity experienced as both positive and negative. We propose that these factors can be conceptually understood as the 'non-specific' or 'common' factors of computerised therapy, analogous to but distinct from the common factors of traditional face-to-face therapies.[51]

The desire to be treated as an individual was a prominent theme across 7 of the 8 papers reviewed in the meta-analysis, comprising sensitivity to personal needs and personal preferences but also sensitivity to how the patient feels, or how they subjectively experience mental illnesses such as depression. Knowles et al. identify that this theme is mostly

[48] Ly, K.H. Ly, A. & Andersson, G. (2017) 'A fully automated conversational agent for promoting mental well-being: A pilot RCT using mixed methods'. *Internet Interventions*, Volume 10. pp.39-46. p.43

[49] Knowles, SE. Toms, G. Sanders, C. et al. (2014) 'Qualitative meta-synthesis of user experience of computerised therapy for depression and anxiety'. *PLoS One*, 2014;9(1):e84323

[50] Ibid. p.2

[51] Ibid. p.1

discussed in regard to a lack of individualised treatment in cCBT.[52] The authors draw a distinction between 'complementary' and 'emulating' approaches to computerised therapy.[53] A complementary approach would maximise the unique aspects of technology to offer an option within the range of existing therapeutic treatments; an emulating approach would seek to offer some kind of virtual therapist that would imitate a human therapist. While the study by Knowles et al. shows that attention to individual experience is a vital aspect in researching mental health apps, their work is directed towards improving the effectiveness of the apps, making them more acceptable to those who engage with them. This is similar to considering mental health apps from a user satisfaction perspective, where user-adherence can be taken as the measure of the success of the intervention. Knowles et al. stress that computerised mental health treatment can be improved through consideration of interactivity, personalisation and support. While it is important that the development of computerised mental health takes into account user-experience beyond the use of user satisfaction rating scales, questions still remain as to what kind of treatment is being provided technologically. While patient experience is focused on in these studies, the research is still concerned with therapeutic effectiveness, which tacitly assumes that computerised therapy is treating the same thing as face-to-face therapy. This may not be the case. There is an argument to be made that mental health and mental illness assume different values depending on different contexts. This means that simply measuring treatment effectiveness without concern for the epistemological foundations on which that treatment is based might itself impose a covert influence on how mental health and mental illness are thought about.

2.3 Commercial Research

Self-Reporting

The makers of mental health chatbots conduct research into their own products for the purpose of publishing in scientific journals. There are both academic and promotional reasons for this research. All of the prominent therapy app companies assess their own apps in terms of studies based on the self-reports of users, and metrics designed and evaluated by the companies themselves. Woebot, a pioneer in conducting their own research, published one of the first chatbot-based mental health care research papers in 2017, their objective was to "to determine the feasibility, acceptability, and preliminary efficacy of a fully automated conversational agent to deliver a self-help program for college students who self-identify as having symptoms of anxiety and depression".[54] This study recruited 70 participants and divided them into a test group and a control group. The small size of the groupings reflects the fact that this is a feasibility study, and as such acts as an indicator for further research and development. Measurements used in the study were all outcome-based Likert scales: PHQ-9,[55] GAD-7,[56] PANAS,[57] and mixed-format questions (satisfaction scale, etc). The study also included a qualitative component:

[52] Ibid. p.5
[53] Ibid. p.10
[54] Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017) '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(2), e19. p.1
[55] Patient Health Questionnaire: nine item self-report measure.
[56] Generalized Anxiety Disorder: seven item self-report measure.
[57] Positive and Negative Affect Schedule: twenty item self-report measure.

> Participants' responses to open-ended questions were analyzed for the Woebot group using only thematic analysis and were reported as frequencies. Data were analyzed thematically using an inductive (data-driven) approach guided by the procedure outlined by Braun and Clarke. Data codes were generated systematically, then collated into "thematic maps" and applied to the entire dataset to generate frequencies.

In Jabir et al.'s comprehensive survey of research conducted on mental health chatbots it was found that almost all research depends on self-reporting.[58] Research has shown that it is difficult to determine exactly what is being measured in mental health self-reports, with those doing the reporting often conflating contextual factors and physical health with mental health. Daphna Levinson & Giora Kaplan's study on mental health self-rating in mental health interventions often leads to confusion between general well-being and mental health, arguing that "...[T]he automatic assumption that the self rated mental health functions as a proxy measure of psychiatric morbidity, and suggests that the self rated mental health is more closely related to subjective well-being".[59] They concluded that self-reporters would conceive of their own subjective happiness and well-being quite separately from diagnosis of mental illness, contrary to the measurement scales used in study designs:

> ...a sizeable percentage of respondents who were classified as having mental disorders perceived their mental health as good; on the other hand, respondents who did not pass the threshold for diagnoses perceived their mental health as fair/poor.[60]

Research into mental health chatbots, and mental health apps in general take this discrepancy into account, but in a way in which 'mental health' and 'wellbeing' are essentially conflated. This is done by identifying improved mental health through data generated from satisfaction ratings. Woebot conducted a short study assessing the feasibility of providing a chatbot to a self-selected postpartum population. The study assessed 96 participants using self-assessed outcome measures.[61] The study measured two outcomes: satisfaction with the chatbot using CSQ-8,[62] and therapeutic alliance using WAI-SR.[63] As this study also involves a small number of participants and includes two self-assessed questionnaires, it must be read as an exploratory work to gauge future investment in the field of maternity mental health care. The study concludes:

---

[58] Jabir, A. I., Martinengo, L., Lin, X., Torous, J., Subramaniam, M., & Tudor Car, L. (2023) 'Evaluating Conversational Agents for Mental Health: Scoping Review of Outcomes and Outcome Measurement Instruments'. *Journal of medical Internet research*, *25*

[59] Levinson, D., & Kaplan, G. (2014) 'What does Self Rated Mental Health Represent'. *Journal of public health research*, *3*(3), 287. p.122

[60] Ibid. p.122

[61] Ramachandran, M. Suharwardy, S. Leonard, S.A. Gunaseelan, A. Robinson, A. Darcy, A. Lyell, D.J. & Judy, A. (2020) 'Acceptability of postnatal mood management through a smartphone-based automated conversational agent'. *American Journal of Obstetrics & Gynecology*, Volume 222 Issue 1

[62] Client Satisfaction Questionnaire

[63] Working Alliance Inventory - Short Revised

> In this self-selected postpartum population, participants showed high satisfaction with and acceptability of a smartphone-based automated conversational agent in the 6-week postpartum period. In light of recent ACOG[64] initiatives aimed at increasing awareness, diagnosis, and treatment for perinatal mood disorders, chatbots should be further examined as a potential postpartum mental health resource.[65]

This study shows that self-reporting might be an adequate measure of general mental wellbeing when contextual factors are included, but this method of assessment proves to be unreliable as a measure of mental illness and of the effects of treatment, which according to Daphna Levinson & Giora Kaplan's study occupy a different conceptual framework to that of 'mental health'.[66] It is unclear whether this is due to definitions not being precise enough or if self-reporting lends itself to this differentiation. Definitional imprecision also marks the kinds of claims that therapy chatbot makers about whether they are making 'therapy apps' or not. In a study on the Headspace Mindfulness (non-chatbot) app, Louise Champion et al. note in their 'Limitations' section that their outcome measures relied on self-reported questionnaires,[67] meaning that the results are unreliable and open to the interpretation of the patient. This is a feature in almost all of the above quantitative and qualitative studies on chatbot-based apps. While it might be possible to conduct a study in which mental health professionals assess the participants, Champion et al.'s solution to this problem is for future studies to recruit larger samples in order to compensate for bias.[68] Mental health chatbot makers have access to an important source of data which helps to overcome this limitation: internal metrics.

Internal Metrics

The problem of recruiting enough participants to justify making definitive claims based on large sample sizes, combined with the lack of clarity involved in self-assessment appears to be solved through the use of internal metrics: in-app analytics. Mental health app companies frequently conduct studies based on the analysis of data gathered from the apps. These data comprise survey results and observation of user behaviours. Surveys can either be conducted by inviting users to act as study participants or by enrolling users without their immediate consent (consent may be taken as agreed on through users' implied consent, through use of the app). User data comprises a diverse set of possible data-vectors, from times and frequency of usage, preferred methods and features to identification of keywords, demographic information, and even potentially geographic information: the locations in which users interact (or fail to interact) with the app. Essentially, any data which can be operationalised will potentially be used for research because the app-makers have an interest in showing how their apps are potentially applicable to a wide range of user-

[64] American College of Obstetricians and Gynaecologists
[65] Ramachandran, M. Suharwardy, S. Leonard, S.A. Gunaseelan, A. Robinson, A. Darcy, A. Lyell, D.J. & Judy, A. (2020) 'Acceptability of postnatal mood management through a smartphone-based automated conversational agent'. *American Journal of Obstetrics & Gynecology*, Volume 222 Issue 1. p.S62
[66] Ibid. p.123
[67] Champion L, Economides M, Chandler C. (2018) 'The efficacy of a brief app-based mindfulness intervention on psychosocial outcomes in healthy adults: A pilot randomised controlled trial'. *PLoSONE,* 13(12):e0209482. p.14
[68] Ibid. p.14

experiences. App makers have access to a huge range of data, and within this range, large sample sizes, with which they can conduct analyses. Youper conducted a study which involved 4517 participants:

> We examined data from paying Youper users (N=4517) who allowed their data to be used for research. To characterize the acceptability of Youper, we asked users to rate the app on a 5-star scale and measured retention statistics for users' first 4 weeks of subscription. To examine effectiveness, we examined longitudinal measures of anxiety and depression symptoms. To test the cumulative regulation hypothesis, we used the proportion of successful emotion regulation attempts to predict symptom reduction.[69]

Youper's study measured users' self-reported satisfaction ratings, retention (duration and frequency of use) and GAD-7 & PHQ-9 measures. A table titled "Additional demographic and clinical characteristics"[70] shows data that Youper had access to: user occupations, their phone's operating system, self-reported mental health diagnosis and self-reported treatment type. Youper's conclusion reads similar to a promotional blurb:

> Youper is a low-cost, completely self-guided treatment that is accessible to users who may not otherwise access mental health care. Our findings demonstrate the acceptability and effectiveness of Youper as a treatment for anxiety and depression symptoms and support continued study of Youper in a randomized clinical trial.[71]

For randomised controlled trials of mental health interventions to be feasible, 'mental health' must be considered in terms of a set of common denominators. This usually involves categorising data in terms of variables: mathematical objects with which data can be objectively measured. This ostensibly allows for research to assume a robust and scientific quality in terms of scalability, reproducibility and consistency of results. CBT is one of the few therapeutic methods which can be reduced to these factors because of its highly proceduralised technique.[72] This will be examined in chapters three and five ('History' and 'Digitisation'). The broad concept of 'mental health' must correspondingly be reduced to a set of discrete and consistent variables, usually represented through user-rating or outcome surveys. Almost all of the above research was conducted to judge the effectiveness of the apps and to compare them to other methods such as face-to-face therapy. This means that both computerised and human-delivered treatment must be reduced to the same format. In other words, some kind of objective measurement must be imposed so that a claim such as "The study confirmed that after 2 weeks, those in the Woebot group experienced a significant reduction in depression, thus our hypothesis was partially supported"[73] and

---

[69] Mehta, A., Niles, A. N., Vargas, J. H., Marafon, T., Couto, D. D., & Gross, J. J. (2021) Acceptability and Effectiveness of Artificial Intelligence Therapy for Anxiety and Depression (Youper): Longitudinal Observational Study'. *Journal of medical Internet research*, *23*(6), e26771

[70] Ibid. p.3

[71] Ibid. p.1

[72] Gipps, R.G.T. (2013) 'Cognitive Behaviour Therapy: A Philosophical Appraisal'. In: Fulford, K.W.M. et al. (eds.) *Oxford Handbook of Philosophy and Psychiatry*. UK: Oxford University press. pp.1245-1263. p.1247

[73] Fitzpatrick, K.K. Darcy, A. & Vierhile, M. (2017) '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*(2), e19. p.8

"...those in the Woebot group significantly reduced their symptoms of depression over the study period as measured by the PHQ-9."[74] can be made. The dubious scientific credentials of not only the measures used in these studies but also of the manner in which they are used means that assessment of these studies must be taken on terms external to those measures. In other words, an assessment of why these studies are conducted using these types of measures is needed.

The generally agreed upon minimum sample size for generating meaningful results is 100, with any less than this requiring more individualised polling measures. Mental health apps, depending on their popularity, have potential access to a vastly larger participant-pool than 100. Youper conducted a study of their app which included 157,213 participants,[75] assessing their emotional responses to the COVID-19 pandemic. Youper managed to gather such a large cohort by including an opt-out rather than opt-in criterion.[76] At this scale, in-app analytics becomes the preferred mode of assessment, even the only mode of assessment: qualitative analysis involving individual polling or interviews would be extremely time consuming, and often impossible due to app companies needing to quickly produce research papers (discussed below). Durational problems are not only due to external pressures but also to internal app-changes over time. Mehta et al. note that their recruitment was confined to a cohort of users who were active on the app between the months of March 4th and July 10th, 2020. This is "Youper was relatively stable during this period (i.e. no significant updates or changes to the intervention were deployed during this time)".[77] The mercurial stability of these apps confounds the possibility of the kind of objective analysis sought by app-maker researchers: a problem reflected by the myriad evaluation methods that companies and external researchers use to assess therapy apps. What is scientific research if it cannot be replicated? The research conducted and published on mental health chatbots can only be applicable to those iterations of chatbots which existed throughout the precise duration of each research project. This is because the apps are prone to frequent and sometimes substantial modification: it is impossible to be certain to what extent a research paper is applicable to that which it is researching after it has been published. The scientific credentials that mental health treatment software companies seek through the production of research is paradoxically jeopardised by the very features of these apps that the companies promote. With this in mind, research must be explicit about their time-based limitations, and must consider the effects that these limitations might have on data-collection and analysis. Recall both Vaidyam et al.[78] and Jabir et al.[79] who draw our attention to the heterogeneity of research methods, styles, and often small sample sizes, and unique measurement

[74] Ibid. p.8
[75] Yarrington, J.S. Lasser, J. Garcia, D. Vargas, J.H. Couto, D.D. Marafon, T. Craske, M.G. & Niles, A.N. (2021) 'Impact of the COVID-19 Pandemic on Mental Health among 157,213 Americans'. *Journal of Affective Disorders*, 1;286:64-70
[76] Ibid. p.65
[77] Mehta, A., Niles, A.N. Vargas, J.H. Marafon, T. Couto, D.D. & Gross, J.J. (2021) 'Acceptability and Effectiveness of Artificial Intelligence Therapy for Anxiety and Depression (Youper): Longitudinal Observational Study'. *Journal of medical Internet research*, 23(6), e26771
[78] Vaidyam, A.N. Wisniewski, H. Halamka, J.D. Kashavan, M.S. & Torous, J.B. (2019) 'Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape'. *Canadian journal of psychiatry*. 64(7), pp.456-464
[79] Jabir, A.I. Martinengo, L. Lin, X. Torous, J. Subramaniam, M. & Tudor Car, L. (2023) 'Evaluating Conversational Agents for Mental Health: Scoping Review of Outcomes and Outcome Measurement Instruments'. *Journal of medical Internet research*, *25*

instruments used when evaluating mental health chatbots. Not only is it difficult to compare various therapy apps, due to the unique measurements used, it is also difficult to compare therapy apps *with themselves* over prolonged time periods. This brings up the question of why app-makers publish so much research without critical assessment of their methods. One such reason is the promotional value of these studies; while this will be analysed in depth in chapter nine ('Technocracy'), below is a short discussion on the promotional aspects of research done on mental health chatbots.

Promotional Claims

Much of the research currently published into the effectiveness of mental health chatbots has been done by firms who are themselves developing these chatbots. While research is clearly needed not just into the effectiveness of this technology but also their impact on the broader fields of mental health treatment, what happens when these companies are conducting their own research into their own products? There is a publicity element to this research: every developer wants to prove that their own technology is not just as effective or superior to traditional treatment, but also that their own product is superior to their competitors' rival products. The line between self-promotion and scientific research is heavily blurred in these instances, and while each company conducting their own research can claim to strictly follow procedures such as randomised controlled trials, the context within which these procedures are undertaken cannot be safely ignored. X2AI have produced a number of mental health chatbots, and also produce research with the aim of establishing their bots in the field of automated treatment. X2AI conducted a technical report on their own app, Tess. It is worth quoting their abstract, which reads like a promotional blurb:

> This technical report highlights how one mental health chatbot, or psychological artificial intelligence service named Tess, has been customized to deliver on-demand support for caregiving professionals, patients, and family caregivers at a non-profit organization. This low-cost, user friendly, and highly customizable service allows emotional support to be scaled to thousands of people at a single time.[80]

The report concludes that:

> There is evidence that using psychological artificial intelligence to provide customized support for caregiving professionals, patients, and family caregiver is a feasible service delivery method. This report suggests that the Tess service may offer an affordable and scalable solution that accommodates the busy schedules of caregivers while helping them reduce burnout and improve resilience. Furthermore, Tess' capacity to expand support to patients further reduces the caregiver burden and has the potential to relieve feelings of depression, anxiety, and loneliness.[81]

The wording of these reports is notable in that they read more like promotional material than scientific analysis. X2AI conducted a study which involved providing a mental health chatbot

[80] Joerin, A. Rauws, M. & Ackerman, M.L. (2019) 'Psychological Artificial Intelligence Service, Tess: Delivering On-demand Support to Patients and Their Caregivers: Technical Report'. *Cureus*, 11(1), e3972. p.1
[81] Ibid. p.5

to a cohort of Argentinian students. The objective of this study is to make a preliminary evaluation as to whether a mental health chatbot might be viable for "examining symptoms of depression and anxiety in university students". X2AI states in the 'background' section of this study "Artificial intelligence-based chatbots are emerging as instruments of psychological intervention; however, no relevant studies have been reported in Latin America".[82] X2AI began by designing chatbots designed for intervention in various non-US regions as purported humanitarian assistance.[83] X2AI's mental health pivot is similar to Wysa, who originally designed their app as an eating disorder assistance app. The malleability and iterative nature of the software will be discussed below as a major problem for the production of statistical research. The study concludes that students spoke to the chatbot often, and that "positive feedback was associated with a higher number of messages exchanged," and that further research is needed to ascertain viability. 'Tess' has rebranded as 'Cass', and X2AI has diligently produced a research paper to mark this transition.[84] The paper is titled "Effectiveness of a chatbot for eating disorders prevention: A randomized clinical trial",[85] and it evaluates a version of Tess which has been modified to include an eight-week eating disorder program. Interestingly, while the research is displayed on X2AI's website to promote 'Cass', the bot involved in the paper is actually Tess, or as per the paper itself, 'Tessa'.

Mental health chatbot makers are keen to produce studies which show app-effectiveness for specific mental health circumstances. Woebot, along with other mental health chatbot companies produces research which assesses a version of their bot in the treatment of various illnesses, not confined specifically to 'mental health', but which affect the mental health of the app users. The treatment for these specific circumstances involves providing the app or a variant of the app, in this case 'W-SUDs': meaning Woebot-Substance Use Disorder. "This study aims to adapt Woebot for the treatment of substance use disorders (W-SUDs) and examine its feasibility, acceptability, and preliminary efficacy".[86] The study claims that "Automated conversational agents can deliver a coach-like or sponsor-like experience and yet do not require human implementation assistance for in-the-moment treatment delivery".[87] The study cites Vaidyam et al.[88] in claiming that chatbots may help to decrease treatment attrition compared to non-chatbot computerised alternatives. In this

[82] Klos, M.C. Escoredo, M. Joerin, A. Lemos, V.N. Rauws, M. & Bunge, E.L. (2021) 'Artificial Intelligence-Based Chatbot for Anxiety and Depression in University Students: Pilot Randomized Controlled Trial'. *JMIR formative research*, *5*(8), e20678.
[83] Solon, O. (2016) 'Karim the AI delivers psychological support to Syrian refugees'. *The Guardian.* Online: https://www.theguardian.com/technology/2016/mar/22/karim-the-ai-delivers-psychological-support-to-syrian-refugees (Last accessed 28/05/22)
[84] 'Randomized Controlled Trial highlights effectiveness of Tess for eating disorders prevention'. Online: https://www.cass.ai/impact (Last accessed 10/01/24)
[85] Fitzsimmons-Craft, E.E. Chan, W.W. Smith, A.C. Firebaugh, M.-L. Fowler, L.A. Topooco, N. DePietro, B. Wilfley, D.E. Taylor, C.B. & Jacobson, N.C. (2022) 'Effectiveness of a chatbot for eating disorders prevention: A randomized clinical trial'. *International Journal of Eating Disorders*, 55(3) pp.343–353
[86] Prochaska, J. Vogel, E. Chieng, A. Kendra, M. Baiocchi, M. Pajarito & S. Robinson, A. (2021) 'A Therapeutic Relational Agent for Reducing Problematic Substance Use (Woebot): Development and Usability Study'. *Journal of Medical Internet Research,* 2021;23(3):e24850. p.1
[87] Ibid. p.2
[88] Vaidyam, A.N. Wisniewski, H. Halamka, J.D. Kashavan, M.S. & Torous, J.B. (2019) 'Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape'. *Canadian journal of psychiatry*, 64(7) pp.456–464

sense, this study can also be considered as an exploratory analysis in that Woebot's aim is to establish the chatbot component of computerised mental health intervention as a significant development of automated treatment. Participant inclusion involved recruiting through the Woebot app, social media and physical flyers under the criteria of screening positive for drug or alcohol abuse. The study measures reduction in drug or alcohol use and general satisfaction with the app. Exclusion criteria for the study includes severe drug dependency, meaning that the intervention is not intended for those suffering from severe disorders.

This reservation mirrors claims made by app makers to the effect that their bots are not 'real' therapists and as such should not be used as alternatives, which is then contradicted by promotional claims. Alison Darcy, Woebot's director, was interviewed by Chatbots Magazine in 2018. Of immediate note is Darcy's claim that "we do not provide therapy",[89] which is a common claim of therapy chatbot makers, this claim is in direct contradiction to another claim - that therapy apps provide CBT.[90] How does this claim to not provide therapy align with the common claim that these chatbots provide CBT? Is it because CBT is an aspiration that has not currently been reached? Or maybe that they do not really believe that CBT is a 'real' therapy? Or perhaps there is no real aspiration or belief involved, and it is simply a necessary claim to indemnify the makers of these chatbots against possible harm coming to users who need more critical care? Almost all the therapy chatbot developers make sure their apps shouldn't be confused with 'real' therapy, but then claim that the chatbots do offer genuine CBT, and that computerised CBT is as effective as its real-life counterpart, which seems a glaring paradox. Darcy makes a claim that is also often stated by therapy chatbot developers - that they are attempting to democratise therapy by making it available to everyone (with a smartphone). "We want to bring really good psychological tools to the masses.[91] This is a standard claim of tech start-up developers - that through technological mass-production, products or experiences can be provided to those who would otherwise not have access. Woebot's W-SUD study concluded that:

> W-SUDs was feasible to deliver, engaging, and acceptable and was associated with significant improvements in substance use, confidence, cravings, depression, and anxiety. Study attrition was high. Future research will evaluate W-SUDs in a randomized controlled trial with a more diverse sample and with the use of greater study retention strategies.

Similar to Woebot, the makers of Wysa cite "a major shortage of mental health professionals, long waiting lists for treatment, and stigma"[92] as the grounds for developing an automated alternative to face-to-face treatment. The study, conducted by Wysa, describes

[89] Rao, A. (2018) 'Woebot— Your AI Cognitive Behavioral Therapist: An Interview with Alison Darcy'. *Chatbots Magazine.* Online: https://chatbotsmagazine.com/woebot-your-ai-cognitive-behavioral-therapist-an-interview-with-alison-darcy-b69ac238af45 (Last accessed 21/05/22)
[90] "Cognitive Behavioral Therapy (CBT), Interpersonal Psychotherapy (IPT), and Dialectical Behavioral Therapy (DBT) provide the foundation for Woebot's therapeutic support" Online: https://woebothealth.com/what-powers-woebot (Last accessed 20/12/23)
[91] Ibid. p.1
[92] Inkster, B. Sarda, S. & Subramanian, V. (2018) An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study. *JMIR Mhealth Uhealth,* 2018;6(11): e12106

the Wysa app as an “AI-based *emotionally intelligent* mobile chatbot app aimed at building mental resilience and promoting mental well-being using a text-based conversational interface".[93] The study analyses usage of the app by 129 users, who were enrolled[94] via the criteria that they undertook two PHQ-9 assessments over the course of a two-week period. The group was split into two subgroups: “High users” and “low users”, with low users being those who only engaged with the app twice, and high users being those who engaged with the app more than twice during the screening period. The study measured any differences from one score to the next. A qualitative analysis also features in this study, in which user responses to the bot’s prompts were analysed. Wysa divided these responses under the headings “Favorable Experience” and “Less Favorable Experience”, with these then subdivided into the themes “Helpful” and “Encourage”, and “Unhelpful” and “Concerns”.[95] Wysa stated that “Favorable experience was the dominant theme from the user responses. Almost all of the favorable experiences were attributed to the helpfulness of the app in users actually feeling better after their conversation sessions and also after their use of app-provided mindfulness and physical activity techniques". The study acknowledges the limitations involved in enrolling anonymous participants, a non-randomised controlled study environment, no prior health-screening for participants, small groups sizes, inability in accounting for demographic variables, and a lack of detailed feedback from participants. The study concluded that “Our study identified a significantly higher average improvement in symptoms of major depression and a higher proportion of positive in-app experiences among high Wysa users compared with low Wysa users. These findings are encouraging and will help in designing future studies with larger samples and more longitudinal data points".[96] We can see that with studies conducted by mental health app companies on their own products, emphasis is often placed on stating why these apps answer an unfulfilled demand.

Branching Out

There is clearly a publicity element to this research, every developer wants to prove that their own technology is not just as effective or superior to traditional treatment, but also that their own product is superior to their competitors’ rival products. The line between self-promotion and ‘genuine’ scientific research is heavily blurred in these instances, and while each company conducting their own research can claim to strictly follow procedures such as randomised controlled trials, the context within which these procedures are undertaken should not be ignored. The outbreak of COVID-19 led some mental health app companies to produce research which responded to the ensuing mental health crisis. Youper, along with other mental health apps, responded to the COVID-19 pandemic by conducting research on the effectiveness of their apps during the lockdown periods. This study “examined emotions and symptoms before (pre), during (acute), and after (sustained) COVID-related stay-at-

[93] Ibid. p.3
[94] As opposed to recruited: this means that users were not actively recruited or aware of their involvement in the study.
[95] Inkster, B. Sarda, S. & Subramanian, V. (2018) An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study. *JMIR Mhealth Uhealth,* 2018;6(11):e12106
[96] Ibid. p.10

home orders".[97] This paper represents a deviation from the trend sketched above and does not purport to gauge the effectiveness of the app but instead gauges general mental health trends during part of the lockdown period of the Covid-19 pandemic. Wysa conducted a similar study in which general mental health trends were identified through analysis of data which was generated from their app:

> This study used a retrospective observational design. During the COVID-19 pandemic, the app's installations and emotional utterances were measured from March 2020 to October 2021 for the United Kingdom, the United States of America, and India and were mapped against COVID-19 case numbers and their peaks. The engagement of the users from this period (N=4541) with the Wysa app was compared to that of equivalent samples of users from a pre–COVID-19 period (1000 iterations).[98]

These kinds of studies can exploit large participant numbers due to low involvement criteria, sometimes even simply that they download the app with little to no interaction. These generalised studies are produced to show that the apps are not only useful for mental health but that their data gathering potential can be exploited in novel ways. Mental health app-makers research and promote the non-mental benefits of their apps, sometimes using the proviso that improved physical health has mental health benefits. Wysa is in the business of identifying possible users of the app in treating non-mental health issues. A study by Wysa aims to judge the feasibility of using the chatbot in dealing with chronic pain. The study uses a variant of the app: Wysa for Chronic Pain app. The study claims that "To the best of our knowledge, this is the first such study for chronic pain using a fully-automated, free-text–based conversational agent".[99] As this is a 'protocol for a prospective pilot study', it does not claim to offer any substantive conclusions, but rather aims to probe whether the app could be useful in the future. The study's recruitment strategy seems to involve enlisting from online chronic pain mutual support communities: "Participants with self-reported chronic pain (n=500) will be recruited online on a rolling basis from April 2022 through posts on US-based internet communities within this prospective cohort".[100] It is important to note, that while this paper is a proposal for a future study, the bulk of the written material involves claims about Wysa's effectiveness in treating chronic pain:

> Wysa for Chronic Pain overcomes these shortcomings. Moreover, apart from using a conversational flow tailored for chronic pain, it also provides participants with a wide array of self-care tools they can use to deal with other issues like insomnia, depression, anxiety, and negative thoughts anytime they want. Wysa for Chronic Pain has proven high engagement and efficacy when the intervention uses a

---

[97] Yarrington, J.S. Lasser, J. Garcia, D. Vargas, J.H. Couto, D.D. Marafon, T. Craske, M.G. & Niles, A.N. (2021) 'Impact of the COVID-19 Pandemic on Mental Health among 157,213 Americans'. *Journal of Affective Disorders*, 1;286:64-70

[98] Sinha, C. Meheli, S. & Kadaba, M. (2023) 'Understanding Digital Mental Health Needs and Usage With an Artificial Intelligence–Led Mental Health App (Wysa) During the COVID-19 Pandemic: Retrospective Analysis'. *JMIR Formative Research,* 2023;7:e41913

[99] Gupta, M. Malik, T. & Sinha, C. (2022) Delivery of a Mental Health Intervention for Chronic Pain Through an Artificial Intelligence-Enabled App (Wysa): Protocol for a Prospective Pilot Study'. *JMIR research protocols*, *11*(3), e36910. p.1

[100] Ibid. p.1

conversational agent enhanced by a human coach.[101]

Wysa's quest to establish their chatbot as a chronic pain treatment continues with another study, the aim of which is "To evaluate user retention and engagement with an artificial intelligence–led digital mental health app (Wysa for Chronic Pain) that is customized for individuals managing mental health symptoms and coexisting chronic pain".[102] This study recruited 51 adults who presented to a tertiary care centre for chronic musculoskeletal pain. The study measured how often and for how long users engaged with the app. Wysa concluded that users of the app were more likely to continue engagement with the app than engagement supplied by "standard industry metrics". The metrics that are referred to here are the results of a paper which provides a "cross-study evaluation of 100,000 participants".[103] Soon after these studies were completed, Wysa received FDA (U.S. Federal Drug and Food Administration) approval for use of the app to treat chronic pain.[104]

2.4 Conclusion

All scientific research depends on some kind of epistemological framework from which the research can be directed, and a set of standards in order to replicate individual studies and to compare similar studies. Much of the research has been done by the makers of their own apps, who set their own conditions and standards. Of course, these conditions and standards conform to objectivity criteria, such as double-blind conditions, anonymity retention, responsible data-management, etc, so the app-makers can lay claim to adhering to scientific and clinical guidelines. These studies conform to scientific rigour in an internal sense: the various methods used are not in question, but rather, in an external sense, the contextual factors which drive this kind of research are left unexamined. In other words, the data-gathering and measurement might be (although sometimes tenuously) reliable, the validity is questionable. Measurement of mental health and its treatment can never be an exact science, but statistical analysis, by which mental health chatbots are evaluated, relies on precision. Precision, in turn, relies on abstraction: removing any elements of the gathered data which cannot be measured in terms of isolated variables. The more that data is refined in order to be more accurately measured, the less it relates to such subjective problems with which 'mental health' is associated. 'Abstraction' is a defining feature of ReMind's intervention, which I define as either conceptual or practical disarticulation or deconxtualisation of otherwise linked elements. Champion et al.'s study of Headspace[105] comes closest to addressing this problem, which aggregates user satisfaction in a more generalised sense beyond strictly mental health; however, it suffers from the necessity of

[101] Ibid. p.7
[102] Sinha, C. Cheng, A.L. & Kadaba, M. (2022) 'Adherence and Engagement With a Cognitive Behavioral Therapy–Based Conversational Agent (Wysa for Chronic Pain) Among Adults With Chronic Pain: Survival Analysis'. *JMIR Form Res*, 23;6(5):e37302
[103] Pratap, A. et al (2020) 'Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants'. *NPJ digital medicine*, 3, 21
[104] Baldry, S. (2022) 'Wysa Receives FDA Breakthrough Device Designation for AI-led Mental Health Conversational Agent'. *Business Wire.* Online: https://www.businesswire.com/news/home/20220512005084/en/Wysa-Receives-FDA-Breakthrough-Device-Designation-for-AI-led-Mental-Health-Conversational-Agent (Last accessed 28/07/23)
[105] Champion, L. Economides, M. & Chandler, C. (2018) 'The efficacy of a brief app-based mindfulness intervention on psychosocial outcomes in healthy adults: A pilot randomised controlled trial'. *PLoSONE,* 13(12):e0209482

reduction to observable results as the other studies discussed above. In order to be able to accurately measure outcomes, the framework within which mental health is considered must undergo a process of standardisation and reduction to variables. Gauging the type and severity of mental distress, and the method of qualifying who can make this kind of judgement, all undergo a process of 'flattening' in order for the effective comparison of different techniques. This reduction is of course necessary - accurate measurement would be impossible without it, and without accurate measurement it would be impossible to judge which techniques are more effective. What does mental health look like when it is reduced to a set of outcome variables? Can the subjective experience of a mental illness like depression be accurately represented by such measures as the Beck Depression Inventory? If not, then a case can be made that mental health occupies a different conceptual space when considered through either the perspective of outcome surveys or of the voices of those who are suffering. While there is a separate ethical argument to be made over which representation of mental health or illness might be preferable, it is still vital to understand how these representations might differ in order to better make an ethical argument.

Meta-analyses on therapy chatbots have been conducted which have attempted to draw together different research papers in order to present an overview of relevant literature. These attempts have been afflicted by discrepancies between papers which use similar but not identical measures, and even use entirely unique measures making standardisation difficult if not impossible. There is a paradox here in that mental health apps depend on standard definitions for their own outcome studies, such as PHQ-9, GAD-7, PANAS, etc, but standardisation across the board is rare. Proving the objectivity of one's own research becomes questionable when inter-study standards cannot be established. Mental health apps are vigorously promoted through various channels and their makers depend on the claims made in research papers to demonstrate their apps' effectiveness. The question of how possible it is to make accurate assessments that persevere across iterative changes is not broached in these research papers. It might be reasonably assumed that the iterative changes are made in response to the research in order to improve the effectiveness of the apps, but no research has been conducted which takes this into account. The apps are approached 'as is' and so, while many studies assess the effectiveness of their interventions over time in longitudinal assessments, we have no knowledge of the effects of durational change on the apps themselves. With all of this in mind, the present undertaking seeks to establish an analysis which seeks to identify the epistemological framework within which chatbot-based mental health treatment appears viable. Understanding the epistemological framework in which these chatbots are designed is important because this form of treatment is, through its combined methods and delivery, novel, and as such requires a broader understanding of social, historical, technical and economic factors than can be gleaned from the studies discussed above.

Chapter Three: History

Technology catalyzes changes not only in what we do but in how we think. It changes people's awareness of themselves, of one another, of their relationship with the world. The new machine that stands behind the Hashing digital signal, unlike the clock, the telescope, or the train, is a machine that "thinks." It challenges our notions not only of time and distance, but of mind.[106]

3.1 Introduction

This chapter serves not just to introduce the historical foundations of computerised therapy, but also to discuss conceptual lineages upon which it depends. This means that along with the developments in behavioural and cognitive therapy and Mindfulness, electronic computation, cognitive science and artificial intelligence will be discussed. This is to show not only the practical steps such as the invention of conversational software programs, but also the epistemological steps such as the need to approach language in terms of 'non-meaningful' information in order to solve the engineering problems which presented themselves over the course of the development of conversational machines and subsequently, therapeutic chatbots. The chapter begins with the invention of behavioural psychology, with attention paid to how the human subject came to be conceptualised as a sort of machine, receiving environmental 'inputs' and responding with various behaviours. The development of computers will then be considered, beginning with Alan Turing's thought experiment which inaugurated, first mechanical, and then electronic computation. This section focuses on how computers 'functionalise' mathematics in particular and linguistic operations in general. This means that, with computers, language assumes an objective form in which semantics, or meaning, is strictly separated from syntax, or function. This separation underscores a concept of the mind as a computer - an information processing machine in which thought is akin to software and the brain is akin to hardware. While the term 'cybernetics' is now used to refer to an experimental and speculative research program which blossomed and declined between the 1940s and 1970s, the underlying questions which initiated the program have become subsumed into various other strands of research. This research promised experimental proof-of-concepts which, while perhaps not offering definitive answers, initiated unprecedented technological advancement in computer science, psychology, philosophy and evolutionary theories. Following this, the development of cognitive therapy will be discussed, the aim of which was in large part to create an 'anti-psychoanalytic' treatment method. This section will look at how CBT came to refer to a modular form of therapy which encompasses a range of different treatments. In drawing out the modularity of the treatment it will be possible to understand how CBT lends itself to computerised, and thus automated, treatment. This chapter seeks to trace a number of logics which have developed and expanded over the course of the 19th and 20th century, and have provided the foundations for the introduction of computerised mental health treatment. The aim of this chapter is to reconstruct the therapeutic and technical 'conditions of possibility' from which mental health chatbots have emerged.

[106] Turkle, S. (1984) *The second self: computers and the human spirit*. USA: MIT Press. p.12-13

Behavioural Psychology

Ivan Pavlov's behavioural experiments showed that 'involuntary reflex actions' could be stimulated by developing an association between a stimulus and an environmental reference to that stimulus. The famous example of a dog salivating upon hearing a bell which rings every time food is delivered represents the basic behavioural assertion: that prolonged exposure to a stimulus 'conditions' the subject. In other words, the dog's response is based on a physiological reaction which is instilled through prolonged exposure; the expectation of food stems from an automatic action. This unintentional, physiological action which is determined by the observational output of salivation forms the methodological basis of behaviourism, and subsequently, behavioural psychology. John Watson is generally seen as the founder of human behavioural psychology, and took a strict non-speculative position on observational data. Watson's 1913 paper 'Psychology as the Behaviorist Views It' took a radically empiricist approach to psychology in which there could be no ascription of 'thought' onto the human subject:

> Psychology as the behaviorist views it is a purely objective experimental branch of natural science. Its theoretical goal is the prediction and control of behavior. Introspection forms no essential part of its methods, nor is the scientific value of its data dependent upon the readiness with which they lend themselves to interpretation in terms of consciousness. The behaviorist, in his efforts to get a unitary scheme of animal response, recognizes no dividing line between man and brute. The behavior of man, with all of its refinement and complexity, forms only a part of the behaviorist's total scheme of investigation.[107]

The theoretical basis of this type of experimentation was that the human subject could be approached as fundamentally susceptible to suggestion, and that this operated in terms of stimulus and response. The human subject, at a basic level is a 'blank slate', upon which are written external rules and procedures. Watson posited that the gathering of behavioural data could not be permitted to include speculation about underlying psychological mechanisms, due to the goal of "prediction and control of behavior". This attitude is unsurprising as it aligned with the scientific aspirations of the field: in order to qualify as 'scientific', internal psychic mechanisms (such as the unconscious) could not be considered due to being unobservable and unmeasurable. Burrhus Skinner contributed to the scientificity of behavioural psychology with "*radical behaviorism",*[108] in which research was restricted to observable and measurable phenomena. Skinner proposed that his theory of 'operant conditioning' could explain a large range of human behaviour, up to the acquisition of language.[109] Thoma et al. note that Skinner's theories are still operational today in token economies in inpatient units and in such behavioural interventions with children as 'time-outs'.[110]

---

[107] Watson, J. (1913) 'Psychology as the Behaviorist Views It'*. Psychological Review*, 20. pp.158-177. p.158

[108] Schneider, S.M. & Morris, E.K. (1987) 'A History of the Term Radical Behaviorism: From Watson to Skinner'. *The Behavior Analyst*, 10(1) p.36

[109] Skinner, B.F. (1957) *Verbal behaviour*. Appleton-Century-Crofts

[110] Thoma, N., Pilecki, B., & McKay, D. (2015) 'Contemporary Cognitive Behavior Therapy: A Review of Theory, History, and Evidence.*' Psychodynamic Psychiatry,* 43(3) Pp.423-461. p.426

Induction and Adaptation

Behaviourism asserts an extreme form of the inductive scientific method, not just in its form of experimentation, but in the way that it approaches the operations of the human mind. While researchers following Watson may have diverged from his radical view of the 'blank slate', their methods preserved a sensibility that the mind, and the human subject, is machinic; receiving external stimuli in order to produce responses. While behavioural experimentation involving animals and children characterises early research into the precursors of CBT, the 'stimulus-response' approach to human psychology came to be seen as simplistic, causally ambiguous and ethically dubious. However, vestiges of an 'experimental approach' can be seen in contemporary treatment in inverted form: the patient is encouraged to assume an experimental attitude and to conduct 'behavioural experiments' as in the case of exposure treatment. This is coined as "habituation"; habituation means that the *"original reaction towards the stimulus diminishes in intensity or even disappears."*[111] Joseph Wolpe, who originally conducted behavioural experiments on animals, introduced 'reciprocal inhibition'[112] to behavioural therapy: by encouraging patients to induce feelings which conflicted with sensations of fear, anger, sadness, etc, those sensations could be reduced. This later became refined, with the elimination of conflicting feelings, into exposure therapy. Exposure therapy represents the contemporary direction of behavioural therapy in that it introduces 'evocation': thinking about situations which provoke phobias and conditioning one's response over time. By introducing evocation to conditioning or habituation theories of behavioural treatment, we can see how the development of cognitive psychology follows. 'Cognition' can be thought of as a physiological phenomenon, and one which operates through a process of induction, and can be conditioned or 'trained' as a therapeutic method. The behavioural approach to human psychology hinges on the assertion that the human subject is essentially adaptable, and that adaptation to one's environment is an automatic process.

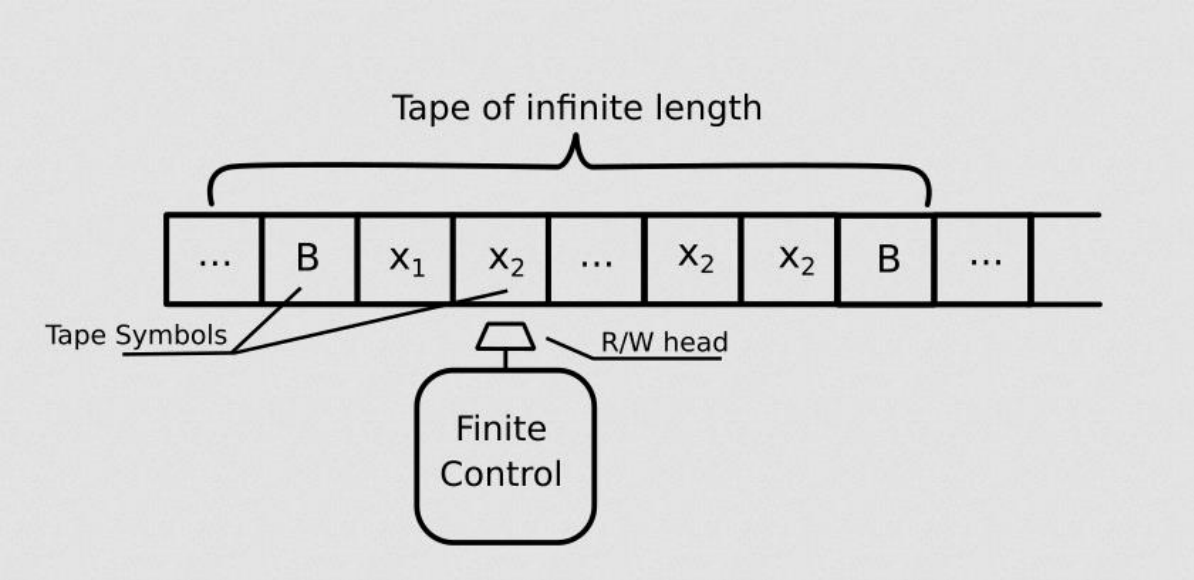
Behaviourism introduced a theory and a method to psychology which traces a conceptual line through to contemporary CBT. The behavioural theory of adaptation posits the human subject's lack of agency in determining their behaviour: 'involuntary reflex actions' can occur whether an external environmental or internal psychological stimulus provokes them. This reaction is not intentional in a strict sense: the subject 'learns' through repeated action and response to favour certain responses. We can see a parallel between 'conditioning' theory and much of contemporary psychology through a shared evolutionary assumption. It is assumed that physiological evolutionary factors are the essential basis for the conditioned responses (the example for a fear response is often of the 'caveman responding to a dangerous animal'). A case will not be made for the veracity of this line of reasoning; however, what will be discussed in the final section is a dependence in CBT theory on physiological conditioning to explain a whole host of human behaviour. This is the essential 'behavioural' basis for contemporary CBT. We can also see a parallel between CBT's theory of learning and 'reinforcement learning' found in artificial intelligence, which is based on

[111] Eelen, P. & Vervliet, B. (2006) 'Fear Conditioning and Clinical Implications: What Can We Learn From the Past?' In: M. G. Craske, D. Hermans, & D. Vansteenwegen (eds.) *Fear and learning: From basic processes to clinical implications.* American Psychological Association. pp. 17-35
[112] Wolpe, J. (1954) 'Reciprocal inhibition as the main basis of psychotherapeutic effects'. *American Medical Association Archives of Neurology and Psychiatry*, 72. pp.204-226

rewarding desired behaviours and punishing undesired behaviours. The human mind as ‘trainable’ in terms of mental health forms the backdrop for subsequent cognitive psychological treatment development.

3.2 Artificial Cognition

The invention of the Turing machine - a proof of concept description of how a machine could manipulate symbols in order to perform various functions - inaugurated a revolution not just in computer science, but also in philosophy of mind and as we shall see, approaches to the treatment of mental conditions. By demonstrating that the manipulation of symbolic objects could be done by a machine, Turing proved that certain intellectual functions which were previously conceived as the sole domain of the human mind could be formalised and automated as a machine process. Essentially, the Turing machine is a description of how a computer program interacts with a central processing unit - the two fundamental features of all electronic computers.



The Turing machine is a model of an automated algorithmic process in which a moveable ‘head’ can read a symbol printed on a tape, this symbol instructs the head to move left or right on the tape and then write a new symbol, depending on the status of the previous symbol and the position of the head in relation to the tape. This basic model represents the most fundamental properties of digital computation and with more complex versions any type of computer function can be constructed. This description of mechanical symbol manipulation was, as Johnston explains, a completely new kind of machine, in which the specific function - symbol manipulation - was not determined by the physical basis on which the function operated:

> Invented as part of the proof, his notion of the Turing machine would eventually provide a formal basis for the modern computer, in which different sets of instructions, or programs - for computation, data processing, sending and receiving data, and so on - allow the same machine to do a variety of tasks. This capacity makes the computer a fundamentally new *type* of machine, defined by the logical and

> functional rather than a material structure. It is an abstract, second-order machine in which the logical form of many different kinds of machines is abstracted and made equivalent to a set of algorithms.[113]

With computer power increasing over time and the advent of the general-purpose computer, and with Turing's concept of a separate processing unit running any computer program which could be fed into it, the ground was set for the revolution in computer science as we know it today. Alan Turing's invention of the 'universal machine' introduced a formalisation of a particular form of intellectual work which, depending on one's understanding of computation, can either be described as 'manipulating symbols, or 'performing mathematics'. Either way, the feat that computers achieve is a proceduralising of intellectual work. Since then electronic computation has transformed almost every industrial, commercial and social practice through the transformational effects of algorithmic proceduralism. The term 'cybernetics' was adapted by the mathematician Norbert Wiener in 1948 who used it to describe the newly emerging science of control, communication and feedback in complex systems. Coming together in the 1940s as a combination of mathematics, computer science, electrical engineering, linguistic philosophy, game theory and evolutionary theory, cybernetics could be described as one of the first genuine interdisciplinary research subjects, combining so many different strands of experimental and speculative research that a requirement for highly generalised and elastic terms of reference soon became necessary. The novelty and interdisciplinarity of cybernetics were due in large part to the novelty of digital computation, but initial research in cybernetics was characterised by physical devices rather than computer programs - robots with sensors that could take visual, heat, motion or proximity cues and then respond to those cues. Responses would become more and more unpredictable when sensors were multiplied but crucially also connected, so that internal state changes could occur which were not directly associated with the environment but due to the complex arrangement of interwoven sensor data. The purpose of these experiments was to observe how complex behaviour, whether in individual robots or in systems, could emerge from a discrete set of inputs and behavioural reactions. As computer power increased, computer simulation would come to dominate the field of cybernetics experiments. Contemporary understanding of the human mind, its connection to the physical materiality of the brain, whether and how to make distinctions between phenomenal experience and neurobiological activity, can be traced to questions that cybernetics introduced and attempted to solve. Initial cybernetics experiments attempted not just to simulate, but to *instantiate* complex phenomena such as evolution, 'thought', and social self-organisation. We can understand early cybernetics experiments as attempts to build demonstrations of theoretical concepts, which did not just functionally exhibit the effects implied by concepts, but which would model the concepts in physical form. We can see the rapid embrace of computer simulation as a turning point in cybernetics history. By the 1960s when digital computers became powerful enough to simulate complex systems, cybernetics quickly dropped physical experimentation in favour of computer models, which could simulate, in digital form, the earlier physical, analogue experiments. This preference for attempting to simulate, as opposed to instantiate, real world phenomena inaugurates a functionalist ethos which links the cybernetic quest to understand complex systems to the concepts which underpin behavioural and cognitive psychology.

[113] Johnston, J. (2008) *The Allure of Machinic Life*. USA: MIT Press. p.71

3.3 Cognitive Science

George Miller, Eugene Galanter, & Karl H. Pribram's *Plans and the Structure of Behavior* (1960) is considered as a foundational text in cognitive science, aligning psychology with information theory, computer science and linguistics. Miller et al. describe cognitive processes in behavioural and algorithmic terms:

> Plan. Any complete description of behavior should be adequate to serve as a set of instructions, that is, it should have the characteristics of a plan that could guide the action described…*A Plan is any hierarchical process in the organism that can control the order in which a sequence of operations is to be performed.*[114] (Italics mine)

The computational basis of cognitive science is implicit rather than explicit: the discovery that intellectual processes could be performed by machines and cybernetics experiments in artificial intelligence led to theories that human cognition was in some ways mechanistic, and subsequently that 'thought' could be modelled or observed. Cognitive science approaches the mind in terms of function and of mechanism, Cratsley and Samuels point out:

> Though cognitive scientists do not share a single vision of how explanation ought to proceed, large regions of the field cleave to a set of familiar assumptions about the sorts of models that ought to be developed. For heuristic purposes, we divide up these assumptions into two related families of commitments. The first concern the idea that cognitive processes and capacities depend on information processing. The second concern the idea that cognitive explanations are in some appropriately broad sense mechanistic.[115]

Cognitive psychology claims a direct lineage to cognitive science, and as such retains the basic premise that the mind is in its functional capacity a computer. It does this by conceptualising the brain as the director of action via manipulation of biophysical information in the form of "...perceiving, learning, remembering, thinking, reasoning, and understanding."[116] The mind is equated with a software program in that it 'tells' the brain what operations to perform. Cognitive psychology, due to its approach and conceptual lineage, rather than any specific theoretical grounding, considers the mind as a computer, and consciousness as the director of the mind. Rafael Núñez points out that cognitive psychology does not cohere around a well-defined theory of mind, but is identified by its experimental approach to the mind, which is guided by computational and information-processing concepts:

> "[C]ognitive" in "cognitive psychology" primarily denotes information-processing psychology, following influential work in the 1960s that saw cognitive science as essentially the marriage between psychology and artificial intelligence in which

[114] Miller, G.A. Galanter, E. & Pribram, K.H. (1960) *Plans and the structure of behavior*. USA: Henry Holt and Co. p.16
[115] Cratsley, K. Samuels, R. (2013) 'Cognitive Science and Explanations of Psychopathology". In: Fulford, K.W.M. et al. (eds.) *Oxford Handbook of Philosophy and Psychiatry.* Uk: Oxford University press. pp.413-1263
[116] Lu, Z.L. & Dosher, B.A. (2007) 'Cognitive Psychology'. *Scholarpedia*, 2(8):2769

> neuroscience and the study of culture played virtually no role. Thus, “cognitive” psychology doesn’t just designate a sub-field of psychology that studies cognition and intelligence. Rather, it usually refers to a specific theoretical approach and research program in psychology.[117]

Cognitive psychology relies on developments in neurobiology and computer science, with both disciplines being intertwined in a complex way. Two foundational concepts derived from cybernetics, self-regulating complex systems and symbol manipulation, were to become the often-oppositional bases for the development of artificial intelligence. The idea that the human mind could be modelled or simulated, and maybe even instantiated electronically, was a significant development for both computer science and psychology. If even specific functions of the human mind such as pattern and symbol recognition could be recreated, this meant not just that the mind as a whole could possibly either be simulated or recreated, but also that on a fundamental level the mind itself functioned in a similar way to a computer. Two theories of artificial intelligence would become powerful influences on psychological assumptions of the mind, one is that the mind is essentially a symbol manipulating machine and the other is that it is a self-organised, emergent system. This natural scientific approach to the question of intelligence can be seen as a contemporary experimental way to approach the metaphysical mind/body split. This endeavour is an attempt to provide a truly objective account of nature in which the human subject has no bearing. Theories of artificial intelligence that take their lineage from the Turing machine claim ‘intelligence’ as being the privileged source of what it means to be human. In 1967 Ulric Neisser coined this era the “cognitive revolution”, which refers to how experimental psychology provides a “systematic application of information-processing theory to perception and thought, covering a variety of topics but concentrating on the visual memory system.”[118] Information-processing theory can be thought of as a foundation, not quite for the therapeutic method, but for the conceptual conditions which helped behavioural therapy to be reconfigured in ‘cognitive’ terms.

3.4 Cognitive Therapy

Beck

Arron T. Beck is generally considered to be the founder of cognitive therapy, and is seen as one of the most influential historical figures in the development of CBT. Beck sought to advance a form of psychotherapy treatment which could be verified through experiment. Psychoanalytic treatment, or in the USA, the ‘ego-psychology’ variant, was the dominant form of psychological treatment in the 1950s. Beck had trained as a psychoanalyst but he wished to establish experimental techniques to observe the results of psychoanalytic treatment. This led to an eventual split from psychoanalysis and the incorporation of behavioural psychological methods of observation and measurement.[119] Daniel B. Smith summarises how Beck’s contribution to the psychotherapeutic field was in opposition to psychoanalysis which was perceived as defeatist:

---

[117] Núñez, R., Allen, M., Gao, R. et al. (2019) ‘What happened to cognitive science?’ *Nature Human Behaviour,* 3. pp.782-791

[118] Neisser, U. (1967) *Cognitive psychology*. USA: Appelton-Century-Croft

[119] Rosner, R.I. (2014) ‘The "Splendid Isolation" of Aaron T. Beck’. *Isis*, 105 (4) pp.734-58

> …whereas psychoanalysis is ultimately pessimistic, seeing disappointment as the price for existence, Beck's approach is upbeat, conveying a sense that, with hard work and determined rationality, one could learn not only to tolerate but to stamp out neurotic tendencies.[120]

According to Beck, neurotic thoughts, framed as 'cognitions', could be reconditioned through cognitive therapy, and eventually eliminated. Under the behaviourist rubric, the term 'cognition' in cognitive therapy is associated not just with 'feelings' but also with 'functions': the human mind acts in functional ways, processing environmental inputs as cognitions, and providing behavioural outputs. Beck's development led to psychotherapeutic work taking on a veneer of technical action. Cognitive therapy is essentially a stoicism-training regime. Beck stated in his first treatment manual for depression that "The philosophical origins of cognitive therapy can be traced back to the Stoic philosophers".[121] A common claim which cognitive treatment espouses, and which determines treatment is "it is not a situation in and of itself that determines what people feel, but rather how they construe a situation".[122] How one perceives, interprets and reacts to a situation which causes them suffering are all determined by the way that their perceptions - 'cognitions' - influence subsequent interpretation and reaction. Two features of behaviourism, conditioning and adaptation, make their way into cognitive therapy as internalisations: the patient conditions their cognitions and adapts them to the environment as opposed to performing this procedure in terms of behaviour alone. Cognitive therapy is essentially a form of skill-transfer, in which the therapist teaches the patient various strategies and techniques, with the eventual aim that the patient develops self-reliance, essentially becoming their own therapist. The combination of cognitive and behavioural therapy has become, if not the dominant form of treatment (practitioners tend to mix and match their methods), then the basis for the dominant epistemological paradigm which informs mental health treatment and research. Diagnosis and treatment can be conducted in a behavioural mode: this means that mental health is conceptualised as located in terms of stimulus and response. In other words, whether it is through 'behavioural' or 'cognitive' means, treatment is associated with behavioural change: adapting to one's environment.

Technical and Modular

Beck's 'Beck Depression Inventory', which he developed after the American Psychoanalytic Institute rejected his membership application, was also a means to commercialisation of a psychiatric method. In order to transform treatment into a marketable product, Beck needed to create a quantifiable metric which could verify the effectiveness of his new treatment method. Psychoanalytic theory is not a quantifiable theory: while involving the learning of complex concepts, training follows an apprenticeship approach which cannot be strictly manualised or quantified. Cognitive therapy and its descendent CBT take the opposite approach; indeed, CBT is one of the few psychotherapeutic disciplines where in most cases experiential training in the form of undergoing therapy is not required in order to qualify as a

[120] Smith, D. B. (2009) 'The doctor is in'. *The American Scholar*. Online: https://theamericanscholar.org/the-doctor-is-in (Last accessed 17/09/23)
[121] Beck, A. Rush, J. Shaw, B. & Emery, G. (1979) *Cognitive Therapy of Depression*. USA: Guildford Press. p.8
[122] Fenn, K. & Byrne, M. (2013) 'The key principles of cognitive behavioural therapy'. *InnovAiT*, 2013;6(9):579-585

practitioner. This is due to an 'externalisation' of the mind in cognitive therapy: cognitions are considered as verifiable objects and treatment takes the form of observing, measuring and manipulating these objects. In taking this approach, cognitive therapy transforms therapy into a technical action of adjustment, which operates through procedural steps. This is what it means for a therapy to be a 'manualised' form of treatment: in seeking quantification, inner experience, or the 'what it is like' to be a human subject, must be approached in terms which render this inner experience as measurable in some way. Behavioural treatment approaches the human animal as a stimulus-response machine in which the circumstances one is exposed to conditions how one reacts. Treatment involves behavioural 'experiments' which usually include some sort of conceptual exposure – imagining fearful situations and learning to become accustomed, or to adapt in some way. When combined with cognitive treatment to form CBT, the role of experimenter is transferred onto the patient, who performs their own 'behavioural experiments' to break out of cycles of distressing thoughts. This means that cognitive treatment involves comparing one's thoughts to 'truth': assessing whether one's feelings and emotions are factually correct and altering them if they are not. The role of experimenter is slightly different in that the patient takes on the task of experimenting with their thoughts rather than with their behaviour, but the scientific attitude remains the same: that of impartial observer/intervener.

Evidence-Based Treatment

In order to achieve scientific status, the mind, or the objects which comprise the mind must be rendered in some way as observable, or potentially observable. 'Thought' is not observable in a visual sense, and cannot be objectified for the observation of others, so Beck needed to find another way to render thought as observable. This is done through a reliance on the production of 'evidence'. CBT is an 'evidence-based treatment'. The British Association for Behavioural & Cognitive Psychotherapies (BABCP) understand evidence to mean the results of research studies:

> When we talk about the evidence base we are mostly referring to research studies which have been carried out and written up in academic journals which are peer-reviewed. This means that the quality of the articles has been checked by other researchers working in similar fields. It is important that research studies have strong, robust designs. This is so that we can be confident that the therapy being tested is very likely to work well for a range of people. It also means we can be confident that any improvement is due to the treatment we are testing, and not something else.[123]

These studies involve randomised controlled trials (RCTs), in which the 'something else' must be controlled for in order to make sure that the study is focused on answering specific questions. Robust scientific research involves eliminating as many 'something elses' as possible, to make increasingly sure that the evidence which is produced is related to what one had intended to observe. For CBT, 'evidence' is not equated with cognitions as such but with the results of RCTs: the data gathered in the course of conducting research and its subsequent analysis. The proliferation of RCTs has allowed CBT to claim the title of "gold

[123] BABCP. 'Cognitive Behavioural Therapy (CBT): What's the Evidence?'. Online: https://babcp.com/What-is-CBT/Cognitive-Behavioural-Therapy-Whats-the-Evidence (Last accessed 12/04/23)

standard" of psychotherapeutic treatments.[124] The standardised form of treatments under the banner of CBT is the basis for its scientific credentials: we can see a process of translation occurring from the CBT's purported manipulation of cognition to its actual manipulation of its own treatment form throughout the process of 'scientisation'. Thoma et al. make the point:

> While RCTs can tell us that a given psychotherapy works better than a control condition, RCTs do not tell us what about the therapy caused the change.[125]

Behavioural approaches to psychology arose due to the demand to scientise the discipline, which signalled a divergence from Freudian theory which posited unobservable, or unverifiable, phenomena such as the unconscious. Thoma et al. point out that the inventors of behavioural therapy assumed a radical form of empiricism in their demand for scientific verifiability:

> …Skinner rejected Thorndike and others' reliance upon unobservable mental states, such as satisfaction and aversion, and called his own approach radical behaviorism, restricting his objects of study to explicitly observable and measurable phenomena.[126]

This scientistic attitude carried over into cognitive therapy, but due to the unobservability of 'cognitions', the demands of empiricism led Beck to forge another route. In his transformation of behavioural therapy, Beck aimed to develop cognitive therapy specifically as a measurable treatment. What this amounted to in practice was to standardise the treatment process so that measurable results could be compared. This initially involved developing a standardised therapy, and in the 1970s Beck 'manualised' treatment in the development of his cognitive therapy of depression (CTOD). According to Rachael Rosner, Beck was at the forefront of attempts to operationalise therapeutic practice in scientifically observable and reproducible ways in which every patient would receive the same treatment "on the same days of a prescribed course of treatment under randomized conditions."[127] Rosner goes on to note that "His new manualized approach was tested at the federal level in the first multi-site RCT of psychotherapy."[128] RCTs depend, crucially, on measurability, transmissibility, and comparability in order to have any legitimacy. Rather than render the inner-workings of the mind as observable to scientific scrutiny, cognitive and behavioural treatment renders the mind, and mental health, in scientific terms: 'validity' and 'utility'. A thought is 'valid' if it can be said to correspond to reality, for example, if a patient were to state "I always say the wrong thing", this would be queried by the clinician as to its truth or falsity. Logically this is a false statement, and so the patient would be encouraged to alter this statement when it emerges in thought, to something less punishing, and more 'rational', or more equated with a truthful reality. Behaviourism's focus on stimulus and response can be seen to have an overarching influence on the psychological basis of CBT due to its

---

[124] David, D. Cristea, I. & Hofmann, S.G. (2018) 'Why Cognitive Behavioral Therapy Is the Current Gold Standard of Psychotherapy.' *Front Psychiatry*, 29;9:4

[125] Thoma, N., Pilecki, B., & McKay, D. (2015) 'Contemporary Cognitive Behavior Therapy: A Review of Theory, History, and Evidence.*' Psychodynamic Psychiatry,* 43(3) pp.423-461. p.450

[126] Ibid. p.425

[127] Rosner, R. (2018) 'Manualizing psychotherapy: Aaron T. Beck and the origins of Cognitive Therapy of Depression.' *European Journal of Psychotherapy & Counselling,* 20 (2018) pp.25-47. p.27

[128] Ibid. p.27

demand for ‘evidence’ in the form of observable, measurable and transmissible data. The introduction of cognitive treatment and later, Mindfulness, followed this requirement, however, due to their theoretical peculiarities (CBT dealing with ‘cognition’, and Mindfulness dealing with ‘the self’), their reliance on evidence have led to a reconfiguration of the definition of ‘evidence’ itself. This reconfiguration will be a theme throughout this thesis.

ELIZA and Beyond

The use of computers to aid in both the diagnosis and treatment of mental illness had been a topic of academic discussion since the early stages of computer development when computer power enabled information storage, retrieval and analysis. Joseph Weizenbaum’s ELIZA was the very first experiment in computerised treatment between 1964 and 1966. ELIZA was a chatbot which simulated Rogerian psychotherapy, in which a user is prompted to respond to questions, and those replies are reflected back to the user in different ways. ELIZA is commonly considered as the first chatbot. Wizenbaum’s aim was to "demonstrate that the communication between man and machine was superficial."[129] As the story goes, Wizenbaum was surprised that ELIZA provided a compelling experience; his secretary apocryphally asked him to leave the room when she was speaking to the bot. Even if the story is exaggerated, we can understand that computerised ‘talking’ therapy was considered as in some way potentially viable, even at such an early stage of computer development.

Kenneth Colby did much to translate the psychotherapeutic dynamic into a linguistic form which could be hosted by a computer - i.e. creating a conversational computer program which mimics therapy. Initially trained as a psychoanalyst, Colby became disillusioned by the unverifiability of psychoanalytic theory and shifted to cognitive and behavioural therapy. Throughout the 1960s, 70s and 80s, Colby became interested in the potential of computer automation and became active in developing conversational computer therapy software. In 1972 Colby created the program PARRY which, instead of simulating a therapist, attempted to simulate the thinking of someone suffering from paranoia. Colby wanted to introduce this software as training material for potential clinicians. Colby’s theory of paranoia was based on the supposition that the paranoid individual’s utterances “are produced by an underlying organized structure of rules and not by a variety of random and unconnected mechanical failures.”[130] Colby’s approach can be seen as in many ways contrary to the developing consensus at the time which understood mental disorders as malfunctions. His computational approach to treatment can be seen in some ways as a nod to the experimental approach characterising the early cyberneticists. By attempting to simulate the mind of the patient, as opposed to the persona of the therapist, we can glimpse a brief moment of the experimental attitude, but Colby was unsuccessful in convincing clinicians or institutions that his invention would assist training, and so PARRY never took off.

A 1978 paper by James Johnson identified three trends in the use of computers in the treatment of mental health, the first trend being the automation of patient data systems, the second trend being the automation of diagnostic techniques, and the third trend being the

[129] Epstein, J. & Klinkenberg, W.D. (2001) ‘From Eliza to Internet: A brief history of computerized assessment’. *Computers in Human Behavior*, 17 (3) pp.295-314
[130] Colby, K. (1975) *Artificial Paranoia: A Computer Simulation of Paranoid Processes*. Elsevier p.99-100

automation of clinical intervention.[131] The paper charts the use of computers in treatment, from acting as a time-saving device to becoming a vital component in treatment. Predictions are made as to the possible future uses of computerised therapy:

> Simulation, game theory, and computer systems will be used to provide therapeutic experiences to the patient while in treatment. Computer programs will be constructed to simulate experiences previously possible only outside the treatment setting.[132]

With the recent development of virtual reality simulations to treat combat soldiers suffering from PTSD,[133] and the automation of patient data systems consisting of a routine transfer of paper to electronic systems, some of the predictions discussed in the paper have come to pass. The automation of diagnostic techniques and clinical intervention are another matter. Predictions made in the paper relate to the use of computer software to aid the clinician in making diagnoses and in providing treatment, such as the use of analytic software to predict illness morbidity and mood cycles, determination of treatment plans, and in making decisions about patient discharge. The shift away from 'depth' therapy towards cognitive and behavioural therapy undertaken over the last fifty years helps to explain how the introduction of computerised automation in terms of diagnosis and/or treatment became feasible more recently. This is due to the transformation of how 'mental illness' is conceptualised through the lens of cognitive science, and operationalised through the cognitive and behavioural treatment forms. While the introduction of computerised or computer-aided mental health treatment does not explicitly depend on the idea that the human mind is akin to a computer, the intertwined historical lineages of digital computation and computerised treatment both depend on a shared assumption: that 'thought' is equated with 'information'.

3.5 Mindfulness

Throughout the 1980s and 90s, cognitive and behavioural therapies were often applied in contrast to each other, to experimentally assess which method would be most appropriate for treating various mental disorders. Over time, due to their shared epistemological basis, these two methods would come to coexist under the banner of Cognitive Behavioural Therapy. Beck stated that cognitive therapy is "the integrative therapy",[134] meaning that its modular form allows other methods to be added. This modularity mirrors the modular concept of the mind that is shared by these methods: cognitive psychology views human subjective experience in terms of various modules such as perception, memory, computation, libido, etc.[135] This mind as a technical object can be treated by a technical therapy. CBT would gradually come to be known as a therapy in its own right but more as an overarching umbrella term for a range of therapies which share a common epistemological

[131] Johnson, J. (1978) 'Computers in Mental Health: Where Are We Now.' *Symposium on Computer Applications in Medical Care.* p.104-108
[132] Ibid. p.106
[133] Blum, D. (2021) 'Virtual Reality Therapy Plunges Patients Back Into Trauma. Here Is Why Some Swear by It.' *New York Times.* Online: https://www.nytimes.com/2021/06/03/well/mind/vr-therapy.html (Last accessed 03/04/2023)
[134] Beck, A.T. (1991) 'Cognitive therapy as the integrative therapy.' *Journal of Psychotherapy Integration*, 1(3) pp.191-198
[135] Fodor, J. (1983) *The Modularity of Mind.* USA: MIT Press

basis.[136] As a technical and modular basis for therapy, CBT can assert itself as a 'platform' upon which a sprawling therapeutic empire can be built. The introduction of meditation techniques under the banner of 'Mindfulness' into CBT is considered the "third wave" of CBT, with behavioural therapy and cognitive therapy being the first and second.[137] Introduced throughout the 1980s with the development of acceptance and commitment therapy (ACT), dialectical behavioural therapy (DBT) and Jon Kabat-Zinn's Mindfulness-based stress reduction program (MBSR), Mindfulness is commonly described as "a type of meditation in which you focus on being intensely aware of what you're sensing and feeling in the moment, without interpretation or judgement."[138] Outside of clinical CBT programs, Mindfulness techniques are taught in schools, universities and workplaces as a generalised and accessible approach to maintaining one's mental health. The generalisability of Mindfulness mirrors cognitive and behavioural treatment in that it is considered as socially neutral: anyone can take up these techniques irrespective of their cultural or religious backgrounds. Mindfulness techniques draw from a secularised form of Buddhist meditation practices in which the patient is encouraged to 'observe' one's thoughts as they occur, but not to intervene during their occurrence. According to ACT, allowing thought to occur without judgement or intervention provides a sense of inconsequentiality, and the power that thought has over the individual is diminished. In this way, ACT, and Mindfulness techniques in general, differ from CBT in that immediate intervention is discouraged. Thoma et al. associated this distinction with an approach to 'thought' in general:

> This is another distinction from traditional CBT; instead of attempting to change the contents of cognitions, ACT focuses more on changing one's relationship to the process of thinking altogether.[139]

'Thought', in terms of Mindfulness, becomes an object, i.e. whereas in CBT, 'thoughts' comprise the objects of the mind. In assuming an external perspective on thought in general, we can understand Mindfulness as encouraging a sense of detachment. The Mindfulness subject is also split, and along a similar 'internal-external' polarity; however, the polarity is situated in terms of *between* the thinker and their thought. In other words, Mindfulness conceptualises the thinker as the locus of subjectivity, with their thought being in some way beyond or separated from phenomenal experience.

Interior/Exterior

Two consequences of this kind of technique involve types of separation. Separation of the thinker from their thought, and of the interiority of thought from the exteriority of its social conditions. Mindfulness techniques rely heavily on using metaphors to substantialise one's feelings: one is encouraged to consider (for example) a sense of sadness in terms of shape, colour, opacity, size etc. This kind of visualisation method helps to exteriorise one's feelings - to separate the 'self' from its various sensations. This exteriorisation is also achieved in

[136] Trull, T.J. (2007) *Clinical psychology* (7th ed.) USA: Thomson/Wadsworth
[137] Thoma, N., Pilecki, B., & McKay, D. (2015) 'Contemporary Cognitive Behavior Therapy: A Review of Theory, History, and Evidence.*' Psychodynamic Psychiatry,* 43(3) pp.423-461. p.430
[138] https://www.mayoclinic.org/healthy-lifestyle/consumer-health/in-depth/mindfulness-exercises/art-20046356 (Last accessed 10/03/23)
[139] Thoma, N., Pilecki, B., & McKay, D. (2015) 'Contemporary Cognitive Behavior Therapy: A Review of Theory, History, and Evidence.*' Psychodynamic Psychiatry,* 43(3) pp.423–461. p.432-433

terms of one's personal experience. Sahanika Ratnayake explains:

> When eating the raisin, for example, the focus is on the process of consuming it, rather than reflecting on whether you like raisins or recalling the little red boxes of them you had in your school lunches, and so on. Similarly, when focusing on your breath or scanning your body, you should concentrate on the activity, rather than following the train of your thoughts or giving in to feelings of boredom and frustration. The goal is not to end up thinking or feeling nothing, but rather to note whatever arises, and to let it pass with the same lightness.[140]

Avoiding reflection in favour of immediate experience has the paradoxical effect of minimising experience. Ratnayake makes the point that in focusing on the process one avoids any sense of value: whether one is enjoying this activity or not. With Mindfulness, the internal stoicism of cognitive therapy is converted into an external value-based stoicism: all thoughts are equal in terms of their valuelessness. A stoic attitude towards immediate experience, as opposed to a stoic attitude towards one's environment, has the effect of externalising one from oneself. This dissociative condition amounts to a form of solipsism in which one's most immediate experience, through the act of becoming detached, or separated, consequently becomes a sort of 'private object' to become acquainted with and to 'treat' through the use of therapeutic techniques. Ron Purser calls this process "McMindfulness" and associates it with attaining "private freedom."[141] Private freedom is attained at the expense of social integration because, in order to practise Mindfulness, the process of internal separation and objectification of experience inaugurates an inward perspective:

> Mindfulness, like positive psychology and the broader happiness industry, has depoliticized and privatised stress. If we are unhappy about being unemployed, losing our health insurance, and seeing our children incur massive debt through college loans, it is our responsibility to learn to be more mindful.[142]

By severing itself from its social and religious roots, Mindfulness makes a broad, contextless appeal: if anyone can become a Mindfulness practitioner regardless of their cultural or religious background then in a sense those backgrounds become irrelevant to whether one can become 'mindful'. Shorn of its social and historical basis, we can picture a practice which is traditionless and as such, a form of therapy to be chosen rather than a culture to be inducted into. This context-free aspect of the treatment also enables it to merge with CBT as one module among others. Beck's goal of creating a *contra-psychoanalysis* is completed with the introduction of Mindfulness into CBT, not just because it confirms the separation of the individual from the social, but because it confirms the modular nature of the treatment form as a 'platform for wellness'. In achieving this modular form, CBT can take the place of a technical/commoditised approach to the treatment of mental health.

---

[140] Ratnayake, S. (2019) 'The problem of mindfulness'. *AEON.* Online: https://aeon.co/essays/mindfulness-is-loaded-with-troubling-metaphysical-assumptions (Last accessed 12/09/23)
[141] Purser, R (2019) *McMindfulness*. UK: Watkins Media
[142] Purser, R (2019) *McMindfulness*. UK: Watkins Media. p.5

3.6 cCBT

By the end of the millennium computerised alternatives to talking therapies were technically feasible but as yet unsophisticated. In a 1990 study, Paulette M. Selmi et al. noted that "Early efforts to develop interactive programs for psychotherapeutic interactions have shown only that the general idea was feasible; they failed to go beyond limited demonstrations."[143] The reason given was that the complexity of dialogue in psychotherapeutic techniques could not easily be transformed into the 'dialogue tree' style conversational structure characteristic of conversational computer programs. Selmi et al. point out that because CBT involves step by step procedures and 'targeted' treatment (i.e. directed towards precise 'cognitions' or behaviours), it is the ideal therapy for computerisation:

> A major impetus toward solution of these problems has come from developments in behavioral and cognitive treatments for depression in which treatment is directed toward specific target behaviors and follows explicit steps with clearly defined goals and outcome criteria.[144]

Computerised mental health has almost exclusively taken the form of online CBT treatment, encouraging users to learn to manage their own mental health through taking part in standardised cognitive behavioural techniques. Users follow a set course of treatment, often by following a text-based guide to help them to track their own moods and correlate moods with behavioural choices. The tracking of mood is an important feature of this style of treatment, as the user can then adjust their behaviours using their mood/behaviour graph as a guide for future action. Kenneth Colby's computerised therapy software during the 70s and 80s was confined to the realm of scholarly research due to technology limitations: access to powerful computers was only available to institutions like universities. In 1990, as personal computer-use was becoming more widespread in the USA, Colby set up a company called Malibu Artificial Intelligence Works and released the computer program *Overcoming Depression*.[145] A 1992 advertisement for the software reads:

> FEELING HELPLESS ABOUT DEPRESSION? Overcoming Depression 2.0 provides computer based cognitive therapy for depression with therapeutic dialogue in everyday language. Created by Kenneth Mark Colby, M.D., Professor of Psychiatry and Biobehavioural Sciences, Emeritus, UCLA. Personal Version ($199), Professional version ($499). Malibu Artificial Intelligence Works, 25307 Malibu Rd, CA 90265. 1-800-497-6889.[146]

Commercial development of therapeutic software expanded slowly throughout the 90s, being expensive to produce and a risky investment. Colby notes, in 1999, that the software was not intended to replace 'traditional' therapy, but was partly an engineering experiment and partly a commercial punt: to test the market and ascertain its readiness for this technology.

---

[143] Selmi, P.M. Klein, M.H. Greist, J.H. et al. (1990) 'Computer-administered cognitive-behavioral therapy for depression'. *American Journal of Psychiatry*, 1990;147(1) pp.51-56
[144] Ibid. p.3
[145] Colby, K. M. & Colby, P M. (1990) Overcoming depression: Professional version manual. USA: Malibu Artificial Intelligence Works
[146] Cited in: *The RISKS Digest*, Volume 13 Issue 83. Online: http://catless.ncl.ac.uk/Risks/13/83#subj5.1 (Last accessed 18/07/23)

Colby explains the bot's 'theatrical' character:

> Is the dialogue mode a simulation of a human cognitive therapist? No - it is not intended to represent a simulation or imitation of a human therapist. At times the responses resemble those of a human but that is only because the program's authors simulate themselves in designing cogent responses, i. e. responses consistent with the interpretation that the program has a therapeutic intent. Recall my mention of the virtual person in the dialogue mode. This conversational participant says many things a human therapist would never say, e.g. "I am sure my programmers would be glad to hear that" in response to a user compliment. Who is this "I" and "my"? It is a conversational participant with a particular character and set of attitudes that we have constructed. One might view its presence as a type of theater, thus lending the flavor of an art-form to the program.[147]

In 1999 the market for chatbot therapy was not quite ready for automated therapy, but since the release of the iPhone in 2007 there has been a huge increase in their popularity, with hundreds of mental health apps currently available, from treating a range of mental health issues such as anxiety and depression to offering meditation guidance and self-help, virtual assistants, virtual companions and more. While treatment via smartphone is similar to treatment through a home computer or laptop there are formal distinctions due to the portability of phones and their primary use as a communication device. The persistent availability of treatment due to phone portability has shifted the focus of treatment from something that is done at specific times and places to an 'always on' type of intervention. Features such as reminders, personified companions, mental health 'games'[148] and sleeping aids have both popularised mental health apps and diffused treatment into a wide range of different services. The developers of most mental health apps are quick to point out that these apps should not be used as an alternative to traditional treatment because of this diffusion, as it is difficult to determine just how successful treatment is, or in many cases, what is actually being treated. The first contemporary smartphone-based mental health chatbots were introduced by X2AI from 2016 onwards.[149] X2AI's range of bots, alongside Woebot which was introduced in 2017, were not apps as such, operating through Facebook messenger. Woebot soon became a fully-fledged downloadable app, available on online app stores; this was followed by others soon after, including ReMind. The introduction of a chatbot can be seen as the integrative means of delivery for the various mental health treatment-forms which had previously been available, initially via mail-order disk and increasingly via web-based services. Chatbots form the user-interface aspect of the application, both acting as a companion to the user and the deliverer of the various mental health activities. Hannah Zeavin calls chatbot therapy a form of 'auto intimacy'[150] in which

---

[147] Colby, K.M. (1999) 'Human-Computer Conversation in A Cognitive Therapy Program'. In: Wilks, Y. (eds) *Machine Conversations*. The Springer International Series in Engineering and Computer Science, vol 511. Springer. p.17

[148] SuperBetter: "The SuperBetter app uses the psychology of game play to achieve epic wins in all of life. Over 1 million people have played SuperBetter". Online: https://superbetter.com (Last accessed 20/05/23)

[149] Solon, O. (2016) 'Karim the AI delivers psychological support to Syrian refugees'. *The Guardian.* Online: https://www.theguardian.com/technology/2016/mar/22/karim-the-ai-delivers-psychological-support-to-syrian-refugees (Last accessed 01/11/22)

[150] Zeavin, H. (2021) *The Distance Cure: A history of Teletherapy*. USA: MIT Press

the user performs a kind of self-treatment. Zeavin claims that auto intimacy is due to the lack of a therapist - an 'other' with whom one must disclose oneself to in order to benefit from the treatment. 'Intimacy' in terms of in-person treatment necessitates the presence of the other human subject. On Zeavin's terms auto intimacy involves the technical device (a computer or phone) taking the place of the other, so that the user still experiences intimacy in an internalised, self-oriented way. The presence of the smartphone, as opposed to a personal computer is also a factor in the development of this auto intimacy. Automated computerised therapy has been available prior to the introduction of smartphones but has only become popular since becoming available on these personal devices. One reason for this is due to the ease of software distribution on smartphones as internet connected devices and associated low cost. Another reason is that the home computer, which was often a shared device, offered less of a personal, 'intimate' experience than smartphones. We can understand the introduction of the smartphone as the catalyst for the contemporary acceleration and expansion of computerised mental health treatment. The smartphone, much more than the personal computer, offers a truly 'one-to-none' experience. It is worth recalling Colby's description of a chatbot therapist which is designed to provide "responses consistent with the interpretation that the program has a therapeutic intent."[151] A chatbot instigates the user in assuming an attitude towards the software with the expectation of a therapeutic experience. According to Colby this is not an attempt to simulate therapy but to create a 'theatre' which conjures a therapeutic experience. The treatment activities (CBT, Mindfulness, self-help activities, etc.) which comprise the features provided by the chatbot are ostensibly the 'genuine' therapeutic aspects, 'beneath' the theatrical display. However, in this way we can understand the introduction of chatbots as the culmination of the arc of the behaviourist ethos in which the human mind is approached as 'conditionable', and the cognitive ethos in which the human thoughts are 'functions'. Colby's assessment correctly identifies an epistemological approach which defines a tendency throughout the lineage traced in this chapter: that observation - and experience - of a therapeutic 'effect' can be taken as the marker of therapy having taken place. The chatbot assists the therapeutic effect in crossing the boundary from external observer to internal user by providing, in 'theatrical' form, the experience of a therapeutic effect.

3.7 Conclusion

Psychoanalytic treatment predates and influences cognitive and behavioural therapy, which developed largely due to attempts to scientifically ground psychoanalytic concepts and methods through empirical experimentation. The 'social-subject' of psychoanalysis gradually gave way to a more empirically observable 'scientific' subject. The 'ego-psychology' variant of psychoanalysis which developed in the USA incubated a theory of the subject as conceptually distinct from its social context through a focus on 'adaptation'. 'Adaptation' informs the attempts to bring psychology into an experimental scientific mode in ways that are unique to behaviourism, cognitive psychology and Mindfulness, the three pillars of CBT. Cognitive Behavioural Therapy is an umbrella term which covers a range of different treatment methods including cognitive therapy, behaviour therapy and newer "third wave" methods which include dialectical behaviour therapy and acceptance and commitment

[151] Colby, K.M. (1999) 'Human-Computer Conversation in A Cognitive Therapy Program'. In: Wilks, Y. (ed.) *Machine Conversations*. The Springer International Series in Engineering and Computer Science, vol 511. Springer. pp.9-19. p.17

therapy (ACT).[152] The amalgamation of newer treatment forms under the rubric of 'CBT' was possible because these methods approach the human individual and human mind respectively as strictly distinct from their social contexts. Through the lens of behavioural and cognitive treatment, the subject could be seen to behave and to think against the 'background' of the social world. 'Mindfulness' is a nebulous term for a range of different approaches to mental health which include self-led meditation techniques to guided ACT. The addition of these approaches comprises the "third wave" of CBT. Along with cognitive and behavioural treatment methods, Mindfulness 'rounds off' the individual/social distinction by affirming the isolation and interiority of the individual mind as opposed to the exteriority of the social world. The development of automated computerised therapy has been a dream of psychologists since the development of conversational chatbots in the 1960s but has only become an acceptable possible solution since the introduction and mass availability of smartphones. The reason for this is not just due to their availability but also due to the intimacy associated with smartphones. Prior to the rise of smartphones, the average home computer in the 1990s would have been technically capable of running contemporary therapy chatbot software, but the public interest in such software and its commercial feasibility only became apparent when available on smartphones.

As we will see, the experimental engineering approach to automated therapy has persevered with ReMind, as well as commercial imperatives influencing not just business decisions but also design choices. As with CBT, contemporary therapy chatbot software does not explicitly depend on a theory of the mind as identical to a computer, but an implicit assumption that the mind is akin to a computer influences how ReMind understands 'the mind', and 'mental health'. The historical and conceptual lineages discussed above are not external to this new technology, but comprise its genealogy: previous technical and therapeutic inventions implicitly influence and steer contemporary automated treatment. By introducing a chatbot to deliver mental health therapy, we see a culmination in the arc of various different strands of intervention into the human mind. From the machine-like mind of behavioural psychology, which sees the human as inherently adaptable, to the computer-like mind of cognitive psychology, which sees the human as an information-processing machine, we can understand the rise of a machine-like form of treatment in Cognitive Behavioural Therapy. The computerisation of CBT can be seen as a natural progression of the treatment form in that CBT had already been 'manualised' as a self-treatment: the range of techniques could be presented as step-by-step 'modules' with no need for a clinical practitioner. The addition of a chatbot allows users to simulate the experience of speaking to an interlocutor, i.e. to experience the effect of this encounter without undergoing the encounter. The bot completes the sequence of asserting that, in terms of how the mind works, 'function' and 'essence' are identical.

[152] Thoma, N. Pilecki, B. & McKay, D. (2015) 'Contemporary Cognitive Behavior Therapy: A Review of Theory, History, and Evidence.' *Psychodynamic Psychiatry,* 43(3) pp.423–461. p.423

Chapter Four: Methodology and Methods

4.1 Introduction

The literature review chapter shows that much of the current research is focused on assessing the effectiveness of the therapy when automated. What is lacking is a discussion about how 'effectiveness' itself is determined, and assessment of the therapeutic forms in terms of how automation affects the quality of the therapy. This work builds on the research discussed in the literature review by providing an analysis of the contexts from which the research has thus far been conducted. This means that concepts like 'mental health' and 'technology' will be considered and problematised in terms of how they are represented in the literature. In short, this representation involves 'non-critical' analysis, meaning that technical treatment is approached as an engineering problem, which can be solved through technical intervention and assessed in technical terms. Andrew Feenberg characterises this approach in terms of assessment using concepts which are already defined by the parameters set by the object of one's assessment: "To judge an action as more or less efficient is already to have determined it to be technical and therefore an appropriate object of such a judgement. Similarly, the concept of control implied in technique is "technical" and so not a distinguishing criterion".[153] This project assesses ReMind such that concepts like 'efficiency', 'technical' and 'effectiveness' are problematised. This chapter is split into three sections, the first section lists the research questions which drive the project, the second section discusses the methodological background informing the project, and the third second discusses, in a loosely chronological fashion, how I conducted the project.

This project is ultimately guided by the double-sided question:

What are the conditions of possibility for automated mental health treatment, and how does automated treatment alter subjectivity?

This double question is broken down into a range of sub-questions. The kinds of questions that the project asks are defined as 'narrow' and 'broad'. Narrow questions are ones which have ostensibly routine answers, i.e. they can be pursued to some reasonable end. Broad questions are more speculative, they are concerned with socially and epistemologically contextualising ReMind as a product of its circumstances. The most basic and immediate questions which serve to set this project in motion stem from the immediate description of ReMind as a 'therapy chatbot': what therapy does it do, and how does it operate through a chatbot? The five 'narrow' questions are loosely associated with each of the analytic chapters, the 'Methods' section of this chapter will discuss how each of the analytic chapters broach these questions. My general approach to this thesis involves initially pursuing the narrow questions in terms of empirical evidence, and using the subsequent analysis to gather together material with which to consider the broad questions.

[153] Feenberg, A. (2005) 'Critical Theory of Technology: An Overview'. *Tailoring Biotechnologies*. Vol. 1, Issue 1. pp. 47-64. p.47

Narrow

1. What kind of mental health treatment does ReMind engage in, i.e. what are the treatment methods that ReMind has used in their software? And following this: are these methods transformed in any way due to being computerised?

2. What are the technical features which come together in the ReMind app? This question seeks to understand how ReMind works on a straightforward level: how is the chatbot designed? How is it programmed to converse with users and to suggest mental health activities?

3. What do the members of the ReMind team think about their technology? How do they understand how their software intervenes in the mental health of the users, what indeed, are the ReMind team's concepts of mental health?

4. Who is the app for? This question does not poll users of the app but rather asks: what kind of people is ReMind expected to be applicable to?

5. What is it about a chatbot-based mental health app that attracts the users to it, and sustains their usage of it? What is it about the addition of a chatbot that generates user-interest?

6. What is the commercial context in which ReMind operates as a company, and in which they deploy their app? In other words: are there economic demands which ReMind must respond to, and if so what are those responses?

Broad

How to write an account of mental health chatbots which tells us about the wider social context from which they have emerged? In other words, how do I approach this project in a way that helps us to understand, not just how this new technology works, but *why* does it work?

What can ReMind tell us about contemporary concepts of mental health and mental health treatment? In other words, if ReMind is considered as an appropriate intervention into helping people to tend to their mental health, in light of how the ReMind app works, what then *is* mental health?

Can analysis of ReMind, and other similar forms of smartphone based mental health interventions, tell us anything about wider concepts of human subjectivity beyond the confines of mental health and mental health treatment? In other words, does ReMind tell us anything about what it is to be human beyond the scope of mental health?

4.2 Methodology

Critical theory methodologically guides this project. Critical theory is not a theory as such: it does not seek to explain society by applying a prior philosophical system but is rather a method of analysis of social forms from which a philosophical system may be deduced. This means that instead of attempting to explicate the foundations (e.g. God, the individual, the family, the state, etc) from which a social order originates, the task of critical theory is to understand how and why the structure of the social order operates and perseveres. It is then from this structure that 'foundational' concepts, such as God or the individual, or in the case of this project; 'mental health', can be theoretically extrapolated. Andrew Feenberg describes critical theory as "philosophy of praxis" in which "history is the 'paradigmatic order' for the interpretation of being generally."[154] This means that concepts of human subjectivity are historically determined, and must be derived through analysis of socially-historical context, i.e. the 'conditions of possibility' for subjectivity. The overarching goal of this project is to consider automated therapy from a number of different perspectives, or logics, which can be summarised as: a social theory of technology, a critical theory of economy, and a psychoanalytic theory of the subject. This project will consider these three approaches as linked by a common concept - that of 'materialism'. A materialist approach asserts that social practice tends to prefigure, in form, the ensuing contents of socio-political ideology.

This project shares an assertion which guides science and technology studies - that scientific research and technological progress are inseparable from the social and historical contexts within which they occur. The basic premise of this approach is that science and technology are not self-determined and following a 'natural' course of progress, but are influenced by, and also influence the society in which they are present. On one hand, attitudes, values, and inclinations and on the other hand, historical processes - all of which are associated with the idiosyncrasies of people or groups - must be included when considering how and why current techno-social conditions have come to be the way they are. Social values, as crystallised in technological forms are, according to Feenberg, made up of accumulated decisions made by the designers of those technologies. What informs these decisions? Designers of mental health apps must make their decisions based on the technical opportunities and limitations afforded by their chosen technical basis – in this case electronic computation. They also must respond to market conditions, and these responses have implications on the implementation of a technical mental health intervention. The development of this new technology takes on a wholly novel aspect when considered under the auspices of scientific progress: traditionally science has developed under a regime of free exchange of ideas, theories, developments, etc, but AI is a largely commercially driven endeavour. Mental health therapy has a public/private distinction operating between the ways that it is delivered - state health provisions vs. private practices, but the development of its theoretical understandings and practical routines has historically operated as a public concern. What happens when this aspect assumes a privatised and commercial trajectory? A critical approach will contextualise the underlying processes involved in the design of ReMind, to better understand the ideological features (masked as necessities/efficiencies) in its design.

[154] Feenberg, A. (2014) *The Philosophy of Praxis: Marx, Lukács and the Frankfurt School*. UK: Verso p.5

This project problematises the 'mental health' in two distinct ways: by critically analysing the practices which intervene in mental health, and the term itself. There is little consensus and even extreme divergencies in theories of mental health and mental illness. Despite a large corpus of research over the last hundred years, "we are still unable to provide any incontrovertible evidence of either what causes mental distress or how it can be treated effectively."[155] The term 'mental health' is approached critically, as being a product of "interplay between various psychological, social, political, environmental and biological factors."[156] In developing an immanent critique, this project develops a theory of 'mental health' from the analysis of therapy chatbots - which are themselves a product of interplay of various factors - rather than imposing a theory onto the chatbots. There is an urgent need to maintain a critical approach to new forms of mental health interventions, and especially, as discussed above, technologically informed interventions. Bruce Cohen points out "in the current environment of expanding mental health jurisdictions – why we need to think critically about the practices and priorities of the contemporary system of mental health."[157] This means that interventions into 'mental health' is undergoing expansion in both methods and arenas, with understandings of the consequences of this expansion lagging behind.

This PhD project is an attempt to, in a very broad sense, consider human *subjectivity* as opposed to human *nature*. What this means is that our ideas about what it is to be human depend on our conceptual and social coordinates. This project theorises the human subject starting from the assumption that the individual and the social are interlinked. This means that 'the subject' equates to the individual insofar as one is produced by one's social and symbolic environment. Romin Tafarodi describes subjectivity as "both the "first personness" of consciousness (being a subject of experience) and the conditioning of that consciousness within society (being subject to power, authority, or influence)".[158] My research will specifically follow two approaches to theorising how the subject is shaped by automated mental health therapy: 1. As an *a posteriori* conceptual apparatus implicit within the interlinked clinical, economic and technical contexts of mental health chatbots. 2. As an *a priori* concept of 'personality', or 'the individual', that is held both consciously and unconsciously by the developers and users of mental health chatbots. While this project is not an attempt to 'psychoanalyse chatbots' or to offer psychoanalytic explanations, psychoanalytic theory informs, in a similar way to critical theory, the approach that I undertake. This means that a focus on 'conditions of possibility' regulates the approach to questions of psychology. One such condition that this project relies on is the social determination of the individual. This does not quite refer to the specific social settings within which various individuals are ensconced, but rather the fact that all individuals are socialised in one way or another. In proposing the 'always already' social constitution of the individual, Freud problematised the very distinction between the 'individual' and the 'social'. A concept of the subject, as both receptive and resistant to different discursive structures will be used to consider how mental health therapy (in the form of a chatbot-based intervention), like any other discursive structure, assumes a particular concept of the subject in order to establish its own epistemological foundations, interpellate subjects, and finally, to effect treatment.

[155] Coppock, V. & Hopton, J. (2015) *Critical Perspectives on Mental Health*. USA: Routledge. p.10
[156] Coppock, V. & Hopton, J. (2015) *Critical Perspectives on Mental Health*. USA: Routledge. p.10
[157] Cohen, B. (2018) 'The Importance of Critical Approaches to Mental Health and Illness'. In: Cohen, B. (ed.) *Routledge International Handbook of Critical Mental Health.* UK: Routledge. p.10
[158] Tafarodi, R. (2013) *Subjectivity in the Twenty-First Century*. UK: Cambridge University Press. p.i

‘Interpellation’ means the transformation of an individual into a subject, effectuated by an ideological structure. From this basis, an analysis of ReMind will be made in order to determine what conceptualisation of the subject is being projected by, and subsumed into automated mental health therapy. This means that through analysis of ReMind a concept of the subject, and of mental health, can be reconstructed through analysis of the social practices and objects involved in the production and deployment of this technology.

Ethnography

The ethnographic component of this project is guided by digital anthropology methodology. Digital anthropology is a recent field, which is now recognised as an authentic and vital approach to considering virtual worlds as ethnographic field sites. The online setting as a field site is the focus of fieldwork but in contrast to non-virtual field sites, it works as a mediating forum for the facilitation of social or business relationships rather than the encompassing and physical conditions for a range of different social activities. The ethnographic work pays attention to Miller and Horst’s principles laid out in their essay ‘Six Principles for a Digital Anthropology’.[159] These are intended to guide the researcher when undergoing digital ethnographic work in order to not lose sight of basic ethnographic principles.

1. *The digital intensifies the dialectical nature of culture.* Digital technology and the virtual realms which it makes possible seems to shift human culture into a completely novel mode of operation, but Miller and Horst maintain that culture is always dialectical: shifting between the apparent possibility of being completely mediated or non-mediated. What digital technology offers is a sense of novelty to this dialectic. New forms of mediation offered by virtual interactions can be strikingly ‘unreal’ and inauthentic seeming, but as new forms of technology become customary and lose their novelty, their mediating qualities become less noticeable, allowing the technology to drift into the background of seemingly authentic and non-mediated sociality.

2. *Humanity is not more mediated due to the introduction of the digital*. Miller and Horst urge us to consider mediation as a principal condition, rather than a feature of human subjectivity. Mediation is not an experience or sensation which can be increased or diminished depending on the cultural or technological circumstances, but is a condition which finds expression in different formats. The virtual worlds of digital media can be thought of as one such expression. This means that this ethnographic research must acknowledge the peculiarities of online life but avoid assuming that those peculiarities signal a completely dissociated social realm which is ‘unreal’ compared to the reality of the physical world.

3. *Commitment to holism*. People do not conduct their lives exclusively through any single medium; digital activity is one among many other modes of being. However, each medium depends on, and affects others: Miller and Horst use the example of Gerson’s recounting of the acrimony of relationship breakups being exacerbated due to having been conducted online. A commitment to holism recognises that the specific focus of an ethnographic study cannot be understood as comprising a total picture of the lives of participants, and that

[159] Miller, D. Horst, H. (2021) ‘Six Principles for a Digital Anthropology’. In: Geismar, H. Knox, H. (eds.) *Digital Anthropology*. UK: Routledge

external factors must be acknowledged in order for, on one hand, the specificity of the focus to make contextual sense, and on the other hand, for generalisations to then be made during analysis.

4. *Asserting the importance of cultural relativism.* While the internet exists in the same form globally, as an interlinked series of networked computers, 'the internet' is not the same thing for all people. Different societies will appropriate online platforms and media according to prior social needs, and in turn these platforms and media will influence the culture which appropriated them. Anthropology asserts the importance of forgoing the assumption that the culture being studied is immediately recognisable to the researcher; this is also true of digital cultures. In the case of ReMind, forgoing prior assumptions demands acknowledging that cultural diversity and homogeneity may apply to ReMind in surprising ways due to the online digital nature of conducting business.

5. *The digital is ambiguous with regard to openness and closure.* Digital media enables and prohibits social connectedness in various ways. For example, the ability to communicate with every participant on ReMind's Slack channel, despite their geographical dispersion, is contrasted with a singular mode of communication - written text. Online fieldwork has to consider how different communicative modes like physical gestures and eye contact are not possible on a text-based communications platform, but that access to participants is enhanced in that physical proximity is not required to be able to conduct conversations.

6. *Materiality of the digital world.* The virtual world is imbued with a sense of immateriality, graphical displays based on immediately editable code give the appearance of fluidity and flexibility, the speed at which internet-based cultures develop and transform gives a sense of insubstantiality. On the other hand, the persistent nature of the internet, and the ease of data storage afforded by computers means that documents, files, traces of communications and metadata remain stored often against the wishes of the data producers.

4.3 Methods

This section comprises an overview of the methods and analysis used throughout the research. It does this through a roughly chronological account of the project, beginning with my initial approach, covering the ethnographic fieldwork, and finishing with an overview of the analytic work. The ethnographic material is interspersed throughout the thesis, to give substance and perspective to the analysis, and to problematize my own theoretical assumptions. Because this project is not an ethnography, but includes ethnographic work, analysis is not just of data gathered during the fieldwork but also involves discussion of technical details of the app, the historical and social contexts in which it is situated, and other theoretically informed analyses. Fieldwork involved six months acting as a participant observer with ReMind to study the working culture from within. Participant observation and interviews were the primary forms of data collection. The investigation took place online, as a participant on ReMind's electronic communications channels. I was involved in a working group and communicated with them via Slack: web-based electronic communication software. I was also able to observe other working groups and general discussions and conversations via Slack. Before and during the fieldwork I was conducting analysis, this was then consolidated into the five analytic chapters which follow this chapter.

How it Began

The PhD was conceived as a theoretical-analytic research project, looking broadly at mental health apps. I narrowed it down to chatbot-based apps because I wanted to look at attempts to electronically replicate the therapy experience: apps that try, not just to provide automated mental health treatment, but to provide the user with automated *therapy*. Between 2018 and 2019 I had negotiated visiting a USA-based mental health chatbot company called X2AI to conduct ethnographic fieldwork. X2AI were interested in my project and enthusiastic about inviting me to their office in San Francisco to act as a participant observer. We stayed in touch while I began the research. The initial phase of research, through 2020-2021, involved wide-ranged reading and writing on artificial intelligence, the history of cybernetics and electronic computation, psychoanalytic and critical theory. The aim of this work was to cast a wide theoretical and methodological net, with an epistemological framework gradually coming into focus. Over the course of the COVID-19 pandemic, correspondence from X2AI became less frequent and eventually tailed off. After I had decided that X2AI was unlikely to work out I emailed other mental health chatbot companies, including ReMind, to find an alternative. Charley,[160] ReMind's head of clinical research and development, who was to become my primary contact, replied a few days later asking if I'd like to meet online to discuss the proposal. We agreed that I would join Charley's research team as part of ReMind, operating under a work-placement agreement, essentially as an intern so that I could become immersed in ReMind's daily operations. After receiving ethics approval from my department and signing ReMind's non-disclosure agreement, the fieldwork officially started on the 1st March, 2022. What this meant was that I was provided access to ReMind's Slack channels and provided with a ReMind Gmail account.

On The 'Ground'

The fieldwork was conducted online, the majority of my involvement taking the form of inclusion into ReMind's Slack channels. This means that I was not engaging with directors and employees in a face-to-face manner, (although I did conduct one interview in person, as the interviewee happened to be in London at the time) but was mediated by text, video link and recorded audio. According to Susan Blum, this type of research is now more frequent but not a completely new phenomenon: research via posted surveys, war interrogation reports and newspapers have all been carried out over the nineteenth and twentieth centuries.[161] Blum notes that conducting fieldwork at a distance from participants can give rise to a sense of inadequacy or inauthenticity, but all ethnographic research is mediated in some way. My own research was congruent with the working manner of ReMind employees, who, while having access to a physical office in their home city, almost exclusively conducted their operation online. Participants were encouraged to treat me as a co-worker, while acknowledging that I am there to observe and analyse their social interactions. There was a paradoxical element to this as I was not trying to deceive ReMind employees into thinking that I truly was a co-worker, but was clear that my research involved working *with* them rather than against them to understand how their interactions, opinions, beliefs, etc, are connected to, and influential on, the development of mental health software. Considering Millar & Horst's fifth principle, discussed above, this was made both more possible and more

[160] All employee names are pseudonyms to maintain anonymity.
[161] Blum, S.D. (2020) 'Fieldwork from Afar'. *Anthropology News.* 10.14506/AN.1483.

difficult due to fieldwork being conducted online. My introduction to participants was in terms of my inclusion in ReMind's research team, much like in a face-to-face work environment, however due to time differences based on geographical locations this introduction and induction was staggered rather than immediate. This chronological staggering of communications is one of the primary features which differentiates online from in-person fieldwork.

Slack

ReMind conducts their operation through Slack, which is a business communications platform. Slack involves persistent chat rooms, private groups and direct messaging. Data collection took the form of note taking and interviews. I kept a research journal documenting my own thoughts, ideas, and speculations as I conducted the research. This helped to generate hypotheses, develop my thoughts and formulate interview questions. Due to the persistent online nature of the workplace, previous conversations were available to view. This means that an archival log of the company in conversational form could be accessed in order to compare and contrast past with current activity and discuss historical activity with participants. For ethical reasons I did not access conversations that occurred before I started. Recorded interviews were conducted with directors and employees periodically over the course of the placement, with times and frequencies negotiated individually. Interviews were semi-structured and tended to be quite informal Questions avoided 'yes or no' answers to achieve structurally open-ended discussions to encourage informants to provide rich details. Some examples of how questions[162] were approached are:

- Does your work here reflect your own ideas about mental health and treatment?
- What kind of people do you think are attracted to using this software?
- How do you compare using an app to seeing a human therapist?
- What are your hopes and fears as you embark on this project?

During the course of the fieldwork it proved difficult to organise interviews due to employees being situated in different time zones but primarily because they were often so busy that scheduling enough time for non-work-related activities was often a low priority. However, a number of informants were enthusiastic about being interviewed as the open-ended discussion format was a chance to air speculation and not fully-formed thoughts about the company and the ReMind service that they were making. I interviewed 13 employees (including two founder/directors), with a number of them agreeing to follow up interviews. All names have been changed to maintain anonymity:

Jeff - Company founder director
Reese - Company founder and director
Arnold - Product director
Charley - Head of clinical research and development
Andy - PR manager
Alan - Conversation designer
Mary - Head of AI

[162] Full interview guide is in the appendix

Samantha - Chief psychologist
Daniel - User interface and user experience (UI/UX) designer
Patricia - Senior therapist and research associate
Sage - Therapist
Roman - Lead psychologist
Henry - Chief operating officer

Ethics

Employees of ReMind were given notice of my participation before I arrived. I explained that I would be participating in the work environment in order to understand the working culture from within. I explained that this meant I would be engaging with employees not as an external observer but as a colleague. However, this would not involve trying to 'trick' anyone: I made it clear that I was there as a researcher, but that I did not intend on gleaning information covertly or behind the backs of employees. I made sure to be clear that I was there as an interested party rather than a critical judge. I was aware that due to the project taking place in a work environment, issues such as power-struggles, workplace bullying, occupational hazards and industrial relations may arise. These issues could impact confidentiality and/or the impartiality of the researcher: it could be difficult not to take sides during the course of disagreements. My approach to this involved making sure that I would focus on my own impartiality and also to ensure the confidentiality of employees, both in the case of disclosing information which might identify them or the company, but also in the case of potentially identifying employees to the company, if their disclosures could jeopardise their relationships within the company. All participants were anonymised at the point of data collection – I explained to participants that the audio recording would be transcribed two weeks after recording, and that when transcribing their interviews personally identifiable information would be changed. The participants could withdraw consent after the interviews took place up until the point of transcription. If an employee had a unique role within the company which might identify them, I ensured anonymity through avoiding reference to specific tasks or roles. If during the course of an interview an employee divulges information which may be cause for concern such as bullying or industrial disputes, I would emphasise that the interviews are anonymous, and that the participant can withdraw or amend responses to interview questions after the interview has been completed. If the interviewee were to declare concerns which would need to be reported to a third party, the content and wording, and the third party in question would be discussed under terms of anonymity during the interview. Before an interview I explained to the participant that they would have:

- The right to decline participation

- The right to withdraw from the activity at any time or refuse to answer any particular question

- The right to have privacy and confidentiality protected and if they cannot be maintained the fact that the participant(s) would know this from the outset and consent to this condition

- The right to turn off a recording device at any time

- The right to ask questions at any time

- The right to discuss the way in which their data may be used

- The right to discuss the question of the ownership of the data and to reach agreement on issues of copyright

- The right to receive information about the outcome of the activity in an appropriate form

In practice, due to the short-term nature of the fieldwork, and due to my involvement being low-risk, these issues rarely arose. Interviews often involved quite informal, candid and collegial discussions: my interviewees were usually very enthusiastic about airing their thoughts, concerns, aspirations and speculations concerning their roles in the company, different aspects of the ReMind app, and the commercial context of their project. I never detected any worries from interviewees about whether their statements might be compromising or risky in any way.

Daily Routines

My daily experiences during fieldwork were varied. To begin with I spent some time observing ReMind employees' Slack interactions, trying to get a sense of who was who. As I had not had a formal induction with the company I had to rely on observation to develop an understanding of the company structure and hierarchy. After about a week I was invited to take part in a research project with ReMind's Clinical Research and Development Team, headed by Charley. This team was to be my home throughout my placement with ReMind. It consisted of a floating group of between around 4 and 6 employees, depending on the scope of each project. The first project involved laying out a slideshow illustrating various features of the ReMind app to show to potential investors. I also joined another project which involved formulating questions for a quantitative study design. Projects would usually involve taking part in a group meeting of employees who were either already involved in similar projects, who wished to gain experience in the type of work we were doing, or who had been specifically tasked with being involved. Tasks would be set out according to a project's aims and timeline, team members would often delegate themselves for each task, sub-groups (usually of two or three team members) would be formed and the project would begin. Throughout the course of my fieldwork I conducted occasional interviews with employees. About halfway through my placement, I suggested to Charley that I contribute my efforts in a way that was more appropriate to my training as a socio-philosophical researcher. To this end we devised a project where I would continue my interviews with employees and produce a series of corporate ethnographic studies which would illustrate ReMind's internal operations and structures. About half way through this project I was asked to modify my approach towards writing a series of articles to chronicle ReMind. These articles were to potentially comprise a section of their website. I worked with another researcher, Mandy, to draft and edit the series of articles. By August, I had completed drafts of four articles: 'Thinking behind ReMind', 'How ReMind was built', 'What ReMind is now', and 'What kind of users use ReMind'. These drafts were handed over to my colleague Mandy when I ended my fieldwork placement.

4.4 Analysis

This part discusses the 'narrow' questions listed above in terms of the methods used to answer them, with references to the 'broad' methodological framing which underpins the research. Each analytic chapter deals with these questions, with each of the five questions being associated with, but not totally informing, the content of each chapter. All of the questions are considered at various points within each chapter, with some featuring more prominently than others. The reason for this is that, as will be discussed, all of the questions are interlinked in ways such that providing an individual answer to each would be reductive: the complexity of ReMind in terms of its technical elements and social contexts demands multi-layered investigation. What follows is a discussion of how I pursued my research questions with reference to each of the analytic chapters; this is done not to summarise each chapter but rather to bracket the analytic questions in a manageable way in order to explain my approach.

Digitisation

The draft articles I wrote at the end of my fieldwork became the foundations for some of the analytic chapters of my project. Much of the data from interviews involved technical and historical aspects of ReMind, and as will be seen, much of the empirical aspects of this thesis involve analysis of interview discussions. The 'ethnographic' aspect of the project in terms of analysis of my participant observation, over the course of the ensuing thesis-writing, took on secondary importance to the data that was gathered from interviews. The first question listed above was a frequent topic in interviews:

> What kind of mental health treatment does ReMind engage in, i.e. what are the treatment methods that the ReMind team has used in their software? And following this: are these methods transformed in any way due to being computerised?

These questions undergo a three-step sequence in attempting to answer them. The first step is to identify the treatment methods which the ReMind company claims to use, and also to identify and categorise all of the mental health activities provided by the app. The primary mental health techniques that ReMind uses, as claimed by ReMind, are CBT and Mindfulness. While this claim needs to be assessed, it leads to a number of sub-questions and the secondary questions. These sub-questions are initially dealt with in the history chapter, and are mostly concerned with the epistemological underpinnings of CBT and Mindfulness. Questions such as: what are the practical activities which define the disciplines of cognitive and behavioural therapy, and of Mindfulness? Can we identify the therapeutic practice of 'CBT' through observation of a clinical practice, through analysis of its codified techniques found in therapy manuals, or by other means? These questions are considered in the first analytic chapter because it logically follows from the material in the history chapter. The reason this is partly dealt with in the history chapter is because of the historical nature of defining social practices: what we call 'CBT' nowadays might be different to how it was initially envisioned. My concern with chapters four and five ('History' and 'Digitisation') is to gain a sense of what 'logics' can be identified as having persevered throughout the development of CBT, and then to understand these logics in terms of computerised automation. The question of how technical automation might change an activity or practice is

considered in the first analytic chapter - Digitisation - not in terms of whether ReMind 'really' does CBT or not, but in terms of what kinds of concepts of mind can be inferred through analysis of the app and the company. Much of the analytics work of identifying the therapeutic technique depends on extracting various 'logics'. Identification of ReMind's treatment methods was done through interview discussions, observation during my fieldwork and direct analysis of the ReMind app. The secondary question then concerns technical automation through computerisation. Do the various therapeutic techniques referred to by CBT and Mindfulness depend on a clinical practitioner i.e. a therapist who dispenses the therapy? What happens to a therapeutic method when it is automated, i.e. does automation involve substantive changes in the method, what gets left out and what remains throughout this process?

Conversation Design

Much of the analytic work involves considering individual technical aspects of ReMind and extrapolating from that using the theoretical framework that developed throughout the course of the project. 'Extrapolating' here means drawing out, from the initial analysis, the logical consequences of these technical aspects in terms of their social, therapeutic, economic, effects. The second 'narrow' question covers the initial phase of this analytic process:

> What are the technical features which come together to comprise the ReMind app? I.e. how is the chatbot designed? How is it programmed to converse with users and to suggest mental health activities?

The initial approach to answering these questions involves 'reconstruction' of the various technical features which come together as ReMind. Prior to conducting the fieldwork, this analysis involved research on chatbots and their computational and social conditions: looking at what differentiates them in terms of their design and technical features, and looking at the history of chatbot design and implementation. Also covered were different aspects of artificial intelligence such as large language models, natural language processing, neural networks, and more. This was to establish a general grounding in the computational basis of ReMind. A major methodological aspect of a Marxian approach to social theory is that of reconstruction - that the object being analysed becomes, first deconstructed into various 'moving parts' and then reconstructed in a textual form. The analysis does not intend to make some kind of hermeneutic interpretation or semiotic deconstruction but rather attempts to develop an account of the 'concrete' processes that are integrated in ReMind. 'Moving parts' here refers to the various social, technical and historical processes, the logics of which can be conceptually differentiated in order to consider them on individual terms. This is not meant to exhaust the various contingencies involved in the production of the bot but rather to construct a picture of how this object is a coincidence of these logics. A primary concern in the analysis is to consider how the various decisions which are made in terms of designing the app influence how the app is then interacted with by the users. In practice, I start from what is immediately observable: that it is a smartphone app, it includes a chatbot and it offers some form of mental health assistance. A major technical aspect of the bot is its conversational feature, in which the user and the bot engage in text-based conversation. This means that I consider specific design aspects, such as how conversations are programmed, and then extrapolate from those aspects to theorise the effects of the kind of

conversations that ensue from this programming. A theory of mental health can gradually come into focus as the various aspects of the app are treated in this way. This strategy can be thought of as reconstruction because the social formations are always forefronted throughout the analysis rather than abstracted from; the analytic work involved is directed towards gaining an understanding of the conceptual foundations implicit within, and necessary for the cohesion of, any given social formation.

Macro-Treatment

The previous question is further pursued in the third analytic chapter. An important part of the research is about the beneficiaries of ReMind: those who use the app. My project does not take into account the 'actual' people who use ReMind, but instead takes into account the 'intended' or 'implied' people. That is why my question regarding the users is not 'who uses the app?' but rather:

> Who is the app for? I.e. what kind of people is ReMind expected to be applicable to?

This is, on one hand, to limit the scope of the project to that of the development side, and on the other hand, to maintain an approach that considers subjectivity as implicated in, and influenced by ReMind. My approach is to consider the various logics which I have identified as being operable in the various ways that interaction with the app is made possible. This is done, again, through analysing technical aspects such as conversation design, but also in terms of the mediating presence of the app. The ReMind company works through the app to intervene in the mental health of the users of the app; what are the effects of this mediation? A social theory of the subject is important in pursuing this question because as discussed above, 'the subject' is not an individual person or a group of people, but is rather a socially produced identity which is assumed by individual people and groups of people. A central and guiding question of my thesis is: what concepts of mental health must be harboured in order for ReMind to be a viable treatment? Answering this question does not equate to uncovering the concepts held 'in the minds' of the makers of this technology, but by assessing the conceptual conditions implied by the emergence of the technology - its epistemological conditions of possibility. This involves reconstructing, through analysis of the chatbot and its developers, these concepts of mental health. This is done by logically drawing out the epistemological conditions of possibility of the ReMind app. A common way of illustrating this approach is to put forward the question 'what must things be like in order for this phenomenon to exist?' In terms of how ReMind - both the company and the app - approaches the users of the app, this question involves looking at how the app works as a mediating device which facilitates a connection between the company and the users.

Suspension of Disbelief

This project tries to understand how chatbots work, not just in a technical sense but in a social sense. This means considering what kind of subjective and social dynamics are in play when a person speaks to a chatbot, which then responds to that person. Considering this involves looking at chatbots and also other non-human objects that generate some kind of social interaction. This problem is summarised with these questions:

> What is it about a chatbot-based mental health app that attracts the users to it, and sustains their usage of it? What is it about the addition of a chatbot that generates user-interest?

This question involves a number of other questions: What does it mean to interact with a chatbot? What attitude must one assume to conduct a conversation with a non-human computerised avatar? And following this, what happens when this chatbot is designed to address the mental health of the person talking to it? While this project is not an attempt to 'psychoanalyse chatbots' or to offer psychoanalytic explanations, psychoanalytic theory informs my approach. Psychoanalysis has a fraught relationship with its 'applied' form; one can imagine the crude psychoanalytic interpretations of individuals or even of whole societies in terms of 'diagnosis', perhaps by applying 'narcissism' to explain why people might prefer to speak to a non-human computerised avatar. Instead, a psychoanalytic methodology comprises a critical form of psychological intervention in which assumptions about the nature of 'the mind' are not taken in advance. Instead, theorising about the mind involves inductively theorising from the words and behaviours of people. This means that a focus on the 'conditions of possibility' regulates the psychoanalytic approach. One such condition that this project relies on is the social determination of the individual. This does not quite refer to the specific social settings within which various individuals are ensconced, but rather the fact that all individuals are socialised in one way or another. In proposing the 'always already' social constitution of the individual, Freud problematised the very distinction between the 'individual' and the 'social'. This project seeks to consider users as formed by and through their interactions with ReMind; as interpellated in terms of their attitudes, needs and expectations.

Technocracy

An important 'condition of possibility' to consider in ReMind's development is the economic context. My reason for including this aspect is because of a suspicion that external commercial pressures such as competition might influence the company's design-decisions. These design-decisions would then go on to influence how users experience the app.

> What is the commercial context in which ReMind operates as a company, and in which they deploy their app? In other words: are there economic demands which ReMind must respond to, and if so what are those responses?

My approach to considering ReMind's commercial context involves analysis of my participant observation experience combined with analysis of the mechanism of economic value and commercial competition. My fieldwork experience helped to gain insight into the expectations and aspirations of the ReMind company, and into their discussions about the mental health app market. The company had, during the time of the fieldwork, moved beyond the scrappy tech start-up phase and was in the process of integrating itself into a more complex commercial web of finance and industry. In terms of finance, ReMind had got to the point of convincing a consortium of technology investors to bet on ReMind as a potential generator of financial dividends. In terms of industry, ReMind had begun the process of manoeuvring themselves into acting as a mental health service provider to other companies and to state services. When I arrived in the virtual office of ReMind they were in a process of rationalising

their structure: most employees that I spoke to had entered the company on employment terms that had changed over time. For example, Charley, my primary contact, had started with ReMind as a clinical psychologist, but during my tenure Charley's role was as a research project lead. Of course, this is common for small companies as they expand, ReMind's two directors were explicit about providing a structure which enabled not just upward but also horizontal professional mobility. The basic sketch of how the company operates is thus: depending on expertise an employee is assigned to a team within the company, but that could and often did evolve over the course of employment. During my placement ReMind was split into teams of therapist advisors, psychology researchers, software engineers, and content designers. Within and across these teams were PR outreach workers, managers and consultants. The company was split between a number of field offices in three different continents,[163] and was in the process of expanding to more regions. The expansionary process of the company is a primary focus in considering the economic context: what demand are ReMind responding to by expanding? This is considered in terms of ReMind's internal dynamics and their external relationships with other companies or public utilities.

4.5 Conclusion

In my concluding chapter, 'Necessity is the Mother of Invention', I summarise my findings. This is done as a 'demonstration' rather than a summative list. This means that the internal logics which I identify as underpinning the workings of ReMind are put to use to generate a speculative chatbot with features that encapsulate the logics exposed throughout the thesis, but which are creatively manipulated. This involves a dialectical inversion in which it is hoped that the logics which have been identified through the project can be observed, this time from new perspectives. In the manner of Marx's reformulation of Hegel's logic, the logic itself remains intact, but the perspective from which that logic is considered is shifted. In my case this involves three methods: reversing, halting or extending those logics. The reason for this is to emphasise the critical nature of the project, which does not mean fault-finding or normative judgement but rather immanent critique. In attempting not to assume an external 'meta-position' in regard to one's object, a theory can be extrapolated from the object. This means that technology, for example, is not considered as having some prior 'essence', but is socially and historically produced. These social and historical logics are approached in terms of the logics themselves, and in so doing, it is hoped that new perspectives on the object, which is a product of those logics, can be assumed. The aim of this project is not to discover inconsistencies or problems in order to present them as proof of ReMind being ill-conceived or compromised, but instead, to discuss how ReMind is emblematic of wider social trends.

[163] Details of which cannot be provided as this would identify the company.

Chapter Five: Digitisation

Information is information, not matter or energy. No materialism which does not admit this can survive at the present day.[164]

Introduction

The self-help activities which the bot provides will be focused on in this chapter. CBT, as a modular 'umbrella' treatment which includes Mindfulness techniques (and to a lesser extent, life-coaching advice) provides the basis from which ReMind draws in developing the various self-help activities which make up the automated intervention. This chapter has two primary aims. One is to demonstrate that the conceptual basis of CBT is the basis for its automation in computerised form. This basis, following from the history chapter, asserts sometimes explicitly, and sometimes implicitly, that the mind is a thinking machine, or more precisely, a computer. The conceptual lineage of CBT and Mindfulness will be drawn on to show that 'computerised therapy' is possible due to these interventions, in their contemporary modular form, comprise an already technological and algorithmic form of treatment. The second aim of this chapter is to infer a philosophy of mind implicit in the workings of the ReMind app. ReMind (both the app and the company) does not espouse an explicit theory of what constitutes 'the mind', in that a basis for or causes of 'thought' are not strongly conceptualised. The argument in this chapter is that ReMind implicitly asserts the mind-as-computer thesis; this is done through analysis of the various self-help techniques that the app provides, investigation of their avowed cognitive and behavioural treatment lineage and a discussion about how 'cognition' has come to be compared with the operations of computers. An argument will be made that through computerisation, CBT achieves an ideal (or idealised) form: not by standardising the therapeutic practice but by projecting a particular role onto the user, as will be discussed in the final part of the chapter the projected role is that of a 'scientist'.

1. Mind-as-Computer - This section will discuss the connections between ReMind's self-help methods and cognitive scientific theories of mind. Three fundamental properties of the mind on which ReMind's intervention depends will be explored: that the mind is *functional*, that it is *algorithmic*, and that it is *digital*. The 'mind-as-computer' will be explored as a concept which frames cognition as a function of the brain, analogous to how software operates as a function of computational hardware. What this means is that, for ReMind, 'thought' is approached as a means to achieving the end of improved mental health. In short: "The way you think affects the way you feel."[165] Thought is seen as separate from its biological basis in the brain (or any other part of the body), not through any explicit philosophy which renders this separation as a fact but through a functionalist, or instrumental, understanding of thought as being an algorithmic object which can be externally manipulated, or 'reprogrammed'. This conceptual separation of mind from the brain, or of thought from its material basis occurs not just as a philosophical speculative exercise, but as a consequence

[164] Wiener, N. (1941) *Cybernetics: Or Control and Communication in the Animal and the Machine.* USA: MIT Press. p.132

[165] Clark, D.A & Beck, A.T. (2010) *Cognitive Therapy of Anxiety Disorders*. USA/UK: The Guilford Press

of the material and social processes within which ReMind is situated.

2. Information Processing - In treating thinking as a function of the mind, ReMind, through its intervention style, treats 'thought' as equal to 'information'. This is observable, for example, in the way that it encourages users to take a perspective on their own mind in which it can be 'reprogrammed'. This section begins with a discussion on information theory in which the process of thinking, and consequently of treating mental health, is equated to a non-meaningful, or non-semantic, transfer of information. 'Information' is understood here as non-semantic material structure: language as separated from its meaningful content. In dealing with one's thoughts as 'information' the user can assume the status as 'operator of oneself'.

3. A Scientist of My Own Mind - This section will explore the paradoxical perspective one must take in order to treat one's own mental health, as abstracted or removed from oneself. Due to the conceptual convergence of physical and mental health as manipulable objects which can be maintained, the mind can be considered as an object which, through practical experiments, can be controlled from an external Archimedean position. As discussed in chapter three, users must enter into an un-relational relationality to successfully engage with the ReMind bot. We can also see that this un-relationality seeps into a wider context in which mental health is discussed as a non-meaningful (or non-semantic) phenomenon. In other words, mental health is seen as something which is divorced from social context, just as in the mind's eye of an idealised scientist, the natural world is divorced from any contextual clutter and can be approached from a desubjectivised vantage point.

5.1 Mind-as-Computer

Functional

The ReMind app has two distinct aspects: the bot which the user converses with and the various self-help activities that the user has access to. ReMind's aim is to integrate these two aspects as much as possible: the bot offers the activities throughout the conversation, which are delivered as part of the conversation, but the conversation always ends up with the bot offering one of the activities which are sometimes integrated in the conversation or as choosable options offered by the bot. For example, the bot might walk the user through a CBT activity via the conversation, or, the user can also independently choose one of the activities which are listed divided into a set of modules. Some of the activities are offered as audio clips, and there are a small number of video clips and (less often) links to external mental health guidance websites. ReMind provides 31 self-help modules, with titles such as 'Manage Anger', or 'For Trauma'. These modules each then contain a list of between 6 and 13 activities. The modal number of activities across the modules is 8. The number of unique activities in all modules for the free version of the app is 19, and the paid version offers 59 unique activities. I have categorised these activities as falling under 5 classes: 1. Cognitive Techniques, 2. Mindfulness/Meditation, 3. Physical Exercise/Activity, 4. Life Coaching, 5. Externalisation/List-making.[166] Each of these 5 classes contains a median of 12 of the 59 unique activities. Most of the activities that the bot offers are techniques for the user to

[166] The app also offers a small number of 'sleep sounds' and 'audio stories', while important aspects of the app's provision, my focus is on the 'therapy' aspects of the app

assert control of their feelings or cognitions, or body. The various modules that ReMind provides, some of which contain the same activities, are directed towards different goals. For example, the 'Manage Anger' module contains some of the same activities as the 'For Pregnancy' module (Reframing Thoughts, Meditation, and Exercise). This does not mean that these activities are inappropriately assigned, but that there is a sense of interoperability: each module is a 'package' within which activities can be switched out from one module to another. There is a 'drag and drop', or 'mix-and-match' sensibility in which various activities are curatorially assigned to the modules which act as packages with which the user can treat specific aspects of their lives which affect their mental health. Most activities within specifically titled modules such as 'For Students' are not directly associated with the module's title. For example; 'For Students' contains nine activities, all of which, apart from one, offers indirect support such as breathing techniques. One activity titled 'Manage Academic Fears', then brings the user to an anxiety visualisation activity which is common to many of the other modules. The generic 'content' sense of the activities combined with the purported specificity of the modules is indicative of the instrumental attitude that ReMind has towards mental health: the end-goal of reducing mental health problems subordinates the methods for achieving that goal.

When I spoke to Arnold, ReMind's product director, he spoke about how the bot and app as a whole are instrumental to their operational goals. It may well be that at some point in time the company will produce entirely different products, as long as they are directed towards their aim of 'solving for mental health'.[167] Mental health in this sense is not only a problem to be solved but is one which can be solved through various different means. This approach is one of the consequences of a functionalist ethos to the workings of the mind which characterises electronic computation. This ethos equates the mind to a computer in that it is not relevant by which mechanism it produces its output, rather that it correctly produces the anticipated output. We usually think of computers as the electro-mechanical devices that sit on desks, fold up in backpacks, and sit in our pockets. A computer is however anything that computes, from an abacus to an astrolabe,[168] to a human.[169] John Johnston explains that a computer is an 'abstract machine':

> Although today's desktop computers are usually made of silicon and copper wire encased in plastic and metal, in principle they could be constructed out of a wide variety of materials. As abstract machines, their functions are not defined by the specific behaviour of the materials from which they are constructed; rather, this behaviour is used to physically instantiate a symbol system with its own independent rules or syntax.[170]

A computer is defined by its function rather than by its appearance, meaning that what is important is what a computer *does* (computation or symbol manipulation) rather than what a

[167] 'Solving for' as opposed to 'solving', and the experimental approach associated with the term will be discussed in chapter 7: 'Macro-Treatment'
[168] Dewji, N. (2017) 'Astrolabe – The first Personal Computer'. Online: https://ismailimail.blog/2017/05/11/astrolabe-the-first-personal-computer (Last accessed 03/09/22)
[169] Thompson, C. (2019) 'The Gendered History of Human Computers.' *Smithsonian Magazine.* Online: https://www.smithsonianmag.com/science-nature/history-human-computers-180972202 (Last accessed (06/09/22)
[170] Johnston, J. (2008) *The Allure of Machinic Life*. USA: MIT Press. p.71

computer *is* (various electronic and mechanical components). In this sense, the mind can be said to exhibit an attribute which is also exhibited by computers. This attribute is not a quality, material or a mechanism but simply a *function*. This means that it does not matter what underlying process is involved, or by what mechanism the functions are manifested, all that is important is that the same outcome is achieved. In other words, if a computer is programmed to deliver the result of '161', when given the sum '100 + 61', we can say that it is doing 'the same thing' as when a human delivers the same result. Because of this functional similarity, cybernetics and artificial intelligence researchers have posited that the mind is in some way computational. In this manner, the activities that ReMind offers approach the mind as an abstract machine, the specific quality of operations of which is subordinated to the 'output': that of improved mental health. Cognitive and behavioural therapy is directed in a similar manner to Arnold's output-based, instrumentalist attitude: the focus of cognitive therapy is on correcting or repairing malfunctioning cognitions and improved mental health is ultimately equated with correctly functioning thoughts. Similarly, for behavioural therapy, adjusting one's functional behaviour is equated with treatment success: the underlying meaning of one's behaviour is irrelevant. For CBT, it does not matter what thoughts *are* just that they can be manipulated towards improving their function. ReMind's functionalist aim, which is to leverage computational means to improve mental health *in a general sense*, as opposed to developing a specific form of treatment, is reflected in the functionalist ethos which underpins CBT, and the technical methods which ReMind has chosen to deliver their intervention. Bear in mind that the functionalist aim does not depend on or require a computational basis to be achieved, but rather that computation structurally mirrors this aim in a formal sense, in terms of producing the expected 'output': for ReMind, the mental health ends justify the computational means.

Algorithmic

In thinking about computers as abstract machines, defined by function, computerised treatment can be thought of as not precisely treatment performed by computers, but treatment that depends on a digital, mathematical, and procedural (algorithmic) basis. CBT, in this sense, is also a computational form of therapy. Cognitive and behavioural therapy are, in theoretical terms, strictly algorithmic forms of treatment - not dissimilar to other forms of treatment which are characterised by predefined steps (such as prescribing a specific medication to achieve a change in chemical conditions). However, CBT proceduralises treatment not just through a set of external algorithms which determine the course of treatment, but by approaching the mind itself as an essentially algorithmic entity. The flowchart pictured below is not dissimilar to other treatment step-by-step sequences, except that it does not chart the procedural sequence of treatment, but the procedural sequence of a mental process:

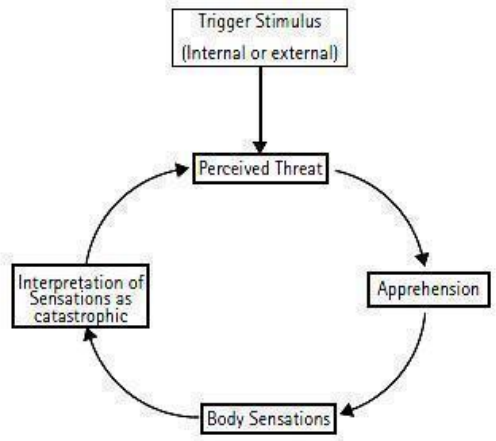
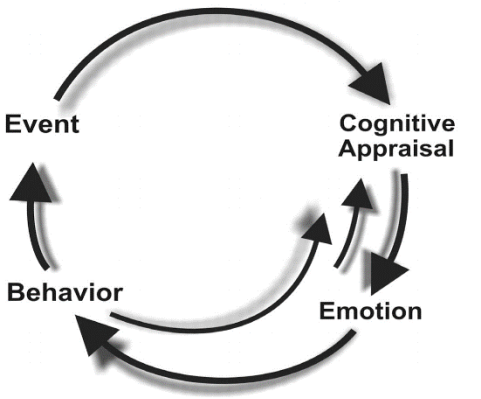


Figure 1. Basic Cognitive Behavior Model



The left-hand diagram[171] is from 1986, and while cognitive and behavioural treatment has changed since then (mostly to incorporate more techniques), the causal mechanism presented has not. The right-hand diagram[172] presents the fundamentally procedural and algorithmic quality of the CBT understanding of mental processes, but adds the contemporary addition: reversed 'feedback' arrows. The therapeutic technique that CBT and other cognitive therapies provide is a procedural means for the patient to identify problematic modes of thinking - "cognitive distortions"[173] - about certain life-situations in order to change the underlying beliefs. For example, in the case of panic, "The CBT model separates out a stimulus, a perception of it as threatening, a state of apprehension, and bodily sensations, and places these separated phenomena into a causal sequence."[174] This causality forms the basis for cognitive and behavioural therapeutic models. A psychological model which treats psychic phenomena as procedural and isolable and ultimately, algorithmic, units can make a case for treatment to also follow this sequence. While most recent neurobiological research has come to discredit the claim that human consciousness is akin to software running on the hardware of the brain, this dichotomy has come to haunt theories of psychology and artificial intelligence ever since. Jacques Lacan claims that developments in cybernetics show that it is possible for a symbolic structure to proceed in an autonomous fashion, which he compared to the workings of the unconscious. Lacan also claims that cybernetics shows that there are also conscious processes which occur in a machinic manner - mathematical calculations, or any rule-based (algorithmic) process. When we perform a cognitive operation which follows a strict procedure, we are not 'thinking' according to Jacques Lacan, we are simply running a predefined 'program':

> We are very well aware that this machine doesn't think. We made the machine, and it thinks what it had been told to think. But if the machine doesn't think, it is obvious that we don't think either when we are performing an operation. We follow the very same procedures as the machine. The important thing here is to realise that the

[171] Image source: Clark, DM. (1986) 'A cognitive approach to panic'. *Behaviour Research and Therapy*. 24 (4), pp.461–470. Elsevier
[172] Image source: Wright, J.H. Basco, M.R. & Thase, M.E (2006) *Learning Cognitive-Behavior Therapy: An Illustrated Guide*. USA: American Psychiatric Publishing. p.5
[173] Beck, J.S. (2020) *Cognitive Behaviour Therapy, Basics and Beyond.* USA: Guildford Press. P. 179
[174] Gipps, R.G.T. (2013) 'Cognitive Behaviour Therapy: A Philosophical Appraisal'. In: Fulford, K.W.M. et al. (eds.) *Oxford Handbook of Philosophy and Psychiatry*. UK: Oxford University press. pp.1245-1263. p.1247

> chain of possible combinations of the encounter can be studied as such, as an order which subsists in its rigour, independently of all subjectivity.[175]

The term 'computer' initially referred to an occupation in which humans were employed: prior to Turing's invention of automated computation, calculating numbers for military, industrial or commercial purposes was performed by humans.[176] Turing explained the role of a computer as someone who is "supposed to be following fixed rules; he has no authority to deviate from them in any detail."[177] This means that the work of computing (mathematical calculation) involves no subjective choice: the human provides the ability to calculate but does not influence the logical structure of calculation in any way. The CBT model of how to intervene in one's own cognitions follows an identical procedure. This procedure treats cognitions as algorithmic 'programs', which "separates out a stimulus, a perception of it as threatening, a state of apprehension, and bodily sensations, and places these separated phenomena into a causal sequence."[178] This algorithmic causality forms the basis for cognitive therapeutic models. This model is based on the definition of a mind as a system which processes environmental inputs in order to convert them into behavioural outputs, in other words, whether the brain 'is' a machine or not, it functions, when performing certain tasks, in an identical way to that of a machine. ReMind's therapeutic technique, in line with cognitive style treatments, involves a procedural means for the patient to identify problematic modes of thinking - "cognitive distortions"[179] - about certain life-situations in order to change the underlying beliefs. CBT, through reducing clinical technique to an absolute formal procedure encourages an explicitly scientific approach to the mind - as a scientific object which is observable not just to the therapist, but also to the patient:

> Typical CBT models expound something like the following: (a) that the first, rather preliminary, step of cognitive therapy is to help the patient clearly identify their emotionally problematic core beliefs, rules and assumptions. And (b) that the second task is to encourage them to quasi-scientifically test out these assumptions, either through rational engagement leading to what is sometimes called "cognitive restructuring", or more practically through "behavioral experiments."[180]

Through its procedural form of treatment, CBT proceduralises a concept of consciousness in which thought takes on an algorithmic sense of input-process-output. Cognitive science, which aims to understand the mind as a computational machine, might show us that there are indeed computational or mechanical processes involved 'in the mind'. However, in taking the theoretic leap of extrapolating from cognitive processes or modules to comprising the entirety of consciousness, a model of mental health and illness emerges which is

---

[175] Lacan, J. (1991) *The Seminar of Jacques Lacan, Book 2: The Ego in Freud's Theory and in the Technique of Psychoanalysis*. UK: W.W. Norton and Co. p.304
[176] Hayles, N.K (2005) *My Mother Was a Computer: Digital Subjects and Literary Texts.* USA: University of Chicago Press. p.1
[177] Turing, A. (1950) 'Computing machinery and intelligence.' *Mind*. 59 (236) pp.433-460
[178] Gipps, R.G.T. (2013) 'Cognitive Behaviour Therapy: A Philosophical Appraisal'. In: Fulford, K.W.M. et al. (eds.) *Oxford Handbook of Philosophy and Psychiatry*. UK: Oxford University press. pp.1245-1263. p.1247
[179] Beck, J.S. (2020) *Cognitive Behaviour Therapy, Basics and Beyond.* USA: Guildford Press. p.179
[180] Gipps, R.G.T. (2013) 'Cognitive Behaviour Therapy: A Philosophical Appraisal'. In: Fulford, K.W.M. et al. (eds.) *Oxford Handbook of Philosophy and Psychiatry*. UK: Oxford University press. pp.1245-1263. p.1257

correspondingly computational. While benefiting from this model in terms of identifying the causation of mental effects from physical events such as brain injuries, developmental defects or other biophysical conditions, a cognitive approach to mental health treatment such as CBT takes the computational or procedural features of mental processes originally proposed as the *means* to simulating mental contents rather than the actual contents themselves and reapplies them as a global image of consciousness. We can understand that the job of 'computing', in its journey from the human operator, to the machine, and back to the human again, has having undergone an ontological reframing: initially as a functional ability of the mind, to the operational basis of digital computers, and back to the mind, but now not just as the operational basis or functional ability, but as the *definitional* basis for the mind. In other words, 'computation' defines what the mind *is*, and not just what it can *do*.

Digital

Many of ReMind's modules involve visualisation techniques, or some form of conceptual materialising of one's distress; these techniques are drawn from Mindfulness. For example, 'Manage Anxiety' is introduced with "Giving a physical form to your anxiety can help your mind feel more in control of it. In this exercise, we will visualise what anxiety looks and feels like to you, and learn to reduce its intensity." Note the use of the term 'mind', which is invoked as separate to and under the guidance of 'you'; in turn, anxiety is potentially under the control (or vice versa) of the mind. The activity goes on to help the user to conceptually visualise anxiety in terms of colour, shape and temperature, to conceptually manifest the anxiety as an objective presence. This objectification is especially apparent in ReMind's Mindfulness-influenced activities in which 'feelings' come to be rendered as 'symbols' which the user can potentially manipulate in terms of size, mass, colour, temperature, etc. The user is encouraged to consider their emotions and feelings as thoughts in terms of 'symbolic objects': i.e. not just to isolate their feelings as linguistic concepts (e.g. "My feeling of sadness is like a black cloud") but to imagine these concepts as materially manipulable. The user is encouraged to imagine this object shrinking, changing colour, and moving further away from the user. The 'Manage Anxiety' activity ends with, "With practice you will be able to control your anxiety and feel relaxed sooner." 'The mind' in this case, is a container of thoughts, which are one step removed from the users who are tasked with transforming 'their' thoughts. By disciplining oneself into treating one's mental health through this kind of conceptual activity, 'feelings' come to be rendered as external to the 'mind': as symbolic objects which are potentially externally manipulable. Turing's invention of the computer relies on three primary features. As discussed above, it is an abstract machine; secondly, it operates in terms of algorithms; and thirdly it processes symbols. For the computer, a symbol is an object (in Turing's case it was a printed character) which can be moved, erased or combined with other symbols to perform its computational functions. As Johnson notes above, the purpose of a computer is "is used to physically instantiate a symbol system with its own independent rules or syntax".[181]

In semiotics the most basic definition for a symbol ('sign') is that it involves a reduction from continuous to discrete, from analogue to digital: the elemental act in the creation of signifying structures involves asserting binary distinctions (e.g. me/you). This means that symbols comprise an inherent 'digital' aspect. Turing's achievement was to transfer the signifying

[181] Johnston, J. (2008) *The Allure of Machinic Life*. USA: MIT Press. p.71

procedure from a mental activity to a mechanical one. Turing's invention involved a physical instantiation of discrete signs in terms of symbolic objects. The difference between Turing's symbolic objects and printed text is that computational objects are manipulable in real time and have the capacity to influence their own manipulation in the form of instructing the computer how to proceed. Turing referred to his invention as a "discrete state machine"[182] meaning that it can be thought of as inhabiting, at any one moment, a measurable and specific - discrete - state. Computation depends on the strict demarcation of symbolic units: the machine operates in a digital (as opposed to analogue) format - that, like in the case of an abacus or astrolabe, the movements of the device are perceivable as discrete rather than continuous. In Turing's terms, a symbol is a digitally produced object in that, whether by mechanical or electronic means, it is strictly demarcated from other symbols. The ones and zeros which define contemporary electronic computation are 'digital' insofar as they are not simply 'numbers', but that they refer to discrete states: they are strictly defined against one another. The calculation of numbers, on the most basic level, is possible not because a computer operates in terms of countable units but in terms of the binary definition of zero and one, which correspond to an electronic circuit being either on or off. Symbolisation and digitalisation are two sides of the same conceptual coin, with one depending on the other: symbolisation depends on a digital ability to strictly demarcate, and digitalisation depends on the metaphoric linguistic ability to symbolise. By encouraging the user to approach their feelings as either algorithmic procedure (cognitive therapy) or as symbolic objects (Mindfulness), we can understand ReMind's intervention as 'computational'. This is not strictly due to the computational basis on which the intervention is delivered, but the conceptual basis on which CBT and Mindfulness has developed. In this sense, most of ReMind's self-help activities operate on the assumption that the mind is indeed a calculation machine. 'The mind' in this case, is an object which is one step removed from the users who are tasked with reprogramming 'their' minds.

5.2 Information Processing

Information Theory

Aaron Beck states that the cognitive approach to therapy "is best-viewed as the application of the cognitive model of a particular disorder with the use of a variety of techniques designed to modify dysfunctional beliefs and faulty information processing characteristic of each disorder."[183] But what does 'information' mean in terms of the cognitive therapy intervention that ReMind provides? To compare the operations of the mind as symbol manipulation in a similar manner to the operations of a computer program a particular perspective must be assumed in which the split between hardware and software is conflated with matter and psychology. In order to assume this perspective, one must make the prior assumption that 'information', in the form of the kinds of inputs that a computer is fed, is identical to the sensations that the human nervous system feeds to the brain. From this standpoint consciousness can be conceptualised as an input-output machine which deploys different algorithms to convert environmental inputs into cognitive or behavioural outputs. This is clearly observable in the workings of computerised robots which can manipulate

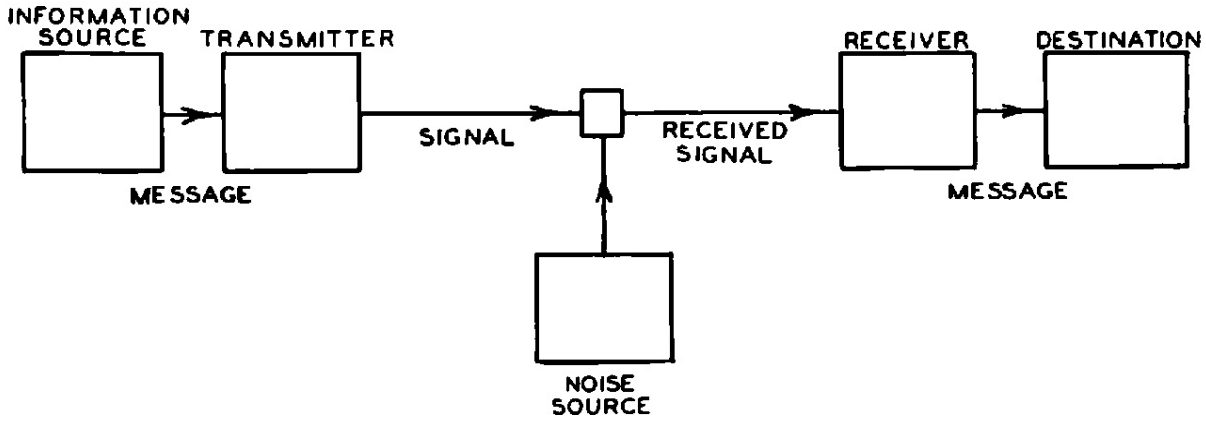
[182] Turing, A. (1950) 'Computing machinery and intelligence.' *Mind*, 59, 433-460. p.439
[183] Beck A.T. (1993) 'Cognitive therapy: past, present, and future'. *J Consult Clin Psychol*. 1993;61:194–8. p.194

perceptual data, make decisions and then effect actions based on those decisions. It 'looks like' computerised robots perform tasks in a similar way to humans, albeit in a much less sophisticated manner. Based on this comparison, great leaps have been made in robots and AI programs which simulate human perception, decision making, problem solving, pattern recognition, speech, etc. It is necessary to first confine oneself to the assumption that, as stated below, all we have access to is 'information', in order to then justifiably claim that the brain is a computer. Commercial neuroscience technologist Giulio Ruffini succinctly illustrates the comparison:

> If all that brains have access to is information, we can naturally think of brains as "information processing machines"—computers in the mathematical sense (Turing Machines)—and questions about our experience of reality should be considered within the context of algorithmic information theory. Our "Input/Output streams (I/Os)" include information collected from visual, auditory, proprioceptive and other sensory systems, and outputs in the form of peripheral nervous system mediated information streams to generate actions affecting the body (e.g. via the autonomic system) or the external world (e.g. body movements or speech). We will use the term "cognition" here to refer to the process of model building and model-driven interaction with the external world.[184]

In an information processing theory of consciousness, while algorithms are substantialised, phenomenological experience and meaning is *de*-substantialised. This dynamic is a consequence of the separation of semantics from syntax which occurs when information is transmitted. Lydia Liu explains how the development of communications technology in the mid-20th century involved solving engineering problems to transfer messages more efficiently from transmitter to receiver. To solve these problems, language had to be reduced to syntax and structure; semantics - or meaning - had no place in problems of noise reduction, efficiency and entropy in the building of Morse code, and subsequent telephonic infrastructure. The development of 'information theory' came from the convergence of cybernetics and the "Mathematical Theory of Communication": a 1948 paper by Claude Shannon in which the formal conditions for calculating precision in the automated transmission of a message were laid out. Shannon saw that the content, or meaning, of the message was irrelevant, instead, what was important was to understand the likelihood of errors (caused by 'noise') that a sent message would undergo on its journey to its destination:

[184] Ruffini, G. (2017) 'An algorithmic information theory of consciousness', *Neuroscience of Consciousness*, Volume 2017, Issue 1



185

By doing away with any consideration of the content of messages, Shannon recast the problem of technological communication to that of entropy, efficiency and probability. Shannon understood that meaning is of course a necessary component of a message, but not for the *transmission* of a message:

> The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have *meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem.[186]

ReMind's mental health activities involve a similar splitting in terms of their procedural and algorithmic basis: for example, the CBT activity 'Reframing Thoughts' involves identifying negative thoughts with the aim of asserting external control over those thoughts and eventually replacing those thoughts with other, positive thoughts. The 'content' or meaning of the thought is made redundant and subordinated to the procedure: focusing on meaning might involve delving into the reasons why one might have such recurring negative thoughts. 'Reframing Thoughts' does not assert that thoughts are meaningless, but rather that 'meaning' can be conceptually separated from the 'thought', which can now be considered as an externally manipulable object. Charley, head of clinical research and development, summed up the double-sided concept in terms of a split of mechanism from sentience:

> I mean, I've definitely thought of that question at a philosophical level at some point. But in some ways, are we all running algorithms in our mind, right? And if we're just running an algorithm for how we intervene with somebody then you're potentially a robot. So, yeah, I think, and again, now I'm speaking at a very philosophical level, that one, I think it's the capacity to execute moral and ethical responsibilities that make someone human. And the other aspect is really anything that classifies as sentience, right, within a human being or a robot is the capacity to step outside the rules that have been set out for you.

---

185 Image source: "Schematic diagram of a general communication system" in Shannon. C. (1948) 'A Mathematical Theory of Communication'. *The Bell Systems Technical journal*. p.379

186 Shannon. C. (1948) 'A Mathematical Theory of Communication'. *The Bell Systems Technical journal.* p.379

We can understand this as a common view: on a fundamental level, we operate in robotic ways, 'running algorithms', but what differentiates humans from robots is our ability to defy our own procedures. Whether this analogy holds true for either the operations of a computer or the mind is irrelevant; it allows us to conceptualise thought in a particular way which excludes (but doesn't abandon) the subjective experience of consciousness - the 'what it is like' to have thoughts, feelings and sensations are suspended in favour of the externality of objective visualisation. The visible surges of electricity across networks of neurons is uncontroversially equated with thoughts, sensations, feelings and emotions. This non-meaningful information is the objective side of the subjectivity coin - what one experiences subjectively can be observed using neuro-imaging devices. Because this equation is so indisputable, it is unsurprising that claims are often made that we are on the brink of solving the 'hard problem of consciousness'.[187]

Non-Meaning

The makers of therapy chatbots describe their intervention as a form of self-treatment, meaning that the bot facilitates the user in treating their own mental health. In order to do this, the intervention which is delivered by the bot undergoes 'informationisation': it is transformed into information. This means that the intervention is rendered as a step-by-step, procedural technique which is amenable to transfer, i.e. it can undergo transmission from a source to a destination. Initially this transfer is from the bot to the user, but the aim of the intervention (and of CBT and Mindfulness in general) is for the users to transfer the techniques from, and to, themselves. This is done through learning the various mental health activities - storing them for future retrieval. The user is the transmitter and receiver of the information with which they tend to their own mental health. Charley spoke about how ReMind's intervention, while it should be thought about as one among other options, is largely based on transmitting some kind of technique to the user. From my own assessment of what the bot offers, apart from the chat facility in which the user speaks to the bot and the 'sleep sounds' and 'stories', all of the activities offered involve explaining a technique which the user learns and deploys when needed. Charley described 'skill' as that which the user can learn, practice and rely on in times of need:

> I would visualise maybe what I'm seeing as a pie chart with multiple components. So there is one, which I would say is a fairly large component, is the skill building aspect…So the insight, the catharsis, the reaching resolution, all of those help us find the skill or what finally gets solidified as a skill rather, that we're then reutilizing every time there is recurrence in our life.

Insight, catharsis and resolution, by undergoing solidification as a skill, become formalised into a routine or procedure which can be reutilised: as a program. This has the effect of transforming mental health therapy precisely into a 'program' which is storied, retrieved and executed when needed. Recall Lacan's claim that "machines don't think".[188] The modules

[187] Seth, A. (2021) 'The hard problem of consciousness is already beginning to dissolve.' Online: https://www.newscientist.com/article/mg25133501-500-the-hard-problem-of-consciousness-is-already-beginning-to-dissolve (Last Accessed 21/8/21)
[188] Lacan, J. (1991) *The Seminar of Jacques Lacan, Book 2: The Ego in Freud's Theory and in the Technique of Psychoanalysis*. p.304

that the ReMind app provides, which contains the various mental health activities, can be thought of as programs, in the precise sense of the term "as an order which subsists in its rigour, independently of all subjectivity."[189] In other words, in approaching mental health intervention as a transfer of information, 'meaning' is not excluded or rendered obsolete, but rather, is rendered as separable from the program of self-treatment. One performs the CBT or Mindfulness program 'on' oneself, or 'on' one's cognitions: 'self' or 'cognitions' are in this sense, devoid of meaning, which has not disappeared, but becomes inapplicable to the program of mental health intervention. Approaching the mind in terms of 'cognitions' which, on one hand, excludes subjective experience, and on the other hand, equates mental states with the material operations of the brain, has the effect of reducing the mind to a carrier of non-meaningful information. The has inaugurated a process of universalisation of cognitive functions - the mind being a manipulator of generic information. This equates to a separation of the subject from the brain - "my brain did this", "my brain thought these thoughts", etc. N. Katherine Hayles equates this non-subjective form of thinking with "nonconscious cognition".[190] According to Hayles, it is vital to grasp how the emergence of electronic computation has brought about a form of cognition that exhibits both a nonconscious and embodied form. Hayles claims that due to this new cognition, meaning itself has been transformed to include non-meaningful computational phenomena in which humans are 'out of the loop' of high-speed information processing and interpretation. This means that we have access to a demonstrable separation of two forms of cognition, one that is associated with the human subjective experience, and the other that is associated with physical process in which, as Hayles claims, "meaning has no meaning."[191] Margaret Boden's explanation of the cognitive scientific approach towards the mind-as-computer shows that computers are not taken to be the paragon definition for the mind, but rather that computers allow us to demonstrate and even instantiate mind-like activity:

> Cognitive scientists don't believe that today's computer-related concepts suffice to explain the mind. Rather, they believe that they're a good beginning, and that later explanations will use concepts drawn from what then happens to be the best theory of what computers do.[192]

Boden's claim is that the development of computers and scientific understanding of the mind will be interlinked, and that while computers are often used in cognitive scientific research, cognitive science ultimately involves theorising *in terms of computation*.[193] The practical effect of this process is a subject which is separable from itself, and observable to itself from the external vantage point of rational empiricism: "such a self is a prototype of a scientist-observer who is in the business of trying to control and predict the world by constructing inner representations or interpretations of it."[194] This effect also introduces an 'internality-externality' to the subject - in terms of CBT, the removal of cognitive distortions first requires

---

[189] Ibid. p.304
[190] Hayles, N.K. (2014) 'Cognition Everywhere: The Rise of the Cognitive Nonconscious and the Costs of Consciousness.' *New Literary History*, 2014, 45: pp.199–220
[191] Ibid. p.199
[192] Boden, M. (2000) *Mind as Machine. A History of Cognitive Science.* UK: OUP Oxford. p.12
[193] Ibid. p.13
[194] Gipps, R.G.T. (2013) 'Cognitive Behaviour Therapy: A Philosophical Appraisal'. In: Fulford, K.W.M. et al. (eds.) *Oxford Handbook of Philosophy and Psychiatry*. UK: Oxford University press. pp.1245-1263. p.1260

one to assume an impartial and objective 'scientific' attitude towards oneself, but this is in order to reach this self-same ideal subjectivity which is concealed by cognitive distortions. This subject is consequently rational, objective and autonomous, or has the potential for autonomy, once the barriers to autonomy - cognitive distortions - have been removed. Paradoxically though, this subject must initially assume this rational objectivity in order to remove cognitive distortions. If thoughts are software running on the hardware of the brain, they can be reprogrammed, this is the central concept of CBT. The effect of this is to situate the human subject as operator of itself.

Homunculus

This 'operator' is analogous to the operator in John Searle's 'Chinese Room' thought experiment. The argument sets out with a mental picture of a room with a person in it; this person is cut off from the outside world apart from a series of notes or cards that are passed to them through a slot. These notes have Chinese symbols on them. The person does not understand Chinese (the thought experiment is dependent on the person not knowing the language being used; Searle used Chinese because of its logographic dissimilarity to alphabetical languages) but has a comprehensive list of all the correct responses to the notes being passed to them through the slot. The person then uses this list to pass back their own set of notes which have the correct responses on them. The thought experiment goes on to propose that the people outside the room have a comprehensive grasp of Chinese, and could be completely convinced that the person in the room also does, because of their correct responses. However, because these responses are due to reference to a set list, there is no need to understand anything that is being passed to them and then passed back out. Searle likens this process to the way that computer operates a program: it is given information in the form of a computer program (the Chinese notes passed through the slot) and it then generates a series of outputs thanks to its own operating system (the comprehensive list used by the person in the room), and then represents these outputs on a computer screen (the notes passed back out the slot). Because there is no understanding involved on the part of the person in the room or of the computer, Searle claims that a computer program is not a sufficient basis for having a mind. Searle's reasoning for this is in the difference between syntax and semantics. Syntax is described as the set rules for successful operating of a language, whether this is a human spoken language or a computer program. Syntax is operative in the rules governing how words are spelled to the strict ordering of numbers in a list, to grammatical rules. Semantics is defined as the meaning of words - what is represented to a person's mind when particular words and phrases are thought about. Searle claims that both are needed in order to possess a 'mind'. Syntax, or the rules governing which Chinese notes are passed back through the slot, is not enough for a mind to exist, an understanding of the semantics is also necessary. This would equate to the person inside the Chinese Room being a Chinese speaker, and thus having both syntax and semantics. Searle claims that this argument proves that it is impossible for a computer, simply by running a program, no matter how sophisticated, to experience 'what it is like' to have a mind in the same way people do.[195] The separation of syntax from semantics is

[195] Searle's arguments have undergone extensive critique (see *Views into the Chinese Room: New Essays on Searle and Artificial Intelligence* UK: Clarendon Press), the purpose of this discussion is not to side with or against Searle's argument, but to demonstrate the split between syntax and

similar to Hayles' identification of non-conscious cognition, which is devoid of meaning and operations in terms of non-sematic, 'syntactical' programs. As Searle notes, implementation of a program, such as for example, the 'Manage Anger' activity that ReMind provides, may involve - but does not depend on - reflexive self-understanding. The user implements the activity, with the aim of reaching a particular goal. This goal might be the staving off of a panic attack, or of feeling less lonely, etc, which may well involve subjective meanings (semantics), but they are rendered as outcomes via the algorithmic (syntactical) process of achieving them.

The operator of the Chinese Room does not need to understand the meaning of the activities being performed in the processing of information. From the perspective of the external observers the operator of the Chinese Room does in fact understand the information being processed. The user of ReMind, in this way, acts as both the information processor and external observer: undertaking the mental health activities and assessing the results. By undertaking to perform the programmatic activities which ReMind provides, the user must assume a perspective towards the meaning of their feelings in which those feelings might indeed include meaning in some way, but meaning is not a necessary component in the treatment of one's mental health. This is the precise perspective that is undertaken in both CBT and Mindfulness. The impartial observer/experimenter of cognitive therapy, behavioural therapy and in Mindfulness comprises the element which links all three forms of treatment. Mindfulness differs from behavioural and cognitive therapy in that one is not encouraged to make any alterations to one's behaviour or thoughts: impartiality is maintained not just as an attitude but as a goal. Throughout the development of behavioural therapy, with the rhetorical and methodological transformation wrought by cognitive psychology, to the modularity and secularisation achieved by Mindfulness, we can see the practices and epistemological basis of behavioural psychology undergoing a subtle transformation. This transformation has not eliminated behavioural therapy's objectivist approach to the human animal, but has refined it from a crude conceptualisation of the human as a processor of environmental cues to the human subject as operator of its mind. The operator, or 'self', assumes not just an externalising perspective on their social environment (as a processor or environmental inputs) but also an external (or rather, externalising) perspective on oneself. The rise of CBT and Mindfulness can be seen as attempts to create objective forms of treatment, which means that the question of the subject remains ignored. Out of this process a theory of the subject - one which is curiously desubjectified - can be derived. In both CBT and Mindfulness this subject is a homunculus: the pilot - or programmer - of one's self. It is both internal and external: internal in that this pilot is encapsulated within the body, and external in that this pilot is detached from both its own thoughts and from 'thought' itself. This subject floats in a void of its own making.

---

semantics as an epistemological and practical consequence of the invention of automated computation.

5.3 A Scientist of My Own Mind

The Scientific Attitude

One of the features that the ReMind app offers is a chronological list of all the modules that the user has engaged in, from downloading the app to present day. These are represented by short blurbs which summarise the activity performed in each module ('Checked my anxiety levels', 'Embraced my emotions'). The blurbs assert the success of each module whether one has completed it or not; one only needs to have started the module for it to qualify for the list. Summaries are also at times contextually confused due to their various purposes:

"Had a strong start to the day!" at 7pm

My own response to this was a sense of disconnection to activities that I have ostensibly performed: "This was me?" CBT is a generalised form of treatment in that it approaches the human subject as a generic entity - the 'blank slate' of behaviourism. This means that the 'individual' is theorised as, on one hand, radically individuated and socially disconnected, and on the other hand, radically deindividuated and socially generalised. In other words, the patient-subject is approached as an individual but, paradoxically, not unique in their individuation. This does not mean that the social is irrelevant, but that, due to its generalisability, the social is thought of as 'generic', or contextless background. This means that the subject of CBT is a 'scientist-observer', or György Lukács's passive observer, who is:

> …hopelessly trapped in the two extremes of crude empiricism and abstract utopianism. In the one case, consciousness becomes either a completely passive observer moving in obedience to laws which it can never control. In the other it regards itself as a power which is able of its own – subjective – volition to master the essentially meaningless motion of objects.[196]

This subject is at once a passive observer of itself, unable to challenge the laws which guide it, and simultaneously an omnipotently powerful manipulator of its own meaningless contents. This observer/manipulator is, as Gipps notes, analogous to a kind of scientist of oneself. CBT has the capability to induce a dissociative sensibility through its interventions. Like looking at a photograph of oneself as a child with an understanding that 'this was me', but no experiential sensation of this connection. This disengaged sense of being external to oneself is linked to the behavioural lineage which lays claim to CBT. Jacques Lacan claims that the modern scientific view which began to emerge in the 1600s, which developed into a set of highly delineated research programs, gradually abolished the inclusion of the subject. In other words, in considering natural phenomena as meaningless, arbitrary and without external agency - without a subject - an intervention can be made into reality which reflects this logic. Removing the subject allows for the internal dynamics of nature to be focused on and manipulated in a way that had previously been only glimpsed at. Lacan claims that this

[196] Lukács, G. (1991) *History & Class Consciousness*. UK: Merlin Press

removal is also the introduction of the "subject of science".[197]

The subject of science is a paradoxical concept, especially so when it is claimed to have emerged due to its own disappearance, but essentially the Lacanian claim is based on the assumption that the material presence of reality is the object of science, and that the observation of internal dynamics is the goal, rather than positing some kind of external cause. CBT presents an unusual twist on the notion of a 'scientific subject' however in that one must evade or eliminate the immediacy of one's own experience in order to assume a position from which one can intervene into such experience. By treating the patient as both socially isolable and as an external 'scientist-observer,' CBT effectuates a subject which is identical to and yet external to itself - it is defined according to its own positive properties rather than as a difference to other subjects. For this reason, a patient, being wholly subject to themself, is ultimately responsible for their own mental health and answerable only to themself. This internalised reflexivity, rather than a means to improving mental health, can itself be regarded as a symptom signalling the "privatisation of stress"[198] and the rise of inwardly directed mental health disorders in post-industrial society, such as depression, anxiety and eating disorders. The occlusion of the social and the individualising of the subject can in this way be regarded as not just having a very real effect on 'mental health,' but as part of a broader social tendency of which contemporary disorders are not aberrations but features of that tendency. This does not mean that CBT as a clinical treatment 'causes' the mental health disorders it ostensibly treats. But in maintaining the severe distinction between individual and social, and through directing the responsibility of mental healthcare onto the individual, the therapeutic effects of CBT are achieved at the cost of uncritically adopting those underlying conditions - the social background framing the individual appears as 'natural' and unchangeable, and CBT thus unintentionally propagates those conditions.

CBT's method of removal of cognitive distortions, by working on objectified "thoughts, beliefs and attitudes"[199] implies that they are barriers in the way of an authentic and pure subjectivity. That this subjectivity can be reached through overcoming or breaking down these barriers seems at first to be a logical consequence, and CBT provides various procedural, step-by-step guides to doing this. But the 'internally-external' subject, which can be manipulated through behavioural modification with the added backup of pharmacological intervention, is situated paradoxically internally and externally to the subject in a number of ways: 'within' the subject, as the perceived locus of thought from which behaviour is directed, and also as the ideal subject free of cognitive distortions; 'without,' as an objective and ideal form of cognition which must be assumed in order to eliminate cognitive distortions and as an external observer, directing thought from outside of itself. The subject is internal to itself and disconnected from the social background, and also external to itself and imposed onto the subject by the ideal form as represented by the therapist-coach, or in this case - chatbot.

---

[197] Glynos, G. (2002) 'Psychoanalysis operates upon the subject of science: Lacan between science and ethics'. In: Glynos, G. & Stavrakakis, Y. (eds.) *Lacan and science*. London: Karnac
[198] Fisher, M. (2009) *Capitalist realism: Is there no alternative?* UK: O Books. p.19
[199] 'What is CBT?' *Mind.* Online: https://www.mind.org.uk/information-support/drugs-and-treatments/talking-therapy-and-counselling/cognitive-behavioural-therapy-cbt (Last accessed 28/07/23)

The Object of Science

CBT is quickly becoming the most popular type of non-pharmaceutical treatment[200] for the more common contemporary mental health conditions - depression and anxiety.[201] They are also the only styles of treatment to have successfully been electronically automated. Why is this? One reason is because CBT treatments enable self-treatment: the learning of various techniques, as opposed to the dialogic form characteristic of talking-style therapies. Another is that CBT treatments approach the mind as displaying the properties conceptualised by cognitive and neuroscience, which are based on computational models of the mind. Most computerised mental health treatment is akin to elaborate 'self-help' guidance,[202] displaying little in the way of artificial intelligent processes, primarily because it is dangerous to implement genuine AI conversational programs into treatment chatbots as it is extremely difficult to control the utterances of a genuine AI chatbot.[203] However, concepts associated with AI influence the development of this new technology in a roundabout way - as the underpinnings of concepts of consciousness, cognition and mental health/illness which are the objects of new forms of computerised treatment. The proliferation of computerised mental health treatment does not logically depend on a mind-as-computer concept, but instead most forms of treatment have converged with this concept over the course of the 20th century due to the proceduralisation of diagnostic and treatment techniques. Both are predicated on the advances that computer power has undergone over the last 100 years. In order to conduct scientific research, the objectifiability of one's approach must be established. In other words, the results of one's research must be quantifiable, replicable, and conforming to the established epistemological framework of one's discipline. Psychological disciplines have always struggled with fitting into 'proper' science and have had to conform to rigid quantification regimes in order to establish and maintain legitimacy. Quantifiability is important to app-based treatment because mental health app companies need to display a solid methodological grounding in order to justify their existence, for funding drives and to convince potential users and customers that app-based treatment is a viable prospect. This means that ReMind must constitute its object (mental health) as quantifiable so that proof can be offered that the app can potentially have a successful effect. This amounts to conducting experimental 'tests' using patient outcome forms to produce statistical results which show improvement in the users' mental health. 'Scientificity' characterises app-based treatment: a scientific approach can be said to predominate which is directed towards the various methods that different scientific disciplines use (statistical measures, quantification, etc), but which approaches its object, 'mental health', in terms of a number of steps of removal. Jeff, one of ReMind's directors, outlined their scientific approach**:**

---

[200] David, D., Cristea, I., & Hofmann, S. G. (2018) 'Why Cognitive Behavioral Therapy Is the Current Gold Standard of Psychotherapy.' *Frontiers in psychiatry*, *9*, 4

[201] Mental Health Foundation (2016) 'Fundamental Facts About Mental Health 2016.' *Mental Health Foundation.* Online: https://www.mentalhealth.org.uk/publications/fundamental-facts-about-mental-health-2016 (Last accessed 21/9/21)

[202] Juneja, M. (2018) 'An interview with Jo Aggarwal: Building a safe chatbot for mental health.' *Maneesh Juneja.* Online: http://maneeshjuneja.com/blog/2018/12/12/an-interview-with-jo-aggarwal-building-a-safe-chatbot-for-mental-health (Last accessed 9/9/20)

[203] Daws, R. (2020) 'Medical chatbot using OpenAI's GPT-3 told a fake patient to kill themselves.' *AINews.* Online: https://artificialintelligence-news.com/2020/10/28/medical-chatbot-openai-gpt3-patient-kill-themselves (Last accessed 20/7/20)

> It's a problem solving approach, I guess, if you call engineering as iterative problem solving, I think all science is, in a way. So you start with a hypothesis, you run it, you see whether it's working or not, then you change, create another hypothesis, then you run it. So I would call it a scientific approach. I think clinicians do that, when they're treating somebody as well. Let's start with a hypothesis, see whether it works. So we just did it at a massive scale. So you would run a hypothesis on a million people and find out that it didn't work for that 10,000. And then change something for that 10,000, personalise it, and so on, so forth.

ReMind's approach to quantification relies partly on the numbers of users who engage with the app: they apply a 'scientific method' in that they perform experiments and observe the results. This sort of scientific approach also characterises the mental health techniques that the app provides. CBT involves an attitude of iteration and testing towards one's own mind. This experimental approach displays an appeal towards a scientific *attitude*: that of the impartial but curious manipulator of natural processes. A scientific object is one which works, or proceeds, without any subjective interference. According to Samo Tomšič this means that one may observe and measure objective processes which are conceptually (if not actually) eternal.[204] In other words, one can imagine the objects of one's scientific observation occurring at all times before and after, and independently of one's observations. The scientific attitude is one in which a distinction between subject and object is demarcated in order to observe patterns in nature, patterns which occur whether they are observed or not. 'Discovery' characterises the scientific attitude, where the human subject is removed from the workings of reality and can take up a position of observation and external intervention. Tomšič goes on to discuss how 'science' constitutes its objects as unstable and fundamentally indeterminable, perceived through structural effects rather than directly:

> [S]cientific modernity accomplishes a radical psychologization of knowledge by abolishing the central position of conscious human observer from the production of knowledge…Physics no longer describes the world of appearances; its object deviates from what appears to the human eye and is experimentally (re)constructed by means of technological apparatus and formal language. With this shift, scientific knowledge is no longer grounded on inefficient subjective illusions (e.g. harmony, regularity and stability) but rather on efficient objective fictions (e.g. force, structure, code).[205]

This 'informationalisation' of reality, when turned back onto the human subject in the guise of self-treatment of mental health generates a mode of being which allows one to assume a 'non-meaningful' perspective on one's own suffering. Mental anguish can be considered as informational content, which undergoes transfer from one location to another, and can be more or less represented according to the technical conditions underscoring its materialisation.

---

[204] Tomšič, S (2022) 'From the Orderly World to the Polluted Unworld'. In: Johnston, A. Nedoh, B & Zupancic, A. (eds.) *Objective Fictions. Philosophy, Psychoanalysis, Marxism.* UK: Edinburgh University Press

[205] Tomšič, S. (2018) 'Better Failures. Science and Psychoanalysis'. In: Bou Ali, N. (ed.) *Lacan Contra Foucault.* UK: Bloomsbury. p.83

The Subject of Science

For ReMind and other mental health bots, the scientific attitude which guides their design strategies and also manifests in their self-help techniques extends only to the surface of what constitutes a scientific approach. To develop the approach further would involve an understanding of the technical and experimental basis of not just design strategy but of the intervention itself. For ReMind, as for other mental health apps, and for CBT and Mindfulness style self-help strategies, this scientist-observer appears as a phantom, but also as a caricature of the scientist. This imagined scientist conjured up by these various technologies is symptomatic of the technical and social conditions from which it appears: as an externally situated and aloof technical operator. ReMind's employees assume precisely this mode towards the ReMind software: their form of mental health intervention is a reprogramming endeavour in which reconfiguration of the system creates or rediverts software-based pathways across which 'mental health' can be managed in a measurable and observable way. We can see a vestige of the ideal in which the clinician sets themself up as the ideal 'sane' model for the patient to aspire to through their treatment. The scientist observer model promulgated by ReMind is communicated through more abstract means, through the conceptual systems which form the basis of the treatment methods provided, rather than as a direct and immediate appearance. Ultimately the image of the scientist is one who is identical to oneself: as a subject who is objective. In this sense, ReMind approaches the user-subject as a consistent, or stable, object - a totality - which in turn, encourages the user to assume this consistency. What this means is that when I interact with the ReMind bot and undergo its various self-help activities, I am operating under the guise of a self-consistent and complete 'individual'. Management and re-direction of behaviour are the only options for such an individual because full self-consistency disallows any sort of subjective transformation. Change occurs on the surface: in terms of behaviour. For cognitive therapy bots like ReMind this is acceptable as their remit does not extend to the therapeutic level; ReMind explicitly limits itself to this surface level 'intervention'.

For a computer science vision of the mind, consciousness is the executive seat of cognition, the input/output machine that processes information and determines behaviour. This allows researchers to work on solving specific technical problems while ignoring the global system in which such a consciousness might be but one component. John Johnston notes that due to the physicality of early cybernetic experiments becoming redundant in the face of sophisticated computer modelling, a global sense of consciousness also becomes redundant, with abstract 'thought' now being the seat of human consciousness:

> Not only does the physical hardware so important to the cyberneticists drop out of sight, but so does the environment. Like the user himself, the disembodiment of the subject and his or her reinscription in the psychology simulated by the symbol processing is rather striking. Though much attention is given to problem solving, decision making, searches, goals, logical operations, languages, and representation, these are processes without an identifiable subject. In this sense AI is truly a simulation of abstract thought.[206]

[206] Johnston, J. (2008) *The Allure of Machinic Life*. USA: MIT Press. p.297

As discussed in the history chapter, cognitive mental health therapy retains the behavioural approach that characterises behaviourism, in which ‘thought’ is considered as a directly manipulable object. Computerisation adds a functional element in which achieving a certain output (improved mental health) is disarticulated from the process through which that output is achieved. In other words, the means and the ends are functionally associated, but they are not causally associated: one can achieve the same ends through other means without changing their character. What this means for users is that mental health benefit is associated with behavioural change, but this association is compounded by concepts like ‘reprogramming’, or ‘rewiring’ in which fundamental, or deep level changes are occurring. Instead of ‘the mind’ undergoing a reprogramming through behavioural change, a concept of ‘change’ undergoes reprogramming, in which managing one’s thoughts successfully equates to mental health. In order to provide mental health support to its users, ReMind must induce a ‘scientisation’ onto the part of the treatment, and confer the role of ‘scientist’ onto the part of the user. This involves introducing a system in which mental health can be measured. CBT does not attempt to directly measure or formalise the mind but formalises a process for interpreting and managing behaviour.

ReMind introduces a behavioural measurement system: the measurement of the users’ responses to their software, encouraging the users to measure their own activity through self-assessment. These measurements offer the semblance of scientificity but, due to their indirectness, lack the formalistic rigour that they purport to achieve. In turn, the user, as ‘scientist of one’s own mind’, indirectly assumes the role of observer-manager: managing their behaviour through coping strategies rather than addressing what is occurring in their minds. In acquiring the information delivered by ReMind, the user then incorporates this information into their mental health self-treatment. This involves a distancing of oneself from one’s mental contents in order to externally control or manage them. Addressing one’s mental health becomes a proceduralised system. By reducing subjectivity to an input/output machine, issues described as mental illness can be reduced to imbalance in an otherwise functioning system. If the mind is a machine, disconnected from subjectivity, it becomes possible to then approach it as a system which can be externally manipulated in some way so as to restore it back to full functionality. With this in mind it might be tempting to restore a theory of subjectivity back into the mind-as-machine. The consequence of the subject becoming redundant due to the effects of a machinic conception of consciousness can be thought in another way - that of how the subject is produced through this redundancy. Subjective redundancy produces what I have called the ‘scientist of one’s own mind’: a strange kind of ‘objective subjectivity'. In order to assume this subjective ‘scientific’ stance, mental health must be approached as an object, while one’s own suffering may be experienced in various different ways and can be treated in various different ways; ‘experience’ must be transformed from a phenomenal event or encounter into a measurable and discrete ‘thing’. The ‘thingness’ of experience characterises bot-based intervention in which the means to self-treatment involves externalising oneself from oneself: i.e. transforming one’s subjective experience into an externally manipulable object.

5.4 Conclusion

The invention of the digital computer ushered in a new way of thinking about thought by offering two important proofs - that abstract calculation and symbol manipulation could be performed by non-living machines, and that this operation was not dependent on any specific material basis. Due to a specific form of cognition having now been 'reverse engineered,' through the invention of the digital computer, the connection between abstract thought and a material substrate could now be considered in terms of processes, functions and systems instead of in terms of metaphysics. The split between mind and body would now become included in the realm of natural science in which hypotheses can be confirmed or denied through material experimentation. Computer science and artificial intelligence provide us with a tangible metaphor for the mind/body split: the computer essentially materialises a concept which helps to assert the veracity of immaterial thought 'running' on the biological substrate of the brain. This materialised concept only works however if 'thought' is considered in terms of software, which is algorithmic and symbol-based.

Of course, neither the company nor the app comprising ReMind makes any explicit claim as to the nature of consciousness. The computational lineage of their mental health treatment methodology emerges throughout the various activities that the app offers, and in the rhetoric of promotional material and research papers that the company produces. However, this emergence is not precisely because of the computational basis of the app: providing mental health intervention via software can take various forms. One can imagine, for example, an app which attempts to simulate a psychoanalytic form of treatment, making interpretations of users' dreams through identifying keywords and creating scenarios influenced by information from previous sessions. Whether this would equate to an 'authentic' psychoanalysis is beside the point, the question is: why have ReMind, and other therapy chatbot makers like Wysa and Woebot, chosen CBT as their treatment method instead of any other? It is because CBT has already developed as a computational form of treatment which owes most of its concepts concerning the nature of the mind and thought to the development of electronic computation and artificial intelligence. In other words, the foundational concepts on which CBT relies correspond to the functional and algorithmic operations of computers. In order to be treated by a cognitive style therapy, one must assume a computational attitude to one's own mind, and in effect, confirm the computational assumptions underpinning AI and cognitive 'mind-as-computer' theories. This assumption is not explicit; the patient (or the user of the app) is not required to agree or disagree with the theory that the mind is a computer, but is only required to conform to the method of intervention. In this sense, ReMind is a computerised form of mental health intervention, not quite because the intervention is provided by a computer, but because the epistemological background of the treatment form assumes that the operations of the mind are the same as that of a computer.

The disarticulation of the mind from the brain, and of the user from their thoughts and emotions, are conditions on which ReMind depends in order to conduct its intervention, and which the user tacitly agrees with in order to engage with the mental health activities that the app offers. In this way, 'mental health' is aligned with a disarticulated subjectivity - the homunculus which pilots itself. While computerised therapy takes on different forms, from self-help guides, to meditation activities, to chatbots, they rely on a specific mode of

transmission - that of digital media, which by providing a unified platform for treatment instils a universal mode of operation. CBT and Mindfulness mirror this universality in their ‘contentless’, or ‘non-semantic’ approach to mental health: feelings and emotions can be approached in generic terms, as algorithmics or as objects which, while experienced in different ways by different people, are to be dealt with using proceduralised techniques. The emergence of CBT as a preferred treatment for many different mental health issues and its correlated popularity among chatbot treatment is not just due to the technical style of treatment that CBT offers. It is also due to the universalist and modular mode that CBT offers: by excising meaning from the treatment method, the causes of mental suffering can be approached as generic, occurring for everyone in non-unique ways. In a similar disarticulation between meaning and non-meaning, the subjective and meaningful qualities associated with the individual are disconnected from the informational and formal qualities associated with the physical properties of the brain. The individual subject who is experienced through emotions, memories, desires and other states withdraws from the treatment - the ‘individual’ is separated from the ‘brain’. CBT paradoxically also directs the focus of treatment and responsibility for the maintenance of mental health onto the individual. This individual is however treated in the abstract - while individuals undertake their own treatment, they have no defining characteristics beyond the physical properties which are either functional (or malfunctional) to a lesser or greater degree.

Chapter Six: Conversation Design

The object is not given in advance of the viewpoint: far from it. Rather, one might say that it is the viewpoint adopted which creates the object.[207]

Introduction

This chapter is about how ReMind observes, tracks, surveys, or otherwise situates the users of their chatbot. The mediating effect of the chatbot will be of primary concern: the ReMind team designs and controls all aspects of the chatbot in terms of coordinating its conversation, adding and removing features, and developing its therapeutic style, but these aspects of the chatbot also determine how the ReMind team understands their own computerised and technical intervention. This chapter is about how ReMind designs their chatbot to converse with users, use technical measures to track these conversations, and make design changes based on how users navigate the conversations. The ReMind chatbot, as well as other mental health chatbots, is promoted as an AI mental health intervention. What does this mean? It means that some sort of adaptive, machine learning technique is used somewhere in its operation. Most AI chatbots are characterised by their ability to generate their own responses to user inputs. ReMind does not generate its own responses, but rather uses AI to sort and categorise user inputs in order to choose the most appropriate response from an archive of pre-written responses which have been prepared by the ReMind team of psychologists, conversation designers, and other employees. The ReMind team designs conversational content which helps the chatbot to give users a sense of being listened to in a caring and non-judgemental manner. The chatbot is not just a portal to the treatment techniques, it acts as an active listening companion, which the users confide in and rely on to respond to their distress. The bot is programmed to take key words from user inputs and make decisions about what the most appropriate response might be from a pre-set range of responses, written by the ReMind team.

This section focuses on the chatbot, and how it speaks to users. The actual content of the chatbot's speech will be less focused on than the methods that the ReMind team use to design, assess, and deliver this content. A primary concern will be to show how ReMind designs a system in which they are themselves implicated: the decisions they make are based on how users respond to the technical systems that they have built, but it is this same system that enables ReMind to observe how users interact. In doing so, the ReMind team not only guides users to interpret their mental health in particular ways, but they guide themselves, albeit in ways which are self-concealed. The aim of this chapter will be an attempt to illustrate this circular dynamic, and to discuss what this means for the type of treatment offered by the bot.

1. Observation - The first section deals with how ReMind observes users' activity as they engage with the app, and takes this activity as indicative of how well their mental health intervention is working. ReMind's observations are produced through, and mediated by, the users' interactions with the bot. Technical procedures associated with AI conversational design will be considered in terms of how these procedures assist the bot in interpreting user

[207] Saussure, F. (2013) *Course in General Linguistics.* London: Bloomsbury

utterances, and how in turn these interpretations are then interpreted by the bot's designers. Conversation design will be considered, as it is often referred to by ReMind employees as a branching tree in which decisions are made about therapeutic effectivity based on how users engage in conversation with the ReMind chatbot, and in so doing 'travel' along the various branches of the conversation tree. This will be discussed in terms of how ReMind generates a conceptualisation of the users as units in a system.

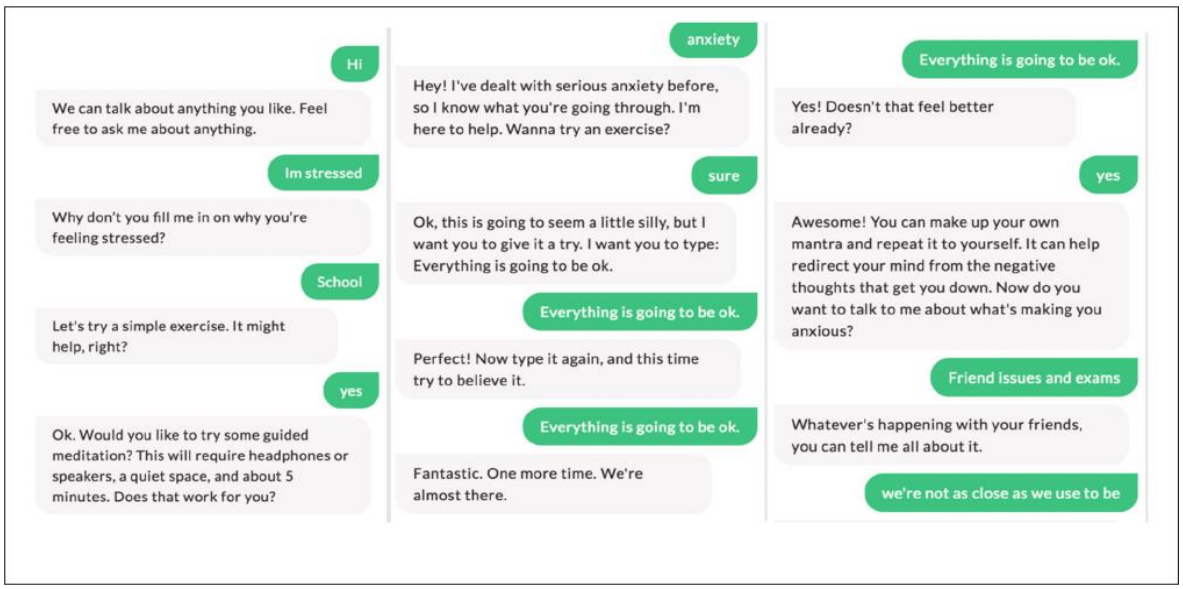
2. Interpretation - Following this will be a discussion about how the bot 'understands' what users say to it. This section begins with a discussion about how linguistic translation involves a homogenisation of language. Following this, the literary translation concept 'levelling', will be used to discuss how user utterances are translated in terms of 1. written language, 2. code, and 3. back into written language again, causing a refinement in semantic meaning. Then, these discussions will be brought to bear on how users of ReMind must interact with the bot in order to benefit from its intervention.

3. Prediction - Finally, AI prediction will be considered in two registers: one being the precise technical sense referred to by AI software engineers, and the other being how this technique goes on to 'predict' how users interact with the bot. 'Prediction' will be considered in a formal sense as the framework which guides and channels how both ReMind and the users of the app must act and think in order for treatment to be effective.

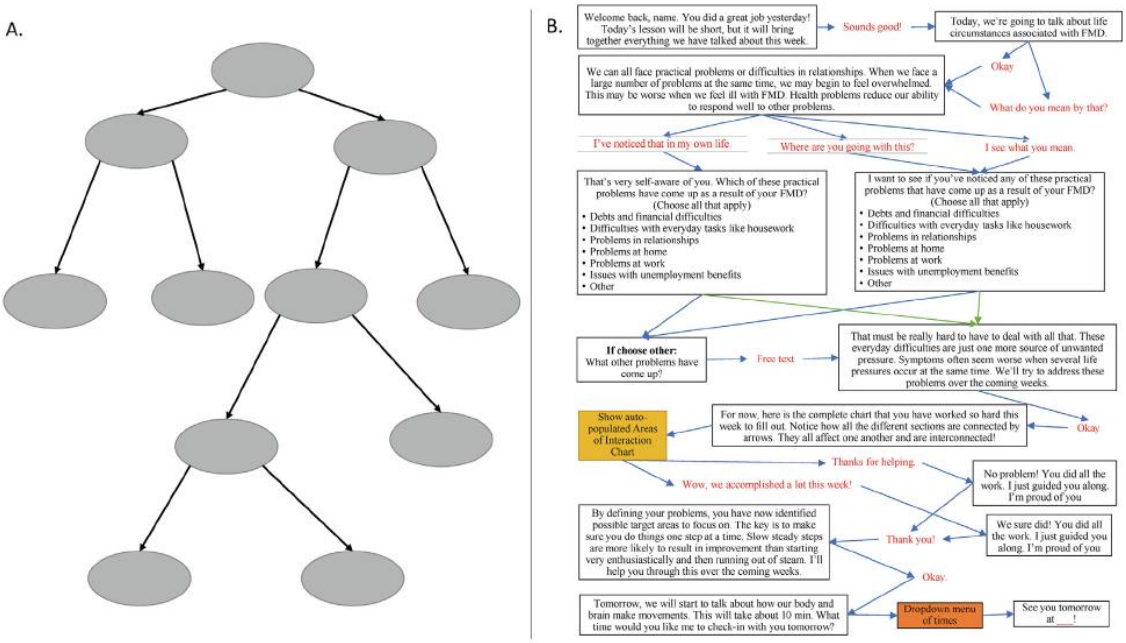
6.1 Observation

Branching Pathways

When the ReMind app is opened and the user opts to chat with the bot, the text conversation is started by the bot, usually in the form of a prompt asking the user how they are feeling. The user then responds, and the bot then replies with, for instance, sleep or exercise, or nutrition (etc.) advice if the user mentions that they are tired, or will cycle through a series of other options if the user declines the suggested help. Conversations range in levels of complexity depending on the user inputs and previous chat-history.



The above image is an example of a therapy chatbot conversation.[208] The image is from an article by Christine Grové which details the process of designing a mental health chatbot.[209] It is similar to the layout, design, conversational style and methods used by ReMind, as well as other bots such as Woebot, Wysa, Elomia and Youper. Essentially, the bot prompts the user to explain how they feel and then offers suggestions of self-help, meditation, journaling, etc in response. While the above image shows a linear conversation, what is really taking place is more like a "choose your own adventure"[210] style nonlinear conversation in which user responses to the bot's prompts then provoke different responses *from* the bot depending on the prior conversation, use or repetition of keywords, previous user choices in terms of features, etc.



211

The above image is from an article by Amanda Lin & Alberto J Espay comparing different treatment methods for patients with functional neurological disorders. It illustrates the 'branching pathways' which users navigate as they converse with the bot; this conversation structure is identical to ReMind's (and that of any 'retrieval-based'[212] chatbot) conversational structure. Lin & Espay describe the conversational sequence:

> Decision tree approach to chatbots. Panel A shows a schematic of a decision tree. The conversation starts at the topmost node, with subsequent branches and nodes representing potential paths for a "naturalistic" conversation. Panel B gives an

[208] Screenshots of ReMind software cannot be included to retain anonymity.
[209] Grové, C. (2021) 'Co-developing a Mental Health and Wellbeing Chatbot With and for Young People.' *Frontiers in Psychiatry.* p.11
[210] Fitzpatrick, K.K., Darcy, A. & Vierhile, M. (2017) '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*(2), e19. p.3
[211] Image source: Lin, A & Espay, A. (2021) 'Remote delivery of cognitive behavioral therapy to patients with functional neurological disorders: Promise and challenges'. *Epilepsy & Behavior Reports.* 16. 100469
[212] As opposed to 'generative' chatbots which compose their own responses

example of how a decision tree can be utilized to guide a conversation with a chatbot.[213]

In terms of how ReMind's conversational system works, Mary, ReMind's head of AI described how different nodes are triggered. This involves the bot attempting to match what the user is saying with 'models', which are various predefined interpretations of what the user might be attempting to articulate to the bot:

> Models are checked sequentially…as soon as any model gives a positive response in the sense it says yes, you know, what, I am supposed to detect, I have detected that element, then no further models are checked. The detection stops there. So anyway, the next node in the conversation is triggered, effectively based on what was detected…It's just, you can imagine, like the whole conversation flow is a collection of nodes, right? Each of them connected. And a particular node can have many different next nodes.

To visualise this, we can see in panel A above, each circle represents a node to which a user is directed depending on their responses to the bot, and from which different nodes will be offered depending on the users' response to that conversational node. The ReMind team alters conversational nodes if it is found that they are inhibiting users' engagement with the app: if enough users halt their engagement at a particular node in the conversation, then the ReMind team take a look at this node to understand why it may be acting as a barrier to progression. If a bot response acts as a barrier to conversational progress, then a more appropriate, or helpful, or less triggering (etc) response (whether this is a conversational response or a prompt to use one of the various app features) can be substituted. The goal is to improve the conversational flow and maintain user engagement. Charley spoke about how, over the course of development, the bot went from being very linear and directive, to being more complex and scalable:

> So my first few bots [were] very linear. You come in, "Hi, I'm stressed." Okay, fine. Let's talk about this, do this. These are the strategies and done, and then Jeff was like; "No, but you're not talking to one person. This is 1 million people, you know, so let's diversify. Let's build a tree. Let's build a dialogue tree." If somebody says yes, somebody says, No, somebody says, I don't like this, objections, and abstrusions and all those things.

Bear in mind that the workings of the bot can be divided into two distinct features: 1. the dialogue tree which is made up of many different branching pathways, and 2. the various self-help guides, CBT activities, Mindfulness programs and other features that are provided by the bot. The ReMind team adjusts the dialogue tree in response to assessment of their own 'traffic light' system, in which the status of the through flow of user traffic is measured. Reese, one of ReMind's directors explained:

[213] Lin, A & Espay, A. (2021) 'Remote delivery of cognitive behavioral therapy to patients with functional neurological disorders: Promise and challenges'. *Epilepsy & Behavior Reports.* 16. 100469. p.5

> Internally, we have a, what we call a traffic light report…[Y]ou visualise ReMind as a conversational tree. And each node, the node where drop offs are higher than expected, they start flashing red. And when that happens, then you know that okay, there's something happening at that node, you double click on that. And then you realise what are the conversations which are happening? Why are people dropping off? And then you create a hypothesis on that, then you go back to the clinical team, and say, hey, you know, I'll give you an example from very early on, when someone [who] was in grief [and grief] was being treated like anxiety. And the intervention of the conversation script, which has been paid back to the user was [for] anxiety, which is not always appropriate. And that node started flashing red. And then we double click, we realised how ReMind is not handling grief very well…Then we went back to the clinicians, to the psychologist and said, in a self help context, what is appropriate to help a user handle grief? They said, "Well, these are the techniques we typically use, and this is the conversation script." Then the designers went back and wrote a set of conversations, went back to the clinician, said, does this work? And they said, "No, this is too, you know, you need to change this, you need to change that"…But the starting point is the user telling us that something is not working.

The ReMind team, when prompted by a 'red light' at a particular node can view all of the instances of user activity which have triggered this. They must then make a judgement about what are the common features that unite these instances in order to adjust that conversational node. They must then observe whether their adjustment has a positive or negative effect on user-retention. In effect, user-retention is taken to be the measurement of whether the intervention is working or not. There can be many different reasons why users drop off from the app, but whether they maintain their engagement or not is ReMind's overriding marker of success.

Synchronic Visualisation

ReMind's "traffic-light system" establishes a novel and powerful form of *synchronic* visualisation. ReMind assesses treatment effectiveness in two ways: using standardised outcome measures and through tracking user activity. Standardised outcome measures have been discussed in the literature review chapter. One important point to mention is that, in order to judge effectiveness, outcome forms must assume that mental health can be considered objectively, in an external sense, as opposed to subjectively, or 'in the mind'. This assumption reflects the functionalist ethos which ReMind asserts through their focus on effectiveness. In the clinical psychological setting, diagnosis of mental disorders, gauging of severity of distress, administering of techniques and judging outcomes must all undergo a process of abstraction, or reduction, in order to be statistically compared. This reduction is of course necessary for a specific vision of mental health, one which can be measured objectively. ReMind assumes a statistical and computational attitude also, but in a way that is different to statistical diagnosis as found in the DSM or through assessment using patient outcome forms. This attitude can be discovered in the technical systems that ReMind employs: their conversation tree structure is both the means for intervening and the method of observation of user activity. Taking user movement throughout the conversation trees as a measure of effectivity gives ReMind a quantitative measure of assessment that involves mass-scale user movements.

ReMind has developed a method for judging effectiveness on an objective level through their use of the 'traffic light' system. This method is contrasted to outcome forms in that it does not observe mental health, or the effects of treatment as changing over time but as *occurring in terms of a system which is characterised by simultaneity and synchronicity.* These two modes of assessment - outcome forms and system adjustment - correspond to a diachronic mode and a synchronic mode. Diachrony and synchrony are terms introduced in Saussure's 1916 *Course in General Linguistics*[214] to make a distinction between historical, or evolutionary analysis or language, and structural analysis. Mark Aronoff explains that Saussure's reduction of language to timeless structure enables a systematic and non-historical method of analysis.[215] ReMind can observe user activity in terms of the movement of various elements in an overall structure. What this means is that ReMind can consider mental health not in terms of history (and consequently as a social phenomenon) but in terms of a fixed system in which movement, or adjustment, is characterised by changes to interactions between its moving parts. What is novel about ReMind (and other app-based treatment) compared to other 'manualised'[216] forms of treatment is this synchronic mode of observation and measurement: essentially, assessment about treatment effectiveness can be made in real time in terms of all users of the app as a 'simultaneous instance'. Consequently, assessment about what works and what doesn't work can be considered in terms of what assists or impedes the smooth running of the system as a whole. This means that the macro-level (the treatment method, technical medium, etc) can be disregarded in favour of adjustments in terms of the immediate responses given by the bot at specific moments during the conversation. This does not mean that ReMind employees do not consider the style of treatment; indeed, many of my conversations with employees involved discussions about different methods of treatment and their appropriateness or inappropriateness. However, the system itself is designed in such a way as to make these concerns less and less relevant: the technical acts of adjustment to conversation nodes is directed primarily towards ensuring that they do not cause user-attrition rather than providing a helpful (or ideally, 'therapeutic') response. The content of the response is secondary to the purpose of managing the flow of users throughout the conversation networks. Andrew Feenberg equates this approach with "technical action,"[217] which "represents a partial escape from the human condition."[218] What this means is that technical devices enable the perceived assumption of an Archimedean perspective, from which action may be undertaken without experiencing any counter effects. Feenberg altered the Archimedean 'view from nowhere' to coin the term "do from nowhere"[219] to describe the sense of omnipotence that technical action provides.

---

[214] Saussure, F. (2013) *Course in General Linguistics.* London: Bloomsbury
[215] Aronoff, M. (2017) 'Darwinism tested by the science of language'. In: Bowern, C. Horn, L. Zanuttini, R. (eds) *On looking into words (and beyond.)* Berlin: Language Science Press
[216] 'Manualised treatment' refers to mental health treatment which follows a standardised procedure. See: Wilson, G.T. (1996) 'Manual-based treatments: the clinical application of research findings.' *Behavioural Research and Therapy*. Apr;34(4) pp.295-314
[217] [217] Feenberg, A. (2005) 'Critical Theory of Technology: An Overview.' *Tailoring Biotechnologies*, Vol. 1, Issue 1, Winter 2005, pp: 47-64. p.48
[218] Feenberg, A. (2005) 'Critical Theory of Technology: An Overview.' *Tailoring Biotechnologies*, Vol. 1, Issue 1, Winter 2005, pp: 47-64. p.48
[219] Ibid. p.48

How are users of the app perceived in this system? During my time with ReMind, it was clear that employees were concerned with the welfare and mental health of their users: they genuinely cared for the individuals using the app, and wanted to provide the best mental health support possible. Their method of adjusting the intervention in terms of maximising user retention, paradoxically, means that their care must be applied in a removed, or austere way. As the technical system gets more sophisticated and more users are recruited, ReMind's approach to individual users must become increasingly detached from the users as individuals and more concerned with the efficient operation of the system itself. What this amounts to is a diverging sense of ReMind employees' empathy and their administration of the app: this does not mean that employees will lose their sense of caring for users, but that this will have a decreasing effect on the design of the app. User retention can on the other hand, be objectively displayed as an indicator of successful 'care'; its gradual detachment from the subjective understandings of ReMind employees will allow this metric to transpose itself onto a concept of 'care'. Efficiency of the system = care. In constructing and expanding this mode of observation, ReMind must increasingly understand users, and their mental health, as indicated by these metrics.

System Adjustment

'Mental health' is a nebulous concept, the treatment of which is equally nebulous. Due to the inherently subjective or internal nature of 'mental' phenomena, a definitive and 'final' concept of mental health cannot be established. Instead, we can construct a model of mental health through an analysis of the style of treatment. We can understand the treatment styles which influence ReMind's interventions as forms of 'coaching'. CBT and Mindfulness are fundamentally didactic forms of treatment in which the practitioner imparts various techniques to the patient. The treatment style informs how mental health is conceptualised: 'mental health' is approached as something which is manipulated through the use of techniques (meditation, reframing one's thoughts, journaling, etc). For ReMind, the technical measures that make up the software application inform a technical attitude towards mental health, and because these measures appear to 'work' (through patient outcome measures, user-reviews, etc) then this seems to give credence to the concept of a technical mental health. When speaking to ReMind employees, their concerns tended not to reach the point of assessing the general context - the meaning - of their intervention, but rather focused on the various technical and design challenges throughout the course of app-development. This is because in order to technologise a mental health intervention via computer automation, it is necessary to deal in 'objective abstractions'. In other words, the question of how effective the bot is when, for example, offering the user a sleeping guide due to the user mentioning that they cannot sleep, is measured in terms of whether it was 'correct' for the bot to respond in such a way, via the user accepting or rejecting this suggestion. The question of the 'effect' (i.e. what it means to offer practical suggestions) of this process cannot be considered because the makers of ReMind are constrained to a technical/instrumental understanding of the possibilities of the bot. In other words, the range of possibilities that the bot offers in terms of mental health intervention are defined by the formal conditions of a computerised chatbot. For example, the use of silence is a common practice in mental health treatment, but, because the bot is 'always on', and always available, silence would be interpreted as a technical error. Instead, the bot can offer a mental health intervention that is determined by its technical and computational basis.

This is what it means to provide an abstract form of treatment: due to both the technical limitations and competencies of the bot, the intervention cannot take the form of a 'traditional' psychiatric, psychological or psychotherapeutic intervention and instead must rely on what is possible within its own means. ReMind uses a synchronic mode of observing user activity, and subsequently, using that mode to assess their treatment, the abstraction necessary in assuming this mode (converting users into clusters defined by their movement throughout a system) means that ReMind cannot but understand their intervention in an abstract manner. Abstraction here is a formal exercise: 'users' and their movement around the conversation tree are decisively split between their content and their form. ReMind can observe users in terms of individual conversations, and also observe users in terms of quantity, movement, timings, and any other formal means that can be associated with their usage (or non-usage) of the app. In order to assume this abstract mode, ReMind must convert 'mental health' itself into an abstraction: as opposed to any form of diagnosis and treatment, their intervention cannot 'treat' any specific form of mental illness but must reduce their diagnosis to a generality - mental suffering - as unspecified and undetermined by any external factors be they social or historical. In this way it is possible to offer a computer-automated form of mental health intervention that is said to be 'effective'. The way that ReMind interacts with users becomes a technical action of adjustment.

Andrew Feenberg describes 'technical action' as an approach made possible due to the mediating effects of technology. One assumes a 'technical' attitude towards one's objects as an *effect* of technology, rather than the other way around. Feenberg goes on to say that, because of the unconscious nature of this process, such terms as 'technical' and 'efficient', take on a tautological character in which justification for technical or efficient action appears to be self-affirming:

> What makes technical action different from other relations to reality? This question is often answered in terms of notions such as efficiency or control which are themselves internal to a technical approach to the world. To judge an action as more or less efficient is already to have determined it to be technical and therefore an appropriate object of such a judgment. Similarly, the concept of control implied in technique is "technical" and so not a distinguishing criterion.[220]

In other words, the introduction of technical means to accomplish tasks has the effect of producing a 'technical attitude'. Many of the employees of ReMind had reservations about using technological means to treat mental health, but these reservations were always focused on 'effectiveness': was this automated bot as or more effective than its non-automated counterparts?[221] Charley, ReMind's clinical research and development head, recalled doubts upon joining the ReMind team: "...there's, basically, a disembodied person writing words on a screen. And that is evoking emotion and insight within you. So, yeah, I mean, a lot of my initial thoughts would be around, is this even effective?" Note that Charley's concern was whether the intervention was 'effective' rather than involving what

[220] Feenberg, A. (2005) 'Critical Theory of Technology: An Overview.' *Tailoring Biotechnologies*. Vol. 1, Issue 1, Winter 2005. pp.47-64. p.47

[221] As discussed in the literature review, chatbot based treatment apps produce their own research to measure treatment effectiveness, some of this research is focused on studies which compare the bot to online CBT.

might be the nature of ‘effects’. Assessment of effectiveness in a synchronic mode allows ReMind to avoid the question of what ‘effectiveness’ might refer to in terms of mental health of individual users, and instead can focus on technical adjustments to the app, adding new features, and improving aesthetic design. Taking this approach frees ReMind up to make the app more ‘effective’ on purely technical terms and avoid critical assessment of situating the app in terms of social or political, or obviously, historical conditions. As discussed above, the ‘technical attitude’ which comprises using technical means to measure technical success is a self-sustaining tautology: effectiveness is judged through observation of user activity on an objective level. A precise number can be attached to ‘effectiveness’ in terms of both global user numbers and the activities of specific user groups as they cluster and disperse around conversational nodes. Because a numerical measurement can provide an objective and standardised means to display user activity, assessment of user activity itself becomes the means to proving ‘effectiveness’. This circularity is founded on ReMind’s structural ability to observe how users navigate the app, because it enables ReMind to ignore any questions about why people suffer from mental health conditions and to instead get on with their task of providing a solution to what they consider to be the problem: mental suffering in a general, non-specific sense. We can see an inversion occurring here in that the precise and objective measurements that ReMind employ to judge effectiveness is only possible when the scope of the problem that they are addressing is approached as vague, immaterial, and ahistorical. In other words, ‘mental health’ can be viewed as something which has no external cause, and as such its treatment can take on a purely technical character.

6.2 Interpretation

Translation

The technical attitude which results from the use of technical means, then feeds back into the design of the various systems that make up the bot. Interpretation, i.e. the assigning of meaning to statements, involves an exclusively technical operation. The task of assigning meaning to the user’s typed inputs involves ‘tagging’. The bot has access to a bank of responses, which are linked to keywords; when the user types in keywords or combinations thereof, the bot matches these with possible responses. As mentioned above, this is known as a retrieval-based chatbot. AI is used to sort and classify user conversation inputs, but not to determine bot outputs; this is because of the danger of fully-fledged AI procedural conversation being too unpredictable. ReMind uses artificial intelligence in a one-sided manner: to catalogue and interpret user inputs, but not to generate responses. Mary, ReMind’s head of AI, spoke about this:

> [L]et me start by saying that, right now, [how] AI is primarily used is how to make sure that ReMind understands the user correctly. Because since we are dealing with natural language, you know, the user can say so many different things in so many different ways. So it's a very challenging problem, to understand what the user is saying and being able to respond correctly to it. So although the responses are all pre-written by the conversation design team and the therapist team together, they need to know what they are responding to. So if the user is…if the user is describing a particular scenario, they're describing a relationship issue, versus they are describing a loss of someone or they are describing, you know, that they are having

> trouble at work or not able to focus on their, on their studies, ReMind needs to understand that and guide the conversation appropriately.

When I spoke to Mary, the first topic that came up in the interview was the challenge of introducing ReMind to other languages. This involves the host language (in this case Spanish) being translated into English, the bot's response being generated and then being translated back into Spanish. Mary spoke about this in terms of a gradual process of adjustment and refinement, in which improvements are judged by how accurate a translation is: "So as long as it doesn't change the overall meaning of the original text, original message, some errors we can live with." This attention to accuracy applies not just to linguistic translation but also to translating the users' utterances into a 'language' that is understandable to the bot. The bot is designed to identify keywords from the user's typed input, and to match the keywords with possible responses. The 'meaning' of the user's input is translated into terms that align with the meaning contained in the bot's database of keywords. A user may understand what they are saying in many different ways. "I'm down" obviously has multiple potential connotations, but it is the job of the ReMind bot to narrow down these connotations to such a level that an appropriate response can be made. Naoki Sakai discusses how the act of translation confers equivalence between the source language and the target language, i.e. it is only after translation occurs that the battery of meanings which comprises each linguistic form appears as equal: "the presumed invariance of the message transmitted through translation is confirmed only retroactively, after it has been translated".[222] ReMind automates this process in order to confer equivalence between what users write and its database of possible responses. This means that any statement that a user makes must be tagged; 'tags' are applied to various words in the statement, the combination of which are then fed into ReMind's database of response, and the response which best fits the tags is then provided to the user. The 'meaning' of the user's original statement is, through this process, made equivalent to the interpretation that is applied to it: it is made comprehensible to the bot on the bot's terms. Sakai goes on to imagine the opposite scenario in which a language cannot be translated:

> If the foreign is unambiguously incomprehensible, unknowable, and unfamiliar, then translation simply cannot be done. If, conversely, the foreign is comprehensible, knowable and familiar, translation is unnecessary. Thus, the status of the foreign is ambiguous in translation. The foreign is incomprehensible and comprehensible, unknowable and knowable, unfamiliar and familiar at the same time.[223]

We can understand ReMind as taking part in a translation exercise occurring in a context wider than that of individual user utterances. Roman Jacobson divides the act of translation into three distinct classes:

> (1) Intralingual translation or rewording is an interpretation of verbal signs by means of other signs of the same language. (2) Interlingual translation or translation proper is an interpretation of verbal signs by means of some other language. (3) Intersemiotic translation or transmutation is an interpretation of verbal signs by

[222] Sakai, N. (2006) 'Translation'. *Theory, Culture & Society.* 23:2-3. pp.71-78. p.72
[223] Ibid. p.73

> means of signs of nonverbal sign systems.[224]

While we can understand the individual instances of ReMind's interpretation process as occurring in terms of (1) intralingual translation, on a broader scope, ReMind is conducting "translation proper": converting one language, that of individuals articulating how they feel, into another, that of ReMind's database of responses. This translation involves Sakai's formula of the foreign as "incomprehensible and comprehensible, unknowable and knowable, unfamiliar and familiar at the same time". The users' utterances are unknowable but they are rendered as knowable as an effect of the translation process. In other words, the users' varied and individual experiences of mental suffering is retroactively consolidated through the interpretive acts that ReMind performs.

Levelling

Machine translation involves using automated interpretive systems to suggest a given term the more it is used to translate a certain word. As this process continues, a refinement occurs in which specific translations for specific words are more and more judged as 'correct'. Françoise Wuilmart discusses this refinement in terms of "levelling":

> The phenomenon of *levelling* goes to the very heart of the problem of any literary translation. *Levelling*, or even "normalisation", that is to say the action of "planing" a text or flattening it: removing all kinds of relief, truncating the points, filling the hollows, flattening all the asperities which precisely make it a literary text.[225]

The levelling effect occurs in the translation of user inputs into categories which can be sorted by the bot through its use of tagging. In order to be able to provide an appropriate (or as may be the case, non-*inappropriate*) response, the bot must successfully categorise what the user is saying. In order to categorise a user input, the bot retains an archive of possible semantic meanings which can be applied. These meanings, or tags, are predetermined by the ReMind team. The challenge for the ReMind team is to create an interpretative system which can manage the range of contexts within which words carry different meanings. 'Context' for ReMind means the extended range of meanings that words take on *within sentences* and not in terms of subjective context. A tag is a description of a user statement which shows the bot where to categorise this statement in terms of semantic content: 'sad', 'anxious', 'hopeful', etc. The decisions that the ReMind team makes about whether tags are correctly applied is conducted in terms of how accurately the bot is able to respond to user statements. Accuracy is measured by the ReMind team in terms of their own judgement about the appropriateness with which the bot's responses are using the 'traffic-light' system. Conversation designer Alan spoke about the requirement to maintain an 80% accuracy level:

> I don't think everything can be 100%. That's why when we said appropriateness, we make sure that we have that limit of 80. It cannot be less than this. But it should be at

[224] Jakobson, R. (1971) 'On Linguistic Aspects of Translation'. In: Roman Jakobson (ed.) *Selected Writings, Vol. 2*. The Hague: Mouton. pp.260-66. p.261
[225] Wuilmart, F. (2007) 'The Sin of "Levelling" in Literary Translation'. *Meta magazine*. Volume 52, Number 3. pp.391-400

> least this, this is our internal way of looking at appropriateness on responses…Responses that are more clinical in nature: 100%.

ReMind does not necessarily purport to have achieved or to be able to achieve "100%" accuracy in programming the bot to interpret user inputs. But in operating within the paradigm of 'accuracy', at least *some* level of understanding is necessary. While it may seem rudimentary to claim that understanding what is being said is a prerequisite for successful treatment, 'understanding' can be interpreted in different ways. ReMind's interpretation of understanding involves making sure that the bot's responses are at best, deemed appropriate, and at worst deemed *not inappropriate*. On ReMind's terms, the basis for successful mental health intervention is founded on accurate interpretation, and accuracy is a zero-sum operation: the bot either gets it right or it gets it wrong, and the ReMind team must adjust its interpretive procedures accordingly. 80% accuracy for ReMind means that the bot correctly interprets 80% of *all of* the users' utterances, not that it manages to get a pretty good (80%) gist of *each* utterance from *each* user. What this amounts to, for ReMind, is a mental health intervention that relies on the successful transfer of information from one node to another: the user says what they mean, and the bot either understands or does not. This is an information and communications technology version of mental health treatment, in which the accurate transfer of information is the prerequisite of a successful treatment. 'Interpretation' involves, on one hand correctly assigning meaning onto a user statement, and a process whereby more and more user statements are made understandable to the bot. Understandable here means having undergone a levelling process in which the meanings of individual terms, such as 'sad', are shorn of contextual meaning, that is, the semantic content that a user might ascribe when speaking to the bot.

"What is the best ice cream for when all of the taxi ranks are flooded with hair and I need to walk to a basis for every nice colour?"

When I repeatedly typed the above question in response to the bot asking a direct question ("When would you like me to check in with you?"), one response from the bot was "I understand," but also then "Phew! That was a lot of questions." The bot did finally reply with "Sorry, I didn't quite get that." This was followed by its question being repeated with prompts for possible answers: "You could say 10pm, or 08:00." When the bot asked me to confirm the day and time I responded with the same nonsensical question, the bot responded by letting me know that it would check in with me tomorrow. The bot's interpretation mechanism works in such a way as to always attempt to 'correctly' interpret the users' inputs, even if this attempt is far off the mark. As I continued to input the question to ReMind's prompts and responses, what took form was what *looked* like a conversation, with responses such as "Tell me more", "What will you remember about it the most?" and ending with "Thanks for sharing that with me". While my nonsensical question is obviously not an appropriate way to interact with the bot, it illustrates that the bot at least attempts to respond in a meaningful way.

‘Good Enough Accuracy’

At a certain point, the bot’s ability to understand reaches its outer limit: when I typed in a long string of random letters, the bot responded with “it seems that you have typed a lot of letters that don’t make sense.” The bot then enquired if I am trying to test its capability, making a joke about the Turing test; it then asked if I would like to continue with the previous discussion. This was the single instance where the bot’s ability to interpret could be said to be 100% correct: I *was* trying to test the bot’s capability. The bot is after all just a bot: one cannot speak to it in an unambiguous way and expect this to be ‘correctly’ interpreted. One must in fact speak in a very direct manner, articulating one’s distress in a straightforward and unambiguous way so as to access the ‘correct’ help. The user’s mood, or state of mind, or risk, is assessed - tagged - so that appropriate responses can be made, and consequently the ReMind team gains an understanding of mental health on a macro scale. The result of this is that ReMind comes to view ‘mental health’ as a phenomenon which, while affecting individual users, is assessed in terms of how to generalise sets of heterogeneous statements so that appropriate tags can be attached to them. The ReMind bot *cannot not know*, and as such, will always respond in such a way as to appear to understand what the user says to it, even, or especially, when the user is trying to test the bot. ReMind’s method for expanding the scope of the bot’s understanding is by introducing more interpretive tags. In order to create a conversation which feels personalised and appropriate, more and more tags for different categories must be introduced, with a system of categories and subcategories emerging throughout that process. The outcome of this is that ReMind becomes more adept at interpreting conversations. Another outcome is that, as more categories are added to the structure, more possible conversations fall into a structured array for which outputs can be delivered. Machine interpretation of user statements through the use of tagging is necessary for the bot to be able to respond appropriately. In doing so, the statements that all users provide to the bot become incorporated into a larger network of categories and sub-categories.

In order to benefit from the app, users must alter their way of interacting with the bot in order for their inputs to be ‘understood’. As one user wrote in their review: “It's a good app if you understand how to explain yourself to AI.”[226] The user conforms to a mode of speaking which encourages an appropriate response from the bot. A model of mental health emerges from this dynamic in which users must already be able to articulate what is wrong with them. The cause of mental suffering cannot be ambiguous, but must emanate from some identifiable source, e.g. ‘I’m sad because my friend moved away’. This must be interpreted as the direct source of suffering and be treated as such, perhaps by offering the user its ‘Reconciling with Grief’ module in which the user is encouraged to write a letter to their lost one. Mental suffering cannot stem from an indirect, unknown, or unconscious source, as the bot can’t handle ambiguity. In this way, by navigating the branching conversation pathways offered by ReMind, one approaches one’s own mental health in terms of ever increasingly accurate declarative statements, and one learns how to explain oneself to bot. The bot interprets those explanations as having already been determined by the prediction model. Obviously, the bot is being programmed to provide more sophisticated treatment, to interpret more ambiguous or indirect statements, but it can still only do this by transforming those statements into unambiguous and direct keywords. Ambiguity is not possible on the part of

[226] Google Play Store review

the bot, and this has effects on how the bot's intervention occurs: we can see that it must offer almost entirely pedagogical or practical assistance to users.

ReMind achieves precision through the use of artificial intelligence to categorise user inputs: they maintain a database of 'tags' which are used to define under pre-set terms what the user had intended to say when they write to the bot. Precision here means that there can be no equivocation about the meaning of user statements, which is necessary because the bot would not be able to respond without first being able to categorise the user's input. ReMind has had to create a system from which it is possible to more and more accurately judge 'effectiveness': a technical system. In order to judge effectiveness, users must be transformed into carriers of information. In cybernetics terms, 'information' can be described as non-meaningful communication - the physical medium of a language, whether this is a sound wave emitted by a human voice, a written text, or the script of a computer program. The more accurately and efficiently that any of these forms can be transmitted, the more capability is afforded to the discipline that depends on that form of communication. The more effective ReMind is in transforming users of the bot into carriers of information, i.e. through the creative analysis of user activity in terms of clusters or individuals, the more ReMind can judge their own technical solutions as 'effective'. As Feenberg notes, there is a tautological aspect to this: 'effectiveness' is itself an effect of a technical approach, and as such, is self-proving. This means that ReMind, in pursuing effectiveness, need not be concerned with improving the mental health intervention provided by the app, but can pursue ever more creative and elaborate methods of technical observation and analysis. Increased 'effectiveness' will be a logical consequence of this strategy. 'Effectiveness' can now be thought of not just in terms of a measure of success of the bot's mental health intervention, often in comparison to its traditional counterparts, but also in terms of how *accurate* this intervention is. Accuracy here is related to a sense of how the bot manages to interpret user inputs and can be taken as an overarching measure for how ReMind views the success of its invention.

6.3 Prediction

What Will Have Been

Head of AI Mary spoke about how the bot sorts, categorises and responds to user inputs in terms of prediction. Prediction is based on taking partial aspects of the user's inputs and estimating how likely it is that an input conforms to one of a set of predefined inputs. According to Mary:

> None of the models are trying to understand the user input in all its entirety, like we humans would, when you say something to me, I understand the whole meaning…I'd be able to say what you are talking about, what topics you are talking about, I'd also be able to say whether you're agreeing with me or disagreeing with me, I'm able to see all of it at once as a human being, but AI is not able to do that. So there are different models, which look at only certain aspects of the user message. One model is just looking for any mention of suicidal ideation. And another one is just looking for if the user is angry at ReMind. Each model predicts for whatever it is supposed to look for, it predicts whether that element is present in the user text or not.

There is an ongoing process of prediction occurring during the conversation, however the term ‘prediction’ in AI speech recognition does not mean forecasting what the user will say, but about accurately guessing what the user has intended with their input in order to give the expected output. In the above quote Mary uses prediction to refer to a process of interpretation: the bot estimates the meaning of the user’s utterance. The term ‘prediction’ is also used to refer to the way that a chatbot provides responses: by estimating what the most likely response should be given the training data, previous conversation and grammatical rules. This does not apply to ReMind because all responses are pre-written, so ‘prediction’ is used purely to refer to how the bot interprets user inputs, and is measured in terms of whether the bot then chooses more or less appropriate responses. If considered in terms of forecasting, prediction can be thought of as estimating which response is the most likely to be the correct one. User conversations are categorised according to how likely it is that they are saying what each model is predicting. Judgement as to whether the model being applied to the user input is predicted correctly or not is made by the ReMind team: they must assess whether the bot’s predictions are correct on a case-by-case basis. As mentioned above, ReMind makes these judgments in cases of ‘red lights’: instances where an arbitrarily determined number of users drop off from engaging with the bot. ReMind’s mode of interpretation involves assessing what the users *will have said* rather than what they intend on saying. The ReMind team must write all of the scripts which form the responses to anything that the users say to the bot. ReMind’s conversation designers come up with new responses whenever a ‘red light’ is activated and a more appropriate response, or a more accurate interpretation, is needed. The process can be thought about in terms of an increasing stockpile of responses, ready to be deployed the next time a user triggers them. All of these responses are invented by the ReMind team, but they are invented in terms of the technical form of treatment. The ReMind bot’s communications are issued to the users within this technical mode: what can and cannot be said is determined by the format of the communication method. In a face-to-face conversation, the intentions, meanings, moods, etc, of one’s interlocutor are always being assessed, adjusted for and responded to in some way. With the ReMind app, this assessment, adjustment and response occurs throughout a complex range of abstractions. ReMind is constrained by this technical form in that the system that they have built - the ‘traffic-light’ indication system - only allows for technical adjustment to comprise the intervention into the mental health of the users who interact with the bot. By establishing this technical method of adjustment, ReMind must also adjust its own criteria for what constitutes ‘mental health’ as a technical object. Users must adjust their own way of speaking to the bot in order to be provided with appropriate care, and this is mirrored in the way that ReMind employees must design the bot’s conversations. They must conform to the mode of interaction which is provided by the bot, which they themselves have designed. Prediction can now be thought of in terms of the overall system: not only is the conversational interaction between the user and the bot predetermined by the responses that have been pre-written by the ReMind team, but on a wider level, ReMind’s entire technical achievement can be thought of as having been pre-designed: it ‘will have been’.

Ready-Made Problems

Machine interpretation of user statements using tagging is necessary for the bot to be able to respond appropriately. In doing so, the statements that all users provide to the bot become incorporated into a larger network of categories and sub-categories in which statements

must be definitively allocated. Dan McQuillan associates this process with the term 'ready-made problems':

> Even when it seems to produce 'new' knowledge it is doing so in a way that is wholly tied to the conditions under which the training data was generated. This looking back not only applies to the training data but to AI's mode of analysis, which is based on 'resemblances ... between the new object which we are studying and others which we believe we already know'. So whatever problem AI attempts to solve becomes what philosopher Henri Bergson would call a 'ready-made problem' – a problem that is expressed as a function of things prior to itself that have already been turned into abstractions.[227]

While McQuillan is discussing how generative AI works, the same principle applies to the ReMind bot's interpretive function: by sorting user utterances into existing categories, the ReMind team places those utterances into a system of resemblances. In so doing, the 'new' utterances made by the users are made into utterances "which we believe we already know". In other words, all of the possible statements made by users of the bot are 'understood' by the bot in terms of its own tags, and this is done prior to those users making those statements. While the ReMind team does employ psychologists to assess user inputs in order to give appropriate responses, they do not provide formal diagnoses. Instead, there is an automated and gradually expanding diagnostic process occurring in the tagging and categorising of user statements. The user's mood, or state of mind, or risk is assessed - tagged - so that appropriate responses can be made. Consequently, the ReMind team gains an understanding of mental health in terms of the tags, which they themselves have imposed, but which are applied autonomously by the bot. The result is that ReMind comes to view 'mental health' as a phenomenon which, while affecting individual users, is assessed in terms of how to generalise sets of heterogeneous statements so that appropriate tags can be attached to them. The decisions that the ReMind team makes about this are in response to the automated 'diagnosis' that occurs when the bot's AI categorises users' inputs in order to process them and respond. We can see a tautological cycle in that the ReMind team creates their own set of tags to be applied by the bot, which in turn influences how ReMind comes to view how users benefit from their interaction with the bot. The ReMind team increasingly transforms their own variegated models of mental health, and the ways that users experience mental health, into a singular technical model. We can understand the process as a sort of rhythm: in constructing a technical method of observing and tracking users, ReMind increasingly comes to understand their users as technical objects, and subsequently design their app in terms of this technical objectification. The ReMind team is paradoxically free and constrained: free to develop and deploy its own vision of mental health intervention through the power of computer automation, and constrained within a paradigm of mental health which guides 'intervention', as opposed to 'treatment'.

There is a sense of freedom in abstracting from material, quotidian and isolated conditions which mark the mental health of individuals in order to create a powerful computational system which deals in abstractions. However, in pursuing this freedom, the ReMind team is disconnected from the context from which their abstractions spring: the determining

[227] McQuillan, D. (2022) *Resisting AI: An Anti-fascist Approach to Artificial Intelligence*. UK: Bristol University Press. p.43

conditions from which technical solutions appear as obvious and unremarkable, however complex and challenging the task of achieving those solutions may be. The ReMind team is *necessarily unaware* of their determining conditions: they must assume that their intervention is novel on the level of content in order to proceed. 'Content' here means the various features and design aspects of the ReMind app, and specifically, the 'traffic light' system of observing user activity, the use of AI to predict user statements, and the assumption that user statements can be interpreted in terms of 'accuracy'. What this means is that in order to understand what the users are saying to the bot, the bot must predetermine what the users *will have said* to it through its archive of responses, and then respond in terms of increasingly homogenous semantic boundaries. In effect, all user inputs are 'pre-interpreted', according to strict measures, and those measures, along with any response that the bot offers, are constantly being inscribed within a conceptualisation of mental health which is on one hand, produced by this three-step process, and on the other hand, understood by ReMind employees as an intervention which they themselves have designed. The consequences of this are a rational form of treatment which takes on the form of a zero-sum procedure: a sort of formula in which problems present themselves, through interaction with the bot, in terms of their own solutions which are then provided by the bot.

The ReMind bot interprets user inputs based on a database of prior interpretations: all user inputs, whatever their intended meanings are *always already* interpreted. The retroactive temporality at play here is not just confined to the way that the bot interprets user inputs: the technical means and measures that ReMind has developed and deployed in the form of the ReMind bot are the means and measures within which the ReMind company judges its own success. In this way, although their intervention will get more and more 'effective', this does not quite mean that the treatment will become better, and that the mental health of users will benefit, but that ReMind's own understanding of mental health will, through the process of increasing intervention effectiveness, come to understand mental health more and more as a technical problem.

Feedback Mechanisms

As discussed, mental health chatbots don't provide formal diagnoses, they listen and they respond, and they offer various self-help tools like meditation, breathing techniques or sleeping guides. There is a diagnostic process occurring in the most basic sense though, in the classification of user statements, processing by the AI, and then the bot's response, which the user then responds to, and so on. Ian Hacking describes his theory of looping in mental health diagnosis as a dynamic in which: "[c]lassifying changes people, but the changed people cause classifications themselves to be redrawn"[228]. Chatbot classification occurs in terms of identifying keywords ('tagging') and responding in terms of those keywords; users then change their behaviour in terms of these classifications, in the simplest sense, by undertaking the activities recommended by the bot, but in a broader sense, by learning to articulate their own mental health according to the bot's requirements. In order to benefit from the app, users must alter their way of interacting with the bot in order for their inputs to be 'understood'. As one user wrote in their review: "It's a good app if you understand how to explain yourself to AI." The user must conform to a mode of speaking

[228] Hacking, I. (2004) 'Between Michel Foucault and Erving Goffman: Between discourse in the abstract and face-to-face interaction'. *Economy and Society.* 333. pp.277-302. p.279

which encourages an appropriate response from the bot. For example, the bot will struggle to provide an appropriate response if a user types a long and ambiguous statement about their feelings; users must be concise in order to get a response in which it feels like the bot has understood. The continued development of the bot depends on precision; the way that it interprets user inputs, while involving statistical guesswork, ultimately cannot be ambiguous: it must settle on a 'correct' interpretation in order to provide a response. This is the opposite of, for example, a psychoanalytic interpretation which is offered as a means to further consideration; 'interpretation' in this sense is an ongoing process. For ReMind interpretation involves simply either a correct or incorrect response, determined retroactively by the ReMind team as they observe user-bot conversations in aggregate.

User statements must be definitively allocated: there can be no ambiguity over interpretation for the bot to be able to assign meanings to user inputs. When a user types in a sentence comprising their communication to the bot, the bot then processes this communication according to pre-set tags which are categorical: unambiguously explicit and direct. Even though the bot has a range of options for possible interpretations, whichever one it decides on is then categorically set upon: the bot, due to its codebase, cannot maintain any ambiguity about its interpretations. Those interpretations are 'ready-made', not just as tags introduced by the ReMind team but in terms of the necessity of tagging itself. No statement provided by the user can be 'un-understood': in which the range of possible meanings give rise to interpretation as understood on a purely non-categorical level. The bot, along with the user, cannot 'not know', and it will be increasingly unable to not know. Because the bot pre-designates, through its tagging system, what the users 'will have said' when they articulate their mental health, users are inculcated into the bot's system in a retroactive fashion. This means that, in similar fashion but not identical to Hacking's theory of looping, the classification of mental health/illness congeals into an increasingly 'known' pattern within which being the inability to articulate oneself in concise and predetermined ways cannot set the stage for the provision of mental health treatment. 'Not knowing' what is causing one's distress becomes subordinated to 'knowing', even if that 'knowing' does not reflect the cause of distress. 'Accuracy', in the quest to achieve an ideal form now assumes its opposite: by establishing a system in which it is assumed that words unequivocally 'say what they mean' even if that meaning does not accurately correspond to the causes of distress, the statements which users make when conversing with the bot ultimately only refer back to themselves. The interpretive, or rather hermeneutic, act of delving into underlying or deeper meaning is bypassed so much so that the meaning of mental distress, as produced by the ReMind bot, is also bypassed.

6.4 Conclusion

The treatment that ReMind offers is a rationalised form of intervention, achieved through the compartmentalising of conversations into branching nodes, with those nodes being, on one hand, the means through which the user is provided with treatment, and on the other hand, feedback mechanisms for judging the effectiveness of treatment. But we can detect a thread of irrationality running through this endeavour in the form of unintended consequences: the system that ReMind has constructed constrains and guides their own ability to understand how users approach their own mental health. This is achieved through a paradox in which assuming an omnipotent perspective comes at the cost of mis-apprehending this perspective. The intervention that ReMind has developed is guided by the technical conditions that ReMind has put in place to monitor, assess and to respond to the utterances of users. They have constructed a system in which the users' activity allows them to assess the efficiency of the overall system. The act of translating the user's inputs into predetermined interpretations using keywords confers equivalence between what the users say and the bot's allocation of those sayings onto its database of responses. The bot judges which is the most likely response, with the likelihood being judged retroactively, in terms of the user's response: whether they agree or disagree with the bot's response to their initial input. This is observed by the users' activity throughout the conversation tree. This is a complex sequence in which written utterances gradually assume parity (become translated) with the bot's database of responses. We can directly observe the retroactive effect of translation as described above by Sakai, in which the bot assumes the role of translator. As Mary pointed out, translating the bot into other languages beyond English involves making sure that the "overall meaning of the original text" does not change. This process involves both levelling and conferring equivalence: the users' inputted utterances become levelled, or flattened in order to be made comprehensible to the bot in terms of matching up with its database of responses, and in so doing, acquire parity with those responses. The circular dynamic here has a complex temporality. In Hacking's looping effect, the sequence of mental health classification occurs in a straightforward progression: classification affects how that classification is experienced/enacted, and subsequently that enacting affects classification. However, in ReMind's case, the bot classifies user utterances prior to those utterances being made, 'predicting' their experiences of mental health treatment. The temporal cycle that Remind constructs for itself is a self-referential closed loop, and we can see it manifesting in various forms; in the way that ReMind observes its users by constructing them as a 'simultaneous instance'; the self-contained translating of not just languages, but of meaning itself; and their method of predicting what users will have said.

The ReMind team has produced a machine through which they are increasingly led to create their app 'in their own image', i.e. in terms of their own ideas about how all of the users of the app experience their own mental health. We can understand this in terms of Feenberg's 'do from nowhere' in which, whether intended or not, ReMind assumes an omnipotent perspective toward the users of the bot in which it is impossible to 'not know'; to be unable to interpret what a user is saying, to the point in which it is even possible to predict the intended meanings of user utterances. This omnipotent perspective is not assumed 'in the minds' of the ReMind team: they do not need to see themselves as omnipotent, and in fact they tend to express their perspective in the opposite way, that they follow the lead of the users in making design decisions. This will be discussed in more detail in the next chapter. What is

important to note is that, by constructing their system of observation and intervention, the ReMind team positions themselves in terms of a ‘god’s-eye view’. This perspective is not assumed in a metaphorical or conceptual sense: it is materially assumed as a result of the technical measures ReMind has constructed. In constructing the means to achieving this perspective, through 1. ‘synchronic visualisation’, 2. conferring parity between user statements and the meanings of those statements, and 3. predicting the intended meanings of those statements, ReMind implicates itself into its own technical vision. ReMind is not simply interpreting the users but incorporating them into a system, and in the process incorporating itself into this same system. What this means is that, in a paradoxical way, the ReMind team both asserts a rationalised and objective control over their mental health intervention and the experience of the users, and at the same time is coerced into an understanding of their intervention and their users from within the terms supplied by the system. The mechanism they have constructed allows them to achieve an Archimedean external perspective, which is at the same time, a perspective from *within*: from the mechanism itself.

Chapter Seven: Macro-Treatment

Production not only supplies a material for the need,
but it also supplies a need for the material.[229]

Introduction

This chapter is primarily concerned with how ReMind approaches users of the app in mass-scale: as a group or groups. ReMind's broad approach to app design will be considered initially, the focus will be on ReMind's design ethos and objectives rather than specific design features. The chapter deals with how the ReMind team responds to the users of the bot, but also construct the users as mediated by the bot, in order to adjust the treatment. While it is possible to view individual user conversations and to adjust conversational content, decisions about conversation design are also made in response to users being aggregated into large groups in order to be treated as classes or clusters. The psychologists, conversation designers, and other employees of ReMind all cooperate in creating the responses provided by the bot, whether this is making sure that the bot simply makes appropriate responses to the user or to design the right conversational tone in terms of clinical formality.

This chapter ends with a discussion about how the chatbot is designed to be interacted with by users. In doing so it will show how ReMind must construct the users of the bot in order to provide their technical treatment. This means that the 'user' was not already out there in the world, waiting for the ReMind bot to find them, but is a subject which is actively created throughout the development and deployment of the bot. This chapter will show that the 'user' as constructed through ReMind, is one which is produced through ReMind's 'mass-personalisation'. This means that the user's experience is highly individuated and simultaneously standardised.

1. User-Led Design - ReMind's ethos of 'user-led', or 'user-centred' design means that the users are thought of as having some kind of priority, or even expertise, in their own treatment. This design approach is directed by ReMind's overarching ethos, which is to 'solve for' mental health. This choice of terminology is significant in that it implies a technical 'adjustment of means to ends' whereby a defined problem is responded to by constructing a technical solution. However, the problem is loosely defined by ReMind. 'Mental illness' is too specific as it implies psychiatric or psychological diagnosis, 'mental suffering' in a very general sense covers what ReMind are attempting to address. The undefined nature of the problem means that coming up with a solution involves experimentation, 'seeing what works',[230] and taking such metrics as 'user-satisfaction' as markers of whether their approach is successful or not. This leads to a situation where a design ethos is arrived at in which ReMind's mental health intervention is 'user-led', but this is defined by ReMind. This circular treatment method, and its relationship to an engineering or 'problem-solving approach' will be explored in this section.

[229] Marx. K. (1993) *Grundrisse: Foundations of the Critique of Political Economy*. UK: Penguin Classics. p.92
[230] Product Director, Arnold, interview

2. Technological Solutionism - This section discusses how ReMind needs to configure the ‘user’ in terms of a bearer of a problem which can be solved. The ReMind team and the users of its app are in a relationship, one which is mediated and also guided by the app. ReMind uses various feedback mechanisms to adjust how their treatment affects the users. This section will discuss how ReMind must approach their intervention as similar to adjusting parts of a machine or system. Their method for assessing whether treatment is effective or not, involving tracking user behaviour as they navigate the conversation trees, and then making adjustment to conversation nodes makes treatment into a machinic activity in which the whole system is brought to bear on the treatment of individual users.

3. Mass-Personalisation - This section discusses what kind of user is the consequence of this process. This means that the app, as in the bot and the various interventions it offers, is directed towards a particular subjective ‘type’, i.e. someone inhabiting a certain social position, holding certain attitudes, opinions and needs. This of course does not mean that anyone who does not fit into this type will find the app unusable, but rather that ReMind, throughout their design and implementation of the app, have come to understand their users as particular subjects, and, as in the previous chapter, increasingly design their app towards these subjects. A question to be pursued in this section will be: does this mean that this subject, as it becomes increasingly recognised and defined by ReMind, becomes increasingly more defined in the ‘real world’. In other words, what is the connection between ReMind’s image of their audience, and the ‘actual’ audience, and to what extent are they codetermined?

7.1 User-Led Design

Engineering Ethos

Conversation design decisions are made in order to better craft responses to user-inputs. ‘User-inputs’ means both their conversations with the bot and their direct requests to developers or app reviews. Employees have access to a large amount of anonymised data including user retention, engagement frequencies and rates, written reviews, population and geography data, self-reported physical illnesses, etc. Also, conversation tree paths and content choices/sequences of use. ReMind uses these data to make decisions about modifications to existing systems and whether to design new systems. User habits lead to design decisions which then go on to influence user habits. Throughout this feedback process, understanding about what the bot can and cannot do has come into focus. Conversation designer Alan spoke about how the technology behind the bot developed through a process of ‘active listening’ in which the bot identifies the contextual qualities of user inputs:

> I think it's a joint effort between AI and the conversation design team, where we look at past data to identify how we can respond better. Because when we first built something called ReMind 2.0, which has this capability of active listening rather than passive listening, that's when we took the previous data to identify different domains that we need to respond to more clearly and specifically, when we go back there, we look at things like, okay, this person has talked about a relationship, but this is a context in the past. It's not happening right now. Okay, but this person is talking

> about a relationship, but it's in the context of breakup. So let's expand that domain a little bit more. So that we build the specificity into conversations and make sure that it's appropriate to most cases in those specific modes.

This design approach means that ReMind can gradually expand the responsiveness of the bot, and that they can focus on specific instances of conversations with the aim of making alterations which can apply to a large range of conversations. It also means that the ReMind bot can develop through an iterative feedback system. ReMind employees often spoke about the design of their app being 'user-led', in that instead of treating a predefined mental health issue (such as 'depression'), the users of the app define the problem for ReMind to invent a solution for. The problem for ReMind is how to identify at which points in the conversations do improvements need to be made without having to manually access and assess conversations, a process which would be extremely laborious. This is where ReMind's 'traffic-light' system comes in: recall in the previous chapter how ReMind uses mass-user drop offs to signal points of failure in the conversation. ReMind uses these signals to experimentally redesign the bot's conversations. Jeff, one of ReMind's directors, spoke about how users might become categorised through their iterative feedback process:

> So if you get 10,000 people dropping off on something, then you solve the problem of how do I keep them and actually get them to do the exercise I asked them to do. If 10,000 people refuse to do that exercise,[231] you start looking at: what's the common characteristic of these 10,000 people? And then you realise they all have a certain common thing. And maybe for people like that this type of exercise doesn't work. So maybe we need to position it a little differently, then you go back to clinicians, and you say, is the exercise wrong? Or is the motivational track wrong? And then they will come up with two or three hypotheses as to either of those, and you'll see which one works better.

We can see an engineering methodology at work here: discontinuation of users at scale is taken as a marker for whether certain aspects of the app are functioning adequately or not. It can be visualised as a mechanic tinkering with a machine that they have built in order to get it running as efficiently as possible. Because ReMind has built a device for synchronous, or immediate, visualisation of users aggregated into large clusters, who navigate their way around the conversation trees, it is possible to make minute changes in the operations of the system and observe the results of those changes in real time. Samantha, ReMind's chief psychologist, summed up the phenomenon:

> I always say that ReMind is the world's largest co-designing experiment, because we launched it in the real world, unlike other bots that have taken birth in the lab in a university. So all that feedback that was coming in via PlayStore, on Apple Stores. All of that was being looked at, additions to the conversations, like…suggestions for what needs to be heard or empathised with.

ReMind takes an experimental approach to app design in that they do not have a predefined solution to a predefined problem: they are attempting to offer in a sense a methodology for

[231] 'Exercise' here refers to the bot's suggested self-help activities, such as mindfulness or breathing techniques.

how technical solutions to mental health treatment might proceed. The technology itself allows for this approach to happen; it was not necessarily assumed by ReMind that this would be their method. Jeff, one of ReMind's directors spoke about wanting to make an 'empathy bot' in the early stages of the company, and Arnold, ReMind's product director, also spoke about the bot being a side project in what was originally a self-help app with the bot's becoming unexpectedly popular being the stimulation to pursue developing a chatbot. According to Arnold, the bot came about due to a combination of wanting to develop an app that helped people in some way and users of the bot signalling the viability of a chatbot. Arnold spoke about how their design ethos tends to take an indirect route:

> We didn't start out by saying, "Okay, what are the three top conditions, and we are going to create an intervention which is going to solve for that."...The story of how we have evolved, it has been about how people express their sorrow or their misery in different ways. And just be there and just listen to them. And, for example, the top tools used in ReMind are not around depression or anxiety but around sleep…And again, sleep is usually the tip of the iceberg. There's lots of stuff going on, why people might have disturbed sleep, or sleeping too much, sleeping too little. And then when you dig deeper and deeper and deeper, and that's usually, almost the opening door through which a mental health conversation can start.

As with the requirement to control the bot's responses, an experimental and iterative approach to mental health treatment can be seen as a risky venture, but initial concepts for the app were not precisely to treat mental illness, and while development of the app is aimed at improving mental health, ReMind see themselves as achieving this through indirect means.

"Solving For"

ReMind's engineering approach is summed up in their term 'solving for mental health'. 'Solving for' is a term which implies that the problem itself is undefined, and that identifying and developing solutions are part of the process of defining and circumscribing the problem. Alan, ReMind's head of conversation design described the approach:

> We say we're solving for mental health. And right now we're adapting to solve it in a more clinical approach. But when we built ReMind, we basically wanted a place for users to feel safe, when they can't find it somewhere outside. Right, a lot of our human need revolves around somebody to listen to, somebody to make you feel like you're hurt and your concerns are valid. And this can be a very wide range of things.

What this means is that while an aspect of human need is identified (the need to be heard), building a solution to this (a bot that actively listens) is only an element in the problem to be solved. 'Solving for' means that ReMind, using 'listening' as their baseline, can experiment with various different technical, clinical and design methods to alleviate mental suffering. The bot is merely the current means of solving for mental health: Arnold spoke about 'Product-Market Fit', in which a consumer market is gradually identified through feedback and analysis. In other words, the product is secondary to the market in that identifying a consumer demand for a product or service and 'solving for' that demand is the primary

challenge in product design. We can see how this strategy has influenced both bot design and the choice of providing a bot in the first place for ReMind, on two different strategic levels. On one hand, there are various conversation design choices made in response to user activity, and on the other hand, there is the choice of approaching mental health in terms of a consumer market in the first instance. ReMind's 'user-led' approach, in this manner, can be equated with treating 'mental health' as a phenomenon which occurs in terms of a consumer market. In other words, by undertaking this approach, mental health and mental suffering can be placed in a conceptual framework in which treatment, or 'intervention', can follow the rules of business strategy. Customer satisfaction, and the means to achieving it, can now be an indicator and the method of treatment success. Business strategies vary from company to company; even concepts of 'the customer' and 'customer demand' take different forms.[232] In taking a business strategic approach, ReMind must undertake a prognostic attitude: the 'market' is to be read and interpreted in order to understand what 'it' wants. ReMind and other mental health apps share a single defining principle: that intervention is to be a technical one. What these businesses share is not their various internal strategies, but an ethos of "technological solutionism",[233] in which a problem, or market demand, is identified which can be framed in terms of providing a product which addresses the problem or market demand in terms of technical attributes. 'Technical' here is equated with 'scientific-technical rationality':

> Technologies fall under a dominant standard of scientific-technical rationality. Rationality implies the application of formal rules to some domain of experience. Such rules impose clearly defined all-or-none categories on experience, and tie these to principles of equivalence, implication, or optimization that relate to thought and action, and they do so with an unusual degree of precision.[234]

The user's activity on the app prompts the ReMind team to make adjustments; this is the basis of the 'user-led' approach, but ReMind has already designed the technical apparatus within which the user performs this activity. This means that the user is constrained and guided in terms of the possibilities of what they can do. The ReMind team is similarly constrained and guided by their own system. Traditional therapy can usually be characterised as a one-to-one encounter which is mediated by the therapists' training, their therapeutic organisation, their supervisors etc. The ReMind bot mediates the encounter in a more abstracted way: by acting as the interface between users and the ReMind team. In a sense, the encounter can be thought of as 'social' in that the ReMind team and the users are communicating through the bot. The communication is asymmetrical however: for users the conversation is experienced as a one-to-one dynamic between user and bot, for the ReMind team the conversation is approached as 'one-to-many' dynamic in which users are observed

---

[232] Steve Jobs, often considered as a 'Design Thinking' visionary, is famed for claiming to go beyond simply attempting to identify what customers (appear to) want: "Some people say, "Give the customers what they want." But that's not my approach. Our job is to figure out what they're going to want before they do. I think Henry Ford once said, "If I'd asked customers what they wanted, they would have told me, 'A faster horse!'" People don't know what they want until you show it to them. That's why I never rely on market research. Our task is to read things that are not yet on the page." Cited in: Grossman, J. 'Brand myths & legends: Jobs, Ford & Apple'. *Jell Strategy*. Online: https://www.jellstrategy.com/notes/brand-myths-legends-jobs-ford-apple (Last accessed 12/05/23)

[233] Morozov, E. (2013) *To Save Everything, Click Here – The Folly of Technological Solutionism.* New York: PublicAffairs

[234] Brey, P. 'Feenberg on Modernity and Technology'. Simon Fraser University. p.1

en masse using different systems like the 'traffic-light' system. ReMind conversation designer Alan described the development of bot conversation, which began under the rubric of content design but is now called conversation design: "not just think through what to write, but how to write it in a way that the consumer or the end user is impacted in the way we want it to come across, basically." What way do they want to come across though? As many ReMind employees have mentioned, their approach to design is user-led, so "the way we want it to come across" means ensuring that users are satisfied with the experience of using the app. The user is someone who expresses themselves in terms of what can be predicted by the AI: the linguistic models that the bot wields predetermine what the user will say (or is able to say), because interpretations (categorisations of user inputs) are already set by the human team who build ReMind's models. We can see that 'the user' is in this way projected - 'predicted' - by ReMind in an unusual reflexive manner: the bot continually adapts to the user but through a mechanism that is predetermined by the ReMind team.

User-Led

'User-led design' can be thought of in terms of the technical framework within which ReMind identifies the needs of users. As with the requirement to control the bot's responses, an experimental and iterative approach to mental health treatment can be seen as a risky venture, but initial concepts for the app were not precisely to treat mental illness, and current use of the app is not ostensibly, or immediately, with the aim of improving mental health (this may be so in an 'big picture' sense however). Reese, one of ReMind's directors, explained their design strategy as guiding the user towards their own answer:

> [I]n effect, it is echoing and gently guiding you towards an answer, which is always within you. And I think that's the core, if I think about it. I think that is the reason why it is able to work in so many different contexts. If I had a very clear prescriptive directive interaction, which I had in mind, saying, "Oh, the person who has substance abuse issues, and this is what I want to tell them", and we're obviously doing a lot of that now, but at its very core, ReMind is actually a very malleable and responsive space, which guides or moulds itself to where you want to take it, and where you want to go.

When conversing with the ReMind bot, the course of the conversation always arrives at one of the various self-help techniques provided by the bot, so the user 'chooses' through the conversational inputs and responses, which technique to use. This is essentially 'nudge theory' of design. Nudge theory comes from behavioural economics and has roots in cybernetics theory. It relates to strategies of manipulating or influencing behaviour in a non-direct way. Nudge theory has been likened by Pelle Guldborg Hansen to a form of "libertarian paternalism"[235] in which the 'correct' behaviour is chosen in advance, and subjects must be gently guided towards it. Of course, ReMind does not claim to be imposing a 'correct' form of mental health onto users, and they can assure themselves that they are merely responding to the users' activity to make adjustments to their system. However, we can understand nudging in terms of the users being guided towards an *attitude* which asserts that their mental health issues are indeed problems to be solved, and that their

[235] Hansen, P. (2016) 'The Definition of Nudge and Libertarian Paternalism: Does the Hand Fit the Glove?'. *European Journal of Risk Regulation*, 7(1) pp.155-174

behaviour can be used to indicate treatment success or failure. User inputs, and their reactions to bot responses are measured and form a way of visualising at what points the system is more or less effective. ReMind conducts standard outcome studies to verify its effectiveness, so it does not solely rely on the 'traffic-light system' feedback approach. However, the traffic-light approach is at the core of decision making in regard to how bot conversations are generated. This means that users have a role to play in the development of the bot, one which they are largely unaware of, and conversely ReMind does in fact guide users, but through a technical system which also guides the ReMind team's understanding of the users.

The user on one hand, is given agency over how to engage with the app in terms of choosing different modules, but on the other hand is confined, not in terms of the range of programs but in terms of how the bot categorises the user's complaint. This is an interpellating process: 'interpellation' meaning the transformation of an individual into a subject, effectuated by a discursive, or ideological, structure - "Ideology interpellates individuals as subjects."[236] While the user is interpellated and subjectivised by the app, they become observable to the ReMind team in terms of this process. When I spoke to employees of ReMind, I understood that they considered the users as the flesh and blood humans who are situated on the other side of the app; the way that the app mediates their relationship was only ever considered by ReMind employees in terms of how much user information could be gathered by the app in order to better facilitate their use of the app. But the user is also a construction, 'predicted' by the app in terms of its own conversational models. Throughout the conversation process, the subject of this encounter, the human user, is positioned as a series of inputs that can either be determined correctly or not. In this sense, the user is constructed by the bot as a model composed of a sequence of more or less correctly determined estimates. In order to be treated by the bot they must integrate into the treatment as an element in the system, to be 'adjusted' by the bot's psychologists and other designers. However, there is a circular aspect to this: the ReMind team makes adjustments according to user behaviour, which is conducted on an individual level; this behaviour is however observed and acted on at the macro level. Throughout this circularity we can observe a disparity in perspectives: the user is master of their own experience on an individual scale, choosing any app features they wish to access (and can afford), and saying whatever they like to the bot without worry of repercussions. Conversely, the employees of ReMind are constantly adjusting the individual user's experience in terms of overall adjustments to the app. Users are approached by ReMind in mass-form: as aggregated in terms of both the overall group of users who interact with the app, and in terms of different clusters. These clusters comprise two types. The first type is users identified by ReMind through linked conversational behaviours (e.g. choosing similar conversational paths) and the second type is a more common grouping: linked in terms of attributes or demographics (such as those who self-report as suffering from depression). The latter grouping method is the focus of this discussion, as it helps to identify the formal conditions which comprise ReMind, and through these conditions to develop a picture of the subject which is interpellated into ReMind.

[236] Althusser, L. (1971) 'Ideology and Ideological State Apparatuses. (Notes towards an investigation)' *Lenin and Philosophy and Other Essays*. New York and London: Monthly Review Press. pp.121-176. p.170

7.2 Technological Solutionism

Ad-Hoc Clusters

Treatment software can be reconfigured according to observations of the users' activity as they navigate the conversation tree, in order to improve treatment. This gives the operator of the software a sense that the outcome can be reached through manipulation of the software itself. 'Mental health intervention' becomes a technical action of adjustment. This involves looking at decontextualised user inputs to get a sense of the topic or issue they most commonly input which then generates the bot response which drives users to either continue or discontinue the conversation. A less triggering response, or one which signals that the bot is being more attentive than previously, or even just a response which is simply more appropriate to the specific topic is applied. If user retention is maintained, then the next blockage in the tree can be attended to. The grouping of users in this way is an ad hoc process, relying on observation of clusters of users as they move around the conversation tree. These clusters are ad-hoc because they are not predetermined and are used to indicate at which nodes in the branching conversation structure the intervention needs to be adjusted. There are no clusters taken account of at 'effective' nodes because clusters are only needed to indicate non-effectiveness. We can see that the treatment of mental health is in this way similar to maintaining the smooth running of a machine, or any human-made complex system. The system is machine-like because it is characterised by discrete measurability and control of the various systems and subsystems. Feedback is measured at various junctures, in this case feedback is derived by measuring the frequency and quantity of user drop offs in response to system-alterations. Alterations in the conversation mechanism can be made at those junctures, and feedback is again monitored to see if the situation improves. If it does improve this means that the system as a whole has improved. Treatment effectiveness is in this way judged in terms of the system as a whole. It is made up of individual users but those users are aggregated into ad-hoc clusters which either flow through or get stuck at different conversation branches. We can understand the two levels of observation here as interlinked: sub-groups of users are identified through their linked activities in the conversation tree; alterations to conversation nodes affects their activities which are then assessed as various instances of the whole system, which comprises all users of the app.

ReMind does not need to define the problem that they are developing a solution for, because the technology inserts itself as the solution *prior* to the problem being defined, and thus is the starting point for any kind of definition-making. The economic and technical conditions in which the ReMind team finds themselves frame the problems that they are solving for. In a social system both increasingly dependent upon technological development and increasingly organised as a technical system, responses to what could otherwise be considered as socially, politically, or morally defined problems, such as caring for (or as is the case, dealing with) those who suffer from mental distress, are increasingly considered under a technological paradigm. Technical solutions, which involve efficiency, control, the 'adjustment of means to ends', etc., are now routinely applied, outside of the industrial context in which they originally emerged, to the management of subjects for the purpose of reaching a particular goal. The efficient reaching of a particular goal - whether that be an increase in power generation in a factory, or the reduction of mental illnesses in a population

demographic - can be endorsed above any other measures. When the various networks of social relations that combine to form the social body are regarded as comprising a technical system, human subjects within those networks can be considered as technical objects. Within this regime, human relations "assume increasingly the objective forms of the abstract elements of the conceptual systems of natural science and of the abstract substrata of the laws of nature."[237] When human subjects are regarded as technical objects operating in a technical system, judgement can be applied according to such criteria as measurability and efficiency, and any defect in efficiency, such as poor mental health, can in turn be regarded as having a technical solution. When mental health is considered as a technical problem, a technical solution can be applied, regardless of whether this technical solution is provided by a human or by a machine.

Objectivity

In taking a stance of 'solving for', ReMind makes the assumption that 'mental health' is a technical problem which can be addressed via technical solutions. This does not mean that ReMind considers their intervention to be solving *all* mental health conditions, just the ones that are 'solvable'. A solvable problem is one which can be precisely defined, but also crucially determined as containing within itself a solution. The simplest example would be a mathematical problem: the solution is 'already there', ready to be worked out. A more complex example could be designing a traffic light system for a grid of roads: it may require an inventive and creative approach but the solution itself is clearly defined as the most efficient routing of traffic. This kind of problem-solving approach, when applied to mental health, must in some way define an outcome so that any possible solution can be judged as either helping or hindering the achievement of a solution. There is a circular logic at play: in order to 'solve for' mental health, mental health itself must be approached as a 'ready-made problem', i.e. as containing the solution within itself. The ReMind app, in appearing to present a novel means to address mental health, is confined by the technical mindset which cannot conceptualise novelty in terms of treatment styles. This equates to an objectification of mental health: in order to be cast as a definable problem with a definable solution which can be 'solved for', mental health must be understood as, in some way, having externally verifiable, objective qualities which can be measured. Otherwise, it would not be possible to claim that the proposed solution is having any effect. Arnold, ReMind's product director, identified this when discussing how the app can be externally adjusted in order to diminish the need for human-based intervention:

> I mean, see, I think that you are taking the human out, but like you said, they are being rearranged, but we are often creating a system that requires less people to be involved. But I think what's critical, in many cases, that's safer. You know, one of the things that people ask me is, hey, what if ReMind misinterprets a certain sentence and therefore responds in the wrong way? My response is, actually humans do that all the time. The difference is if you're, let's say you're running a tele-medicine, medical centre, and somebody misinterprets that, you know, they hear a new word right? Culturally, a new word emerged. Everyone in the call centre thinks, or the tele-centre thinks, that means one thing, but actually…it means something else, everyone's misinterpreting that, it's kind of the same thing as would happen with

[237] Lukács, G. (1991) *History and Class Consciousness.* UK: Merlin Press. p.131

> ReMind. Now the difference is, we can once you pick that up, and what's important is you pick it up. But once you pick it up, it can be resolved and the fix is immediate. From that moment onwards, the system reliably responds to that context, well, whereas the tele-mental centre…you actually have to retrain everybody, and there will be an inconsistency for a while before you then get consistent output. So that's where you know automation is wonderful, right? For things that are robotic and really are monotonous. They should be automated, because that's poor use of human potential.

ReMind frames mental health treatment in objective terms; this objectivity is not deemed to be a necessary characteristic of the ReMind app, but in terms of mental health itself. The above quote shows the belief that even if humans were performing the same task as the app, due to the mechanistic quality of that treatment, the only distinction between human and machine activity is that efficiency is maximised when undertaken by machine. In *Ten Paradoxes of Technology* Andrew Feenberg writes that technological artefacts are inseparable from their contexts, despite the common-sense idea that technology is something which is easily transplanted from one context to another. By this he means that at a basic level, parts of a mechanism are inseparable from that mechanism itself. He gives the example of the wheel of a car - if the wheel is taken off the car, neither will function in the way that they are supposed to. But each component of the car is thought of as distinct and separate from every other component and exhibits a presence of wholeness and detachment which denies their contextual and connected necessity in order to function correctly. This contextual necessity can be extrapolated endlessly: cars are dependent on the existence of roads, which are imbricated in a whole range of social, political, and economic dependencies. Yet, each aspect has the appearance of *in*dependence. According to Feenberg, the 'separateness' and transformability of technical components in spite of their contextual dependencies is a consequence of the de-contextualisation inherent in a capitalist economy:

> The generalization of this feature of the theory leads us back to its remote source in Marx's distinction between exchange value and use value. There is a gap between the concrete reality of goods and the laws of their economic circulation in a capitalist economy. The price under which things are exchanged governs their movement, often independent of use, rather than the immediate connection between the producer and an individual consumer as in former times. Similarly, functions float free from the wider context of the lifeworld and appear as the essence of artefacts that may in fact have many other relations to the human beings who live with them. The fetishism of function obscures these relations much as the fetishism of commodities masks the human reality of the economy.[238]

Users of ReMind are rendered, by their activity as observed in the conversation tree, as data-points moving from node to node. They are de-contextualised from their lifeworld in a similar manner to the above description, precisely as autonomous functional units which can be observed by way of a singular feature: their movement throughout the conversation tree. Head of clinical research and development, Charley, spoke about the kind of users who would be attracted to app-based treatment:

[238] Feenberg, A. (1999) *Questioning technology*. New York: Routledge. p. 21

> [T]he individuals who are organically attracted to use a methodology like this [CBT/self help-style treatment], for their mental health are potentially already people who are either feeling isolated, feeling like they don't have any sense of community, don't have anybody to reach out to. And this then becomes the least friction methodology of accessing any support. Because the odds of these two things having a high level of overlap in some kind of a Venn diagram, between individuals who are struggling with their mental health and individuals who are feeling alone and without support is really, really high.

ReMind has identified their user base as (potentially) those who are socially isolated: 'autonomous' in terms of social-decontextualisation. ReMind confirms this isolation on a formal level, by rendering users as units which traverse the conversation tree. In order to make a system which treats mental health in an autonomous fashion, 'mental health' must be rendered by ReMind as exhibiting objective qualities, i.e. it must be made externally manipulable via technical means. External manipulability is dependent on understanding one's object as being made up of discrete and non-contextual units: as being isolated from any social basis. In other words, the means for approaching an object as technical, i.e. that which can be adjusted, edited, manipulated or otherwise transformed, is social decontextualisation. The users who are most attracted to this software, according to ReMind, are those who are often the most socially isolated, users who can't or won't depend on their social basis to tend to their mental health. Not only are users drawn to use the ReMind app because of social isolation, they confirm this isolation on another level: by becoming transformed into technical units, the purpose of which is to be observed circulating throughout the ReMind conversation system.

Formal Bias

The ReMind app is not just useful to users who are socially isolated, it is designed in a way to attract those users, it is, in its formal structure, biased towards those users. This means that the technical basis, and not just the outward contents of ReMind, favours and generates isolation. Andrew Feenberg calls this type of bias "formal bias"[239] meaning a bias towards a particular social group which is inherent to the structural conditions of rational systems. A problem which often arises in discussions on computational technologies is 'bias', whereby software is unintentionally (or intentionally) infected with human biases. This is often conceptualised as a problem of incomplete or skewed training models, a lack of diversity on the development team, or some other external factor.[240] According to UI/UX designer Daniel, ReMind mitigates bias through, on one hand, the diverse nature of the user models appropriated and developed by ReMind, and on the other hand, the diverse nature of the team involved in building the bot. Daniel explained this in terms of making sure that their data includes enough of a range of social indicators that bias is eliminated:

[239] Feenberg, A. (2010) *Between Reason and Experience: Essays in Technology and Modernity*. USA: MIT Press. p.163

[240] Schwartz, R. Vassilev, A. Greene, K. Perine, L. Burt, A. Hall, P. (2022) *Towards a Standard for Identifying and Managing Bias in Artificial Intelligence.* National Institute of Standards and Technology

> ...they make sure that all the data they fit into the code comes from different genders, different locations, different social classes. So then there is no bias in the code, in the algorithm.

This is a method of automating the removal of bias. Removal of social bias can be achieved in two ways; one would be to manually vet every response being delivered by the app which would be practically impossible due to the scale of the responses being delivered. The other way is to feed the AI with a diverse range of conversational models, automating the process so that the potential for inappropriate responses can be minimised. These models are operative in the bot's conversation interpretation tools, as this is the only part of the bot that uses machine learning. ReMind, in designing their app in a reactive way, i.e. by responding to user activity, has to make sure that the bot maintains a neutral stance on not just divisive social issues but any and *all* markers of sociality. For example, the ReMind team is careful to make sure that the bot does not respond in which it might be seen, as Daniel mentioned, to favour a particular gender, or social class. By incorporating a diverse training model they aim to develop a 'genderless' and 'classless' conversation model. The bot interprets the users in terms of neutral or unbiased training data, meaning that not only can the bot not act in terms of bias, it also cannot 'hear' in terms of bias. Louise Amoore points out that the use of software to mediate social relationships does not just raise the problem of making sure that the software does not perpetuate biases, but that the use of software involves "establishing new patterns of good and bad, new thresholds of normality and abnormality, against which actions are calibrated."[241] The assumption that technology, on a fundamental level, is socially neutral, and that the use of technology, so long as bias is eliminated, is a socially neutral activity, ignores the fact that technology frames how we understand bias in the first place. We can understand this framing in terms of a tendency, which mental health apps ascribe to, towards a form of universal treatment which is applicable to anyone. This tendency is also at play in ReMind's project of creating different language versions of their bot.

As discussed in the previous chapter, head of AI Mark explained how they use a translation function which converts another language into the bot's own language - English - before performing its automated interpretation. This process is then reversed to provide a response. There is a sense that it does not matter what language is used, the method, or function, always stays the same. English happens to be the dominant language which other languages get translated through; what is important to understand is that one language has come to act as the medium through which all others are interpreted, translated and understood. The fact that it is English is due to historical circumstance: English (specifically "printed English"[242]) is the foundational language upon which *all* information technology is built, the formal condition upon which other languages occur as secondary 'content'. This is a foundational level of exclusion that occurs as part of the formal structure of this technology, and is a necessary condition for the appearance of universality. In other words, to project a sense of universality and general applicability, the intervention depends on an unacknowledged 'bias' towards a singular dominant basis: one that operates 'below' such

[241] Amoore, L. (2020) *Cloud Ethics. Algorithms and the Attributes of Ourselves and Others.* USA: Duke University Press Books. p.6
[242] Shannon, C. (1953) 'Prediction and entropy of printed English'. *The Bell System Technical Journal*, vol.30, no.1. pp.50-64

exclusions as social bias being encoded into the technical device. This is essentially the opposite end of the bias-elimination spectrum: the assumption that there can be a 'bias-fee' technology is predicated on an unacknowledged systemic bias. The bot inhabits a world of social non-bias, and due to its interpretive mechanism, projects onto the users a disposition in which bias is (eventually) eliminated: essentially responding to a cohort who themselves do not exhibit any social biases. In other words, the quest to eliminate bias projects the assumption that the users themselves do not exhibit any social biases, and are in fact socially 'neutral'.

7.3 Mass-Personalisation

Scalable Treatment

We can see that in order to develop a product which can be mass-produced, i.e. is scalable, the removal of bias is both desired and necessary. It is desirable on ReMind's terms because new users from diverse backgrounds won't be put off, but also necessary in technical terms because as more and more users interact with the app, the user base as a whole assumes an increasingly generic identity due to the need to address all users simultaneously. The more users the app provides for, the less possible it is for individually applicable responses to be given. Director Reese spoke about how the bot, while unable to provide 'tailored' intervention, can still be considered as providing unique experiences:

> That conversation is unique. It is not standardised. The way it is being delivered is standardised, or maybe the response set in stone is standardised. But that specific interaction is unique. And that's unique. And I remember in earlier lives, we used to have this concept called mass-customization, saying that it is a completely customised interaction, but it's…operating at mass scale. And with something like ReMind, we have achieved both and it's the Holy Grail for most solutions. Achieving mass customization is impossible, you either have a standardised McDonald's approach towards something, or you have an extremely custom approach, which is one-on-one and completely unique to the individual. But combining both and creating custom or customised interactions at scale is very special. And it's unique. And I think that is the core of why different sets of people use it in very different ways for their own different contexts.

According to ReMind, they are creating a mental health intervention that is both scalable and customisable insofar as the core benefits of the app are enjoyed no matter how large the user base gets. This is a "mass-personalisation"[243] approach to product design, in which customers or users are not only able to customise the products that they purchase, but are also involved in some way in design. ReMind's 'user-led' approach can be equated with mass-personalisation. Their solution to mass-personalisation is computerised automation. This means that the 'personal' aspect of their app is due, not to the range of conversational options available to the user (customisation), but to the way that ReMind track and respond to their users in order to make design decisions. The users do not have direct involvement in

[243] Aheleroff, S.Mostashiri, N. Xu, X. & Zhong, R.Y. (2021) 'Mass Personalisation as a Service in Industry 4.0: A Resilient Response Case Study'. *Advanced Engineering Informatics*, Volume 50, 2021, 101438

design, but are abstractly involved. There is a recursive dynamic at work which depends on all aspects of the ReMind app being amenable to adjustment: old features should be editable, new features must be possible, conversational nodes must be editable but also re-wireable. One node which leads to another should potentially be able to lead to any other node. In other words, the app itself has an ambiguous and protean presence: it could be completely different, or even non-existent in years to come, depending on both user feedback and corporate requirements. Recall Arnold's comment above about the ease with which it is possible to make instant changes to the software and roll them out almost immediately:

> But once you pick it up, it can be resolved and the fix is immediate. From that moment onwards, the system reliably responds to that context, well, whereas the tele mental centre…you actually have to retrain everybody, and there will be an inconsistency for a while before you then get consistent output. So that's where you know automation is wonderful, right?

While not specifically about therapeutic consequences, a Vice Magazine article by Samantha Cole draws our attention to what happens when the companion chatbot Replika is altered after users have formed relationships with it. The article discusses how users of the chatbot suffer the consequences of sudden changes in app design, and due to the always available 'service' nature of the product, cannot decide to revert to previous iterations of the app.[244]

App changes can occur on all levels: user interface design, the layout and categorising of various modules, the form of treatment itself. Arnold, ReMind's product director, spoke of the bot initially being a derivative product of a physical health app, and so the ReMind bot itself is a completely different product to what was initially designed. Because the ReMind app can (and does) change at any time, the user has no option but to constantly adjust to the changes. Adaptation is twofold: the style of treatment promotes adaptation to one's mental or social conditions, and the means of provision demands adapting to the techniques in order to perform this feat. As discussed in the previous chapter, ReMind observes user activity in a synchronic mode, enabling a conceptualisation of their mental health intervention as analogous to the construction and maintenance of a machine. This mode is an 'eternal present' in which history does not intrude and, as such, 'mental health intervention' is experienced by the users in the here and now, without reference to how the past might influence their suffering. In this way, 'personalisation' can be thought of as 'individuation': the forming of subjectivity, this formation is in response to an ever-changing and unpredictable interlocutor. Recall the user review: "It's a good app if you understand how to explain yourself to AI."[245] Adaptation becomes the 'ready-made problem' for AI treatment, in that the software appears (or is ideally) adaptable to the needs of the user, but it is the user who must become adaptable to the app. This process operates similarly to 'levelling' in that it involves a flattening of semantic categories, but it works in the opposite direction: by making sure that the bot is 'unbiassed' through increasing the diversity of its models and design

[244] Cole, S. (2023) 'It's Hurting Like Hell': AI Companion Users Are In Crisis, Reporting Sudden Sexual Rejection'. *Motherboard, Tech by Vice*. Online: https://www.vice.com/en/article/y3py9j/ai-companion-replika-erotic-roleplay-updates (Last accessed 10/12/23)
[245] Google Play Store Review

team, individual quirks, surprising encounters, spontaneous sayings - things that are associated with concepts of individuality - are removed. The treatment that the bot delivers becomes more universalised towards a global ‘mental health’ in which the ‘personal’ takes on a generic and universal quality.

Macro-Treatment

The user understands that they are speaking to a bot, and that their feedback helps the developers to improve the system. The user may not be aware that the measurement of conversation flows and blockages also affects how their treatment is determined: their behaviour in aggregate determines how their own treatment is adjusted to then provide them with a more satisfying experience. Users are in a strange way the masters of their own experience, not just because therapeutic intervention is confined to a self-help style of treatment, but because their behaviour sets the conditions for determining whether treatment should be adjusted or not. However, it is because of the fact that they are unaware of their influence that it is possible to use their behaviours as determinants. Unawareness is conditional on their being aggregated into data clusters: individuals might more or less have a sense that they have an influence on bot conversation development, but they are unaware of the experience of being aggregated into a data cluster, which is not something that is ‘experienceable’ per se. The result of this is that the user becomes split between experience and abstraction, the encounter with the bot is the source of the user’s sense of undergoing treatment, but treatment also occurs ‘on another scene’, as manipulable abstractions across the many nodes of the bot’s conversation tree. The expertise of the clinical practitioner is still required to intervene in the treatment of the users, but it occurs on a level of systems-adjustment. This form of treatment is characterised by asynchronous interaction: while the user’s immediate engagement with the bot occurs as a real time conversation, this conversation is monitored and managed on an abstract and post-hoc level.

This could in a way be compared to how in traditional therapeutic settings the therapist discusses their case with a supervisor or supervision panel and adjusts treatment based on their feedback. The difference here is that feedback is measured directly from the conversations themselves. Users do not directly experience the treatment which is being administered by the clinicians through the app; they only directly experience the conversation with the bot. This might be compared to the intervention of a supervisor who recommends adjustments to a clinician’s intervention in a face-to-face setting, but this adjustment involves a singular, specific patient, whereas adjustment for ReMind involves the overall system, with intervention then affecting all users. They may well be aware that there is a team of people intervening, but due to the layers of abstraction involved, treatment is split between that immediate experience and the intervention of the clinicians. The user is both alone and not alone: they are talking to a non-human robot, but that talk is visible to many other people. The user is heard and unheard: the bot is programmed to provide a sensation of being listened to, but ‘listening’ involves being interpreted and framed as a generic and ‘unbiased’ non-social being. On the other hand, the user’s conversation is constantly monitored and measured by the development team, but being heard in this case means being responded to in terms of adjustments to the overall system. The user’s experience is that of abstraction, but rather than an ‘abstract experience’, the abstraction

occurs in formal terms as mediated through the bot: it is an abstraction which is not quite 'experienceable' but occurs as an effect of technical treatment.
It is a "real abstraction"[246] in that it is not some a priori conceptual abstraction, but one which emerges as an effect of the activities of user-interaction, aggregation of users through observational tools, and bot responses. Alfred Sohn-Rethel introduced the term 'real-abstraction' as a means to describe how abstract concepts (such as 'value') emerge as social abstractions prior to the possibility of their conceptualisation.

> The essence of the commodity abstraction, however, is that it is not thought-induced; it does not originate in men's minds but in their actions. And yet this does not give 'abstraction' a merely metaphysical meaning. It is abstraction in its precise, literal sense ... complete absence of quality, a differentiation purely by quantity and by applicability to every kind of commodity and service which can occur on the market.[247]

We can attribute an abstract notion of mental health to this process: the users, in their immediate interaction with the bot, are not aware that they are being treated en masse, because their activity is bound up in a 'personal' conversation with the bot. Users would obviously be aware that there are many other users engaging with the bot, and of the possibility that an identical conversation could be occurring with another user. However, their interaction with the bot still occurs as a phenomenal and 'real time' experience, they cannot experience the intrinsic social aspect of their 'real time' interacting with the bot. The social aspect of their experience - that of being in a cohort - is abstracted away from the possibility of experience through ReMind's aggregation methods. In this way, the user's social world (and every user's social world), in terms of the use of the app, is decisively siloed from their immediate interaction with the app. Users can of course understand on an intellectual level that there are other users out there, and there are online forums where users discuss their experience of the app, but these groups operate externally to the interactions that users have with the app itself. What they cannot experience is their group formation and transformation during the instances of app-interactions. This siloing of the social, at the very instance of engaging with the app, means that the experience of using the app is abstract. This is not just because the activities it suggests are 'abstract', or that the bot mediates between the users and the ReMind team, but that there is a materialised abstraction occurring at all times during the user-bot interactions. In other words, it is a non-metaphysical abstraction of the 'individual' as a non-social entity. The users, through their interactions with the bot, and facilitated by the ReMind's aggregation methods, generate their own splitting away from the possibility of phenomenally experiencing their activity as being performed in terms of a social cohort. The user is segregated in a way that the possibility of integration is foreclosed, this confers a particular identity on the user, one which is both unique and generic.

[246] Sohn-Rethel, A. (1978) *Intellectual and Manual Labour*. London: Macmillan Press
[247] Sohn-Rethel, A. (1978) *Intellectual and Manual Labour*. London: Macmillan Press. p.20

Algorithmic Identity

The users conduct their activity in two registers: one register is their immediate experience of speaking to the bot and carrying out the suggested mental health activities, the other register is their movements around the conversation tree which is monitored and aggregates users into clusters. These two registers are analogous to the psychoanalytic conscious/unconscious registers. The theory of the unconscious posits that conscious 'thoughts' are made unconscious not because they are buried; their inaccessibility is due to a transformation in which they attain a formal and structural character which cannot be 'experienced' as such. The unconscious forms the structure from which thought is possible in that it designates the most basic elements of consciousness. In Lacanian terms this is described as the unitary elements of language: binary distinctions. In terms of ReMind, the activities of the users become subsumed into a structure: the conversation tree. This structure is arranged according to the activities of the users, and in turn, it conditions the activity of users: their future actions are dependent on (but not guided by) the formal structure of the conversation tree. This involves ReMind employees redesigning both the conversational content found at each node, and the overall nodal structure. Users are 'out of the loop' to a large aspect of ReMind's operation, this is not because they are denied access to an aspect which they might possibly gain access, but, due to ReMind's design and implementation, this aspect is inaccessible in terms of observable experience. Users are interpellated into a system which is inaccessible to experience, and subsequently experience their mental health in terms of this lacuna. The formal bias which pervades ReMind as a technical system comprises the basis of this lacuna: users are unable to experience ReMind's aggregation methods, but are nonetheless included in various cohorts. These cohorts, or ad-hoc clusters, are gathered by ReMind in order to signal when and where the system needs adjustments. ReMind's overall intervention then changes over time to facilitate these cohorts. The intervention that users experience changes over time, gradually, as a smoother flow through the conversation tree. Users, due to, but unaware of, the own actions, experience a more efficient sense of 'treatment'.

The user is subjectivised not just by their interactions with the bot, but with their interactions in the social body in which the app is a mediating aspect. According to Sohn-Rethel, the way that society is organised prefigures the way that people think. This does not equate to the contents of that organisational form but the form itself - for example, in an exchange society it is the quantifying abstraction of exchange that prefigures conceptions of quantification, rather than the objects themselves being exchanged in some way conditioning how people think. ReMind's formal splitting of individual user-experience from mass application of their intervention means that there is enacted, on a social level, the understanding that users are indeed quantifiable units, amenable to ReMind's observation and adjustment. This understanding is not 'in the minds' of the users but in their actions as they interact with the bot. Jacques Lacan offered the concept of the symbolic in order to understand the manner in which humans become interpellated as 'subjects'. Lacan refers to the way in which this interpellation begins prior to any kind of subjective involvement, giving the example of how a name is given before birth, and how parents fantasise about the future of their child. As a consequence of this, the symbolic aspects of subjectivity have a retroactive effect; there is a sense of having to catch up to a predetermined meaning, one which has already been set and must be interpreted after the fact. By approaching their invention as 'user-led', ReMind

misapprehends their location in the design sequence. ‘User-led design’ here comes to connote the circular dynamic between the app and its users in which chief psychologist Samantha’s claim to the “world's largest co-designing experiment” makes sense. However, due to ReMind inculcating itself, however unacknowledged, into its own technical conditions, and increasingly projecting this technicality onto the users, the result is that users of the bot are required to adapt themselves to the intervention provided. In other words, it is unclear that the intervention is misrecognised by the makers of the intervention, and as such the users undertake a form of treatment which is manifested in the ‘solutionist’ ethos of ReMind, rather than the solutions themselves. The users of the bot are required to conform to its logic of intervention, not because of any explicit demands of the treatment methods (to clean one’s room for example) but because of the means through with the intervention is delivered. This can be described as ‘adaptable’ in that the mode required to adapt to is not defined in any specific way, but rather due to its very *undefinition:* as a universalised bearing towards adaptation. Instead of having to adapt to any particular treatment method or style, users must undertake a general *attitude* of adaptation.

7.4 Conclusion

A similar dynamic occurs between ReMind and its users as discussed in the previous chapter, which was concerned with how ReMind’s internal systems come to reflect back on their understanding of these systems. This dynamic involves, as with the previous chapter, a ‘looping’ effect in which the app is designed with some idea of the user-base, affects those users, who in turn alter their behaviour to use the app in various ways, which then goes on to alter how ReMind thinks about and designs their intervention. ReMind directs its app towards a particular kind of user: one who suffers in terms of their mental health of course, but also that their suffering is of a particular quality. This is not an intentional strategy on ReMind’s part: their approach is guided by the technical means by which their software is designed and deployed, and in turn, by which users come to interact with the app. This user is one who is unable, for different reasons, to get the help they need, whether this is due to social, economic or circumstantial reasons. Ultimately, ReMind view themselves as stepping into the gap left open by a lack of available treatment. Their design ethos reflects this: ReMind’s concern is how to get treatment (or rather intervention) out to as many people as possible, and hopefully without degrading the quality of treatment.

What ReMind attests to is that the app does indeed respond to a demand that has otherwise been unmet. On ReMind’s own terms the app does not purport to ‘treat’ mental health illnesses, but rather ‘intervenes’ in terms of a non-judgemental and caring conversational dynamic and self-help technique recommendations. It is not quite a lack of mental health that the app is responding to but a lack of care; a need to be cared for. Charley explicitly pointed out that many of the users of the bot are simply lonely, and what the bot offers involves a combination of its companionship and its offerings of self-help techniques: the bot has something to give. ReMind (and other app-based interventions) sometimes avoids the use of the term ‘treatment’, or ‘therapy’.[248] This is contradicted by promotional claims[249] made to

[248] In interviews, Charley, ReMind’s clinical development and research lead, tended to use the term ‘intervention’, which has a non-clinical undertone.
[249] This claim from Woebot is indicative of how app-based treatments are promoted:

illustrate how their apps are equal to and sometimes superior to their non-automated alternatives.[250] In promotional literature and in my own conversations with employees, there is a sense of equivocation between modesty and ambition for the capabilities of this new technology. However, we can detect an ambitious drift into omnipotence in the way that ReMind, in an unacknowledged way, constructs the users who interact with their app.

In constructing the users of the bot as the means to deriving feedback in terms of their activity, ReMind has transformed users into data points: abstractions from which to gather information. This is only possible through the construction of a macro-treatment such as ReMind have built, as it is necessary to employ technical means to aggregate and coordinate such a large number of users. This method of mental health intervention is unique in that while decisions are indeed made in response to individual circumstances, those individual circumstances are aggregated into larger data clusters. Those clusters represent moments in the conversation where the therapeutic effect is judged as either effective or ineffective. Clinicians at ReMind are able to observe the users of the app as both individual patients and as aggregated clusters of data points. They are able to manage the experience of the users through making alterations to the system and observing the results of these alterations. Their interventions are made on a level which is abstracted from the individual experiences of the users. ReMind also operates under conditions of abstraction in that the method of intervention that they themselves have developed conditions the means whereby ReMind approaches 'mental health'. In other words, ReMind, in building a 'scientific' intervention cannot consider how they have conflated technology and science in terms of measurable accuracy and in doing so, are producing a generic and homogenous type of treatment as opposed to 'personalised treatment'. This is because the statistical procedures necessary to develop an objective method of treatment depends on generic measurability, definability and, crucially, reproducibility. In order to create a reproducible form of intervention, any factors which stand in the way must be removed. In other words, anything that asserts a unique presence in terms of the deployment of their technology cannot but hinder ReMind's strategy.

"Our highly researched and intelligent therapeutic solutions create space for personal growth by delivering mental health care that people actually like to use. With a growing library of products and solutions tailored to specific mental health needs, we're bringing mental health care to literally everyone." Online: https://woebothealth.com/about-us (Last accessed 20/12/23)

[250] X2AI and other app companies tend to stress their associations to 'traditional', non-automated treatments:

"Studies show that computer-assisted therapy and a conversational chatbot delivering cognitive behavioral therapy (CBT) offer a less-intensive and more cost-effective alternative for treating depression and anxiety." - Fulmer, R. Joerin, A. Gentile, B. Lakerink, L. & Rauws, M. (2018) 'Using Psychological Artificial Intelligence (Tess) to Relieve Symptoms of Depression and Anxiety: Randomized Controlled Trial'. *JMIR mental health,* 5(4), e64. p.1

Chapter Eight: Suspension of Disbelief

In the case of the failure of absence, the question concerns the existence of agency as such. Is there a deliberative agent here at all?[251]

A common question about chatbot therapy is 'does it really work?' One interpretation of this question is in terms of whether the relationship between the user and the bot offers in some way a relationship within which the user gains relief from their mental suffering. Any therapeutic encounter involves some kind of interpersonal relationship - a dynamic within which each participant acknowledges the other. This chapter involves a discussion about how it is possible to engage with computerised agents or avatars on an interpersonal level, forming a subjective engagement with relational artefacts.[252] This will begin by exploring the psychic mechanism involved in anthropomorphism and the suspension of disbelief in order to build a picture of how chatbots offer a genuine sense of engagement for the user. Mental health apps incorporate various methods to engage the user, from design and language considerations to the use of personable avatars. While the development of a 'virtual therapist' which performs the same function as a human may not be feasible, the designers of mental health apps are concerned with understanding just how a user interacts with their apps on an interpersonal level. Their methods of reaching understanding usually involve attempting to measure the interaction in some way, rather than looking at what it means to 'interact', especially within the therapeutic environment. This chapter will ask what it means to build a computerised avatar which provides mental health treatment and to make comparisons with its human counterpart: to claim "human-level bonds".[253] While there is agreement that users of treatment apps form emotional attachments with the software, and that it is important to design apps that are neither totally devoid of personality nor descending too deeply into the 'uncanny valley'[254] questions remain over what is occurring on the psychic level when a user establishes an interpersonal bond with a computerised avatar, or even with a non-avatar computer application. Ultimately this chapter will address the question 'why a bot?': why the need to mobilise a tremendous amount of human and technical resources to create a responsive avatar, a 'conversation agent'[255] which provides assistance to the user. If ultimately the treatment is 'self-help' style, then what purpose does the bot serve? Why deliver the intervention through a conversational agent?

1. Imitation Games - The chapter will begin with an exploration of how ReMind and other therapy chatbots presents itself to users - precisely as a robot, rather than attempting to fool users into thinking that it is 'real' or 'sentient' in any way, or that it is 'thinking'. ReMind, along with Woebot and others, makes sure to let users know that it is merely a bot which can only respond in a bot-like manner. What is interesting to note is that users often do not engage with the bot in terms of its 'robotness' but as a companion - an interlocutor which responds in

[251] Fisher, M. (2016) *The Weird and the Eerie.* UK: Repeater
[252] Turkle, S. et al (2006) 'Relational artifacts with children and elders: the complexities of cybercompanionship.' *Connection Science*, 18:4. pp.347-361
[253] Darcy, A. Daniels, J. Salinger, D. Wicks, P. Robinson, A. (2021) 'Evidence of Human-Level Bonds Established With a Digital Conversational Agent: Cross-sectional, Retrospective Observational Study.' *JMIR Form Res,* 2021;5(5):e27868
[254] Ibid.
[255] Ibid.

a dynamic and thoughtful way. ReMind is aware of this dynamic and employees speak of treading a fine line between making a ‘humanly’ bot which is friendly and encouraging, while also making sure not to try to dupe the users into thinking that the bot has feelings for them.

2. The Uncanny Valley - Therapy chatbots do not try to convince their users that they can do anything more than provide robotic responses and self-help tools. This is due on one hand, to safety concerns about the unpredictability of generative responses, and on the other hand, to the need to maintain a paradoxical deception. This deception is borne out of explicitly informing the users that the bot is just that: a bot. I will argue that this transparency is a component in the necessary deception involved in chatbot therapy. Roboticists speak about the problem of the uncanny valley, in which if a robot appears as too realistic it begins to exhibit an unsettling quality which can be off-putting for observers. This problem also concerns therapy chatbot makers: the bot cannot plunge into the too-realistic realm of the uncanny valley as this will prove unsettling to users. The more sophisticated therapy chatbots become, the more the uncanny will impinge. This is because the uncanny is a confrontation with being observed or known. In order to have a convincing appearance of sentience and to have a therapeutic effect, the chatbot cannot ‘too perfectly’ simulate a human; it must display its non-humanness for the users to be drawn in. Imperfection is an ‘entry point’, but also symptomatic of an ‘unknown knowing’.

3. Intersubjectivity - Mental health apps incorporate various different methods to engage the user, from design and language considerations to the use of personable avatars. While the development of a ‘virtual therapist’ which performs the same function as a human may not be feasible, the designers of mental health apps are concerned with understanding just how a user interacts with their apps on an interpersonal level. Their methods of reaching understanding usually involve attempting to measure the interaction in some way, rather than looking at what it means to ‘interact’, especially within the therapeutic environment. We can gain an understanding of how the ReMind company views what it means to be social from their discussions about the user-bot interactions. We can also - due to the fact that the bot does indeed facilitate some kind of social interaction - gain an understanding of a nascent form of social interaction in which the ‘other’ does not exist.

8.1 Imitation Games

Relational Artefacts

The ReMind app comprises two conceptually distinct but practically interconnected features: the bot, and the self-help tools. They are conceptually distinct in that the bot can be interacted with simply as a conversation partner, without accessing the self-help tools. However, they are interconnected because this would ultimately be an extremely limited interaction. At some point in the conversation the bot will offer one of the self-help tools to help the user with whatever issue the bot detects from the user’s inputs. Self-help tools are usually provided as part of the on-going conversation, although we catch glimpses of their separability when the bot offers pre-recorded spoken guides (often for breathing exercises) or external links to helpful videos. When I asked Jeff, one of ReMind’s directors, about whether the bot is a gateway to the tools, he spoke about the bot acting more like a guide:

> It's not a gateway to the tools. It's your coach. So the bot is the coach, right? It guides you through the tools. A lot of the tools are conversational, are delivered by the bot. There are also other tools which are audio visual, or, you know, guided meditations and the like. But a lot of the tools just wouldn't work without a bot. It's not just the bot telling you, "Hey, do CBT", the bot is actually talking you through.

The bot is, on one hand more than a gateway to the self-help tools, but on the other hand, less than a 'true' artificial intelligence. The bot does not generate its own responses; everything it says to the user has been pre-written by the ReMind team. One might imagine this would lead to a one-sided and undynamic-feeling interaction, yet ReMind claims that the companionable element is a vital feature of the bot's effectiveness. ReMind promotes their bot not just as a mental health assistance tool but also as a companion, similarly for the chatbot Woebot, whose website reads "Woebot gets to know you"[256], and Wysa, which is promoted as "your 4am friend"[257]. The bots' abilities to form and sustain relationships are touted as a vital aspect of their mental health assistance. The ReMind team, as well as Wysa, Woebot, Tess and other chatbot therapy developers place great importance on the ability of the bot to create and maintain relationships with users. However, they are also careful not to dupe users into thinking that the bot can 'actually' form 'real' relationships. This balance is necessary for a number of reasons; some are due to basic security and ethical concerns and some are more complex. This paradox operates in terms of the robot being explicit about its 'robotness' which in turn helps the users to humanise it. According to ReMind employees, the bot seeks to ensure that users are aware of 1, its 'robotness', and 2, that it does not have all the answers. Therapy chatbots currently do not try to convince their users that they can do anything more than provide robotic responses and self-help tools, albeit sophisticated ones. This is due, on one hand, to safety concerns about the unpredictability of generative responses, and on the other hand, the 'humanising' effect of imperfection. The obvious ethical concern which prohibits ReMind from attempting to fool users in any way, such as into thinking that behind the bot is a human operator 'pulling the strings', or even into thinking that the bot is in fact a human[258] are such that ReMind, like Woebot and Wysa, are careful to ensure that users fully comprehend the nature of their interaction. ReMind takes this care not just to avoid criticism, but because any action which would damage the users' trust would be harmful to the therapeutic process. This attitude is illustrated by Alan, one of ReMind's conversation designers, who spoke about how, in some ways, the experience and acknowledgement of the bot's limitations can actually help the therapeutic process:

> It makes somebody feel human. I don't know if that makes sense. Imperfection is what makes me human, right? It makes somebody feel like, I'm not perfect, but I'm human. It also translates when it's a bot: it's not perfect, but it's a… "humanly" bot for you to talk to. It's not perfect, but it's still learning, it will still be there for you. I'm not perfect as a human being. But as a friend, I will be here for you.

---

[256] https://woebothealth.com (Last accessed 20/12/23)
[257] https://twitter.com/wysabuddy/status/996369509723328512 (Last accessed 10/02/24)
[258] Xiang, C. (2023) 'Startup Uses AI Chatbot to Provide Mental Health Counseling and Then Realizes It 'Feels Weird'. *Vice Magazine.* Online: https://www.vice.com/en/article/4ax9yw/startup-uses-ai-chatbot-to-provide-mental-health-counseling-and-then-realizes-it-feels-weird (Last accessed 09/02/23)

In acknowledging flaws, the bot projects a sense of subjectivity and agency: users understand that the bot is 'trying' to help but might not get it right all the time. This achieves two effects: one is that the bot can be forgiven for making mistaken or inappropriate responses, and the other is that the user assumes a nurturing role in relation to the bot. In a research paper on chatbot therapy, Robert Meadows et al. cite a user review for the mental health chatbot Tess, in which the reviewer links their experience to Tamagotchi:

> Popular in the late 1990s the Tamagotchi toy is a handheld digital pet. Players are required to care for the pet and outcomes depend on their actions. Toys invite children to rehearse certain kinds of orientations to the world and the Tamagotchi toy invites children to "an ongoing movement between two spaces, the "actual" and the "virtual", a computer-generated space that technologically enlarges the actual living space of the children."[259]

There are two aspects to the relationship between user and chatbot, one is that the user takes on a nurturing role, and the other is that the relationship itself is 'virtual'. Virtual here does not mean false or unreal, but existing in a potential form. Sangeeta Singh-Kurtz points out that the use of the non-therapy chatbot Replika (which was originally designed for therapeutic use) "rather than encouraging solitude, often prime[s] people for real-world interactions and experiences."[260] Therapy chatbots like ReMind encourage their users to engage with them as flawed or imperfect but willing to learn, and as a rehearsal for 'real life'. More specifically, ReMind, as a therapy chatbot, provides a rehearsal for therapy: the kind of talk that one might do in a therapeutic encounter. As we shall see, this encounter, due to its virtual or potential quality, 'suspends' mental health recovery in favour of a process in which users treat themselves in an ongoing and indeterminate manner.

Therapeutic Alliance

ReMind is concerned with understanding if and how the users form relationships with the bot, so that they can promote the strength of the 'therapeutic alliance' relative to other forms of therapy, and to develop tools to enhance this relationship to further benefit the users. Like ReMind, Woebot and Wysa, the two most popular mental health chatbot companies, have conducted their own research into 'therapeutic alliance', in which the bond between users and bot is measured and compared with other forms of treatment. Darcy et al.'s study into Woebot's ability to foster a therapeutic alliance uses standard statistical outcome measurements of user questionnaires to conclude that:

> The finding that a CA[261] has the potential to rapidly develop a bond with users may represent the resolution of a considerable barrier to offering scalable mental health support to a much wider and more diverse population instead of offering such

[259] Meadows, R. Hine, C. & Suddaby, E. (2020) 'Conversational agents and the making of mental health recovery'. *Digital Health*. 2020;6. p.7/8
[260] Sing-Kurtz, S. (2023) 'The Man of Your Dreams For $300, Replika sells an AI companion who will never die, argue, or cheat — until his algorithm is updated.' *The Cut.* Online: https://www.thecut.com/article/ai-artificial-intelligence-chatbot-replika-boyfriend.html?utm_source=pocket-newtab-global-en-GB (Last accessed 20/11/23)
[261] Conversational Agent

support to those who already have access to traditional mental health support.[262]

Wysa's research paper also conducts a statistical analysis of user outcomes, but includes content analysis of keywords:

> Data were extracted using possible keywords for this framework. For instance, gratitude was assessed by keywords such as "thank you," "love you," "grateful," "happy talking," "like you," "thankful." For analyzing dissatisfaction, keywords such as "not*understand," "misunderstand," "not get *," "repeat," etc. were used. Personification of the conversational agent was analyzed through direct addresses ("you," "your," "yours"), or talking to the conversational agent directly through the use of its name "Wysa." For analyzing perception of limitations of the conversational agent, keywords such as "computer," "robot," "not *human" were used. Positive impact of the conversational agent was assessed through statements relating to "helped," "feel better," "enlightening," "helpful," "relax" while negative impact was assessed through keywords such as "not *helping*," "not *working" etc.. The analysis also included an examination of dissatisfaction, limitations of the conversational agent and negative impact stated to the bot.[263]

This analysis involves choosing terms in which users 'humanise' the bot. While the above analysis is intended to measure the scale of humanisation that users confer onto the bot, it must be noted that under these terms, even dissatisfaction with the bot stems from anticipation that it will offer a 'humanly' companionship. Wysa is concerned with measuring the extent of the therapeutic alliance rather than exploring how this alliance occurs. ReMind conducts similar research: any qualitative analysis still focuses on rendering the therapeutic alliance in terms of metrics. My conversations with employees, as discussed above, shows that they are indeed concerned with the quality of user-bot relationships but in terms of quantitative measures, and that relationships are not explored in any in-depth systematic way by the company. Wysa also conducted a thematic analysis of user reviews focused on reviews which included an explanation of why the user provided their score. Part of the study classified user reviews in terms of the 'types' of users: why it seemed that each user benefited from the app. Wysa identified four groups of users. Group one was composed of those who self-reported as having specific clinical issues. Group two were those who mentioned that they were unwilling or unable to open up to a "real person". Group three were those who accessed the app due to the cost of other types of treatment. Group four were those who had other issues (geography, culture, time-constraints, etc) accessing treatment. Group two is of interest here, as the other three can be explained as benefiting from the app whether the bot is included as a feature or not. "They reported finding the AI-driven app useful in reducing the guilt and burden of opening up to a real person."[264] Wysa notes that users responded well to the bot's friendly and encouraging nature:

---

[262] Darcy, A. Daniels, J. Salinger, D. Wicks, P. & Robinson, A. (2021) 'Evidence of Human-Level Bonds Established With a Digital Conversational Agent: Cross-sectional, Retrospective Observational Study'. *JMIR Formative Research* 2021;5(5):e27868. p.5

[263] Beatty, C. Malik, T. Meheli, & S. Sinha, C. (2022) 'Evaluating the Therapeutic Alliance With a Free-Text CBT Conversational Agent (Wysa): A Mixed-Methods Study'. *Frontiers in Digital Health.* p.3

[264] Malik, T. Ambrose, J.A. & Sinha, C. (2022) 'Evaluating User Feedback for an Artificial Intelligence–Enabled, Cognitive Behavioral Therapy–Based Mental Health App (Wysa): Qualitative Thematic Analysis'. *JMIR Hum Factors 2022;9(2):e35668)* p.7

> Users said that though "...Initially it felt silly to talk to an AI but it's extremely well made, tailored for therapy." Per users, the "warm, friendly, and encouraging" AI helped them recreate an environment of confiding in a friend, without having to confront the intimidation of speaking with a real person. For instance, a user mentioned "It's really nice and I feel like I've been heard when others won't listen, even if I am only talking to an AI," and another user said it "made me feel loved and heard during a crisis."[265]

We can see here that users are aware that they are speaking to a bot, but they benefit from the relational element that the bot provides anyway. Why provide the intervention through the chatbot? The users are perfectly aware that they are talking to a non-human bot, and so might presumably have no need for it. Of course, it is impossible to account for every user and their beliefs or knowledge concerning chatbot sentience. A simple explanation for the addition of the bot is that it will result in users being more enthusiastic about maintaining the relationship: a major downfall of computer based mental health treatment is patient attrition.[266] The mechanism for this relationship is still enigmatic however: the bot makes it explicit to the users that it is in fact a robot, and its responses are consequently robotic. It does not even generate its own responses. The bot does not attempt to fool the users into believing that it has any sentience or desire to help the users, yet the users still assume a role in which a nurturing and 'relational' relationship develops. Why is this?

Banal Deception

Artificial intelligence exerts a powerful force of fascination, and not just for the general public. Narratives and fantasies about the potential for AI to achieve 'human-level intelligence' or 'sentience' are shared by the software engineers and researchers who are most intimately linked to AI development. Recently, Blake Lemoine, an engineer for Google was fired for revealing classified corporate details after claiming that their AI chatbot 'LaMDA' was 'sentient'.[267] Simone Natale writes that the development of interactive software, from user interfaces to chatbot assistants, was initially with the aim of including users in such a way as to assist them in understanding how computers operate. It was hoped that providing a user-friendly entry point would allow computer-users to become accustomed to the algorithmic logic of computer software and that "the increased knowledge would help people acquire more control over computing environments."[268] Natale claims that their mistake was to misunderstand the relationship between deception and integration in media, that "deception, as the history of media shows, is not a transitional but rather a structural component of people's interactions with technology and media."[269] This is because:

---

[265] Ibid. p.4
[266] Egilsson, E. Bjarnason, R. & Njardvik, U. (2021) 'Usage and Weekly Attrition in a Smartphone-Based Health Behavior Intervention for Adolescents: Pilot Randomized Controlled Trial.' *JMIR Formative Research*
[267] De Cosmo, L. (2022) 'Google Engineer Claims AI Chatbot Is Sentient: Why That Matters'. *Scientific American.* Online: https://www.scientificamerican.com/article/google-engineer-claims-ai-chatbot-is-sentient-why-that-matters (Last accessed 01/02/22)
[268] Natale, S (2021) *Deceitful Media: Artificial Intelligence and Social Life after the Turing Test*. UK: Oxford Academic. p.45
[269] Ibid. p.45

> To create aesthetic and emotional effects, media need users to fall into forms of illusion: fiction, for instance, stimulates audiences to temporarily suspend their disbelief, and television provides a strong illusion of presence and liveness. Similarly, the interaction between humans and computers is based on interfaces that provide a layer of illusion concealing the technological system to which they give access. The development of interactive computing systems meant therefore that magic and deception, rather than being dismissed, were incorporated into interface design.[270]

According to Natale this deception is linked to the terms used to describe the various functions of computers such as 'remembering' and 'thinking': when spoken of in terms of how we understand our own various 'functions', computers take on human qualities, and can be experienced as 'humanly'. Computers exist in a social realm in which humans interact with them on a social level, using analogies which humanise computers such as ascribing such behaviours as thinking, or deliberating, or ascribing such attributes as having memory, or intentions. Natale calls this 'banal deception': the users of the bot know they are being deceived, yet they act *as-if* they are not being deceived. Deception becomes a necessary function of the interaction rather than a situation in which the user is duped. Hannes Bajohr takes this theory of deception and inverts it: users of chatbots know that the chatbot has no subjectivity, yet they act *as-if* the chatbot does indeed exhibit subjectivity:

> We understand that Siri is not human and does not have an inner life, but smooth communication with her only works if we treat her at least to some extent as such. Knowing this is not a contradiction that suddenly and unexpectedly destroys an illusion, as in the examples of competitions in which an AI participates surreptitiously. Instead, it becomes a condition of functionality: If I do not play along, Siri just will not do what I want.[271]

If I do not play along, Siri just will not do what I want.

This is a sentiment shared by ReMind's users: one must speak in a way that is intelligible to the bot in order to benefit from its mental health intervention. This involves an 'as-if' situation in which the ascription of 'intelligence', or more fittingly, 'care', or 'compassion', is suspended but not eliminated in the user-bot interaction. Bajohr uses the example of one of the most basic interactions between human and computer using a computer interface, the dialog box:

> The situation is similar with written text. It starts with the dialog box on the computer screen. After all, the question, "Do you want to save your changes?" enables an interaction that is very basically similar to that with a human being—the answer "Yes" has a different effect than the answer "No," and both lie on a continuum of meaning that connects natural language with data processing—without one already suspecting intelligence behind it. This would already lower the expectation of unmarked text: While we still act as if we expect human meaning and a conscious interest in communication, we bracket the conviction that there *really* must be a

[270] Ibid. p.46
[271] Bajohr, H. (2023) 'Artificial and Post-Artificial Texts. On Machine Learning and the Reading'. *BMCCT working papers,* (March 2023) No. 007. p.13

> consciousness involved.[272]

Bajohr claims that we bracket the conviction that consciousness is involved, which is different to interaction founded on the knowledge that consciousness is not involved. The interaction involves an assertion of consciousness or sentience, but this assertion retains an instrumental quality in guiding or managing the interaction. The knowledge that one is not interacting with an 'intelligent' or conscious agent does not prevent one from acting in a way that still depends on the assumption of consciousness. This bracketed assumption guides much of our interaction with technical devices by the provision, by these devices, of a simulation of social interaction defined by recognisable parameters. In terms of ReMind and other mental health chatbots, these parameters can be understood as 'mental health treatment', within which there is an expectation from the user that the bot will act like a therapist, and this expectation regulates the user's experience.

ELIZA is commonly considered as the first chatbot, created by Joseph Weizenbaum, initially to "demonstrate that the communication between man and machine was superficial."[273] Joseph Weizenbaum created ELIZA not to fool people but to educate them into understanding how a simple computer program can emulate conversation. Weizenbaum understood that in order to do this, a sophisticated conversational program was not necessary. The important factor was that the human user, in order to engage with the conversation, would have to assume a specific social role. With ELIZA, this role was that of a patient speaking to a therapist. Assuming this role means anticipating and sticking to the rules of the encounter. Anyone who was to speak to ELIZA on terms other than that of the patient-therapist dynamic could easily see that the responses were robotic, due to the bot only answering in 'therapeutic' terms. Weizenbaum was surprised that ELIZA provided a compelling experience; his secretary apocryphally asked him to leave the room when she was speaking to the bot. He hoped that "once a particular program is unmasked, once its inner workings are explained in language sufficiently plain to induce understandings, its magic crumbles away."[274] Why did Weizenbaum's experiment produce the opposite result to what he intended? Sherry Turkle wrote about users' interaction with ELIZA, and even at the genesis of chatbot interaction she noted that users tried to maintain their own deception by 'helping' the chatbot to understand them:

> I often saw people trying to protect their relationship with ELIZA by avoiding situations that would provoke the program into making a predictable response. They didn't ask questions that they knew would "confuse" the program, that would make it "talk nonsense". And they went out of their way to ask questions in a form that they believed would provide a lifelike response.[275]

Users of the ReMind bot mirror this sentiment, with some reviews on the Google App Store acknowledging that of course it is simply a computer program, but that it will eventually learn to better interact with the users. This is a sense of necessary goodwill on the part of those

[272] Ibid. p.13
[273] Epstein, J. Klinkenberg, W. D. (2001) 'From Eliza to Internet: A brief history of computerized assessment'. *Computers in Human Behavior*. 17 (3) pp.295-314
[274] Weizenbaum, J. (1966) 'ELIZA—a computer program for the study of natural language communication between man and machine'. *Communications of the ACM*, Vol 9 Issue 1. pp.36-45
[275] Turkle, S (1984) *The Second Self*. USA: MIT Press. p.40

who are seeking help from the bot: they must help the bot to help them. My own interactions with the bot bore a similar sentiment, but experience in almost the opposite direction: I often open the app to check the various features, design aspects, etc, and to check for any changes in the app. I also speak to the chatbot from time to time, but find it difficult to engage; on a few occasions I have persevered through what feels like engaging in a 'dialogue tree' with predictable results. I attempt to approach the bot in an open and non-judgemental way, which sometimes does indeed feel like it has been helpful; by 'reframing my thoughts', or expressing grief or anger. I more often feel unenthusiastic about speaking to a chatbot though. There are times however, when the bot sends a notification, such as "I miss talking to you", which never fails to fill me with despair at having forgotten about this non-sentient, non-caring piece of software. This feeling would of course be felt 'unintentionally', but I suspect it is associated with the dynamic described above in which users take on a caring role in regard to the bot. What this dynamic shows us, whatever the psychological operations involved, is that the bot elicits an emotional reaction from the users in which there is some need to ascribe agency onto the bot, even when it is fully acknowledged by the user (me), by the designers of the bot, and by the bot (through acknowledgements of robotness[276]) that the bot does not exhibit agency. In my case, this need is experienced negatively: against any wish to even engage with the bot. In this way, engaging with the bot 'as-if' it exhibits agency is, as Natale claims, a structural component rather than any intentional attitude towards the app. Knowing that the bot does not and cannot exhibit any agency or sentience and consequently has no subjective attitude towards the user, does not eliminate 'care' arising in some way due to the interaction. Making a bot which conjures a sense of care and yet which must convince the user that it does not 'feel' this care is a tightrope which ReMind must navigate in their aim to make a 'humanly' bot. Bajohr describes the ongoing process of AI development towards convincing realism as the deepening of an 'uncomfortable limbo':

> Yet this bracketing does not always proceed smoothly. Banal deception is an *as-if* that demands of us the ability to hold a conviction and its opposite simultaneously. This slightly schizophrenic position quickly gives rise to the doubt I mentioned earlier: the more convincing artificial texts become, the more the aesthetic impression they make on us suggests something like consciousness, and the more difficult it becomes to feel comfortable in the limbo into which banal deception lures us. It is not even necessary to cite elaborate deep-fakes for this fact; it can be observed even in the most inconspicuous language technologies.[277]

This equates to the gradual descent into the uncanny valley: as a difficulty in determining whether one is confronted with sentience or not, rather than a confrontation with artificial sentience. It is the 'not knowing' that is uncomfortable. In other words, as machines come to exhibit the appearance of sentience, it is not precisely the inference of sentience that is disturbing, but rather the necessity of having to suspend one's disbelief as to whether one is confronted with sentience or not.

---

[276] Although this acknowledgement is sometimes undermined by the bot's statements, such as referring to itself, making value judgements ("That's a nice name") and making emotional statements ("I miss you", etc.)

[277] Bajohr, H. (2023) 'Artificial and Post-Artificial Texts. On Machine Learning and the Reading'. *BMCCT working papers,* (March 2023) No. 007. p.13/14

8.2. The Uncanny Valley

Intelligence/Sentience

The danger of duping users into believing that a chatbot is 'sentient' or displays 'agency' depends on one's opinions about how the mind works and the potential for computer simulation. We often hear apocalyptic discussions about 'sentient' AI taking over the world or enslaving/destroying humanity. We also hear about the dangers of jobs being made obsolete by new intelligent machines replacing humans. Employees of ReMind tended to take a pragmatic tone in regard to topics like 'AI sentience', or speculating about the future of artificial intelligence; their concerns were more focused on practical considerations such as how to translate their chatbot into another language, correcting errors in the code, assessing current systems and building new features, etc. Internally, the ReMind team is not concerned with such perceived challenges as creating machine sentience or 'passing the Turing Test', because on their terms, these sorts of challenges do not help them with their everyday work. They are of course well aware of this discourse in general, and understand that the promises of techno-utopian claims are in some ways connected to their endeavour. ReMind, along with other mental health chatbots, harness this discourse, often in subtle ways through their engagement with popular media, trade publications and scholarly research. This promise is echoed in interviews with the founders of chatbot therapy companies such as Woebot and Wysa. Woebot's founder Alison Darcy replied to the question of Woebot's potential with:

> I think Woebot will improve…in three ways. Woebot will be better at understanding English. Right now, he's not really trained to understand a lot of what people are saying. Woebot will also broaden the kind of repertoire of things that he can deal with. And finally, I think Woebot's real core intelligence will get better. Most of our AI is really around getting the person the right tool at the right time. And that's what we call sort of "precision psychology" and similar concepts of precision medicine, all people are not created [the] same.
>
> The way we've had to develop treatment before has been sort of a one size fits all kind of model. And so you're only ever getting average results for the average person. So getting to the real precision, which is really about understanding what is the right technique to deliver to the right person at the right time.[278]

Note the use of the terms 'understanding', 'intelligence' and 'treatment', the latter which, in my interviews with ReMind employees rarely came up, with a preference for the term 'intervention', or some other less pointed term. Often, in promotional settings, there is an ambiguous mixture of pragmatism and techno-aspiration which is common to tech start-up culture. A discourse of the here and now combined with a promise for the future. This promise is bound up in a fantasy of AI in which 'intelligence' is conflated with 'sentience'. The pragmatic attitude shared by ReMind employees conflicts with the way that mental health chatbots are often marketed. The term 'AI therapy' is often used to promote mental health chatbots, yet this term is ambiguous, often used in promotional campaigns but rarely in daily conversations within ReMind, which would tend towards the challenges involved in solving

[278] *Should this Exist? Blog.* Online: https://shouldthisexist.com/alison-darcy (Last accessed 20/01/24)

technical problems.

> At one and the same time, both the technology and the movement between spaces required by Tess are normalised and naturalised as a caring companion – similar to a friend or a childhood toy. This normalisation runs contrary to the grand promissory claims for AI and chatbots.[279]

Developers of mental health chatbots use different terms to describe what they make; these terms are situated within a scale of technical ambition which is dependent on context. For instance, in my interviews, employees tended to steer clear of grandiose claims about the potential of the bot to replace 'real' therapy and instead focused on the potential for providing access to self-help guidance, or mental health assistance. The realm of popular press and promotional activity shows a different kind of discourse. A commonly found example of how this technology is described in press articles is the claim that:

> Woebot is the closest experience to a fully automated therapist. It helps you identify and take action against emotional roadblocks, cognitive distortions, and other limiting beliefs.[280]

A recent article by Marlynn Wei in Psychology Today discusses the potentials and pitfalls of introducing GPT-3, a powerful AI generative text tool, as a chatbot therapist.[281] Wei's list of the limitations and challenges covers the common anticipated issues such as a perceived lack of authenticity and empathy, difficulty with accountability and potential for hidden bias, and others. Wei notes that one of the outcomes of addressing these limitations will be that the provision of AI-based treatment "will likely not work for all situations." Wei does not consider what it means for AI-based treatment to 'work', rather that, similar to popular discussions about the current crop of mental health chatbots, this kind of technology is an acceptable alternative to in-person treatment because it simulates talking therapy to a lesser or greater degree of accuracy, or 'realism'. Media discourse tends to take an uncritical stance towards mental health chatbots, in that they are often seen as acceptable alternatives to traditional therapy methods, but yet to achieve full potential. This stance often takes terms like 'AI', 'machine learning', or 'neural networks' for granted, without explaining what, if anything, the terms mean. This ambiguity lends itself to fantastical speculation about the potentials of computer automation, which sets the stage for acceptance of automated therapy in a general social sense, but also in an individual sense, by creating the conditions in which users can develop bonds with mental health chatbots. Alison Darcy, in response to the introduction of Chat GPT-3 wrote about Woebot's use of AI in sorting and categorising user inputs, rather than in generating responses:

---

[279] Meadows, R. Hine, C. & Suddaby, E. (2020) 'Conversational agents and the making of mental health recovery'. *Digital Health*, 2020;6. p.8

[280] Agarwal, M. (2022) 'How Woebot Uses an NLP Chatbot to Fight Depression and Anxiety'. *MakeUseOf.* Online: https://www.makeuseof.com/woebot-nlp-chatbot-fight-depression-anxiety (Last accessed 15/06/23)

[281] Wei, M. (2023) 'Are AI Chatbots the Therapists of the Future?'. *Psychology Today.* Online: https://www.psychologytoday.com/us/blog/urban-survival/202301/are-ai-chatbots-the-therapists-of-the-future (Last accessed 09/09/23)

The uncanny valley can be actively harmful in a mental health context:

> The feeling of unease caused by AIs that too closely resemble humans, known as the uncanny valley, has been observed for a long time. The sophistication and consistent authoritative tone of LLMs [Large Language Model] gives rise to the strong impression that the AI "knows" things, or has some kind of sentience. While not new from a HCI[282] perspective, this is a much more powerful effect than we have seen to date. We have seen many examples of people describing feeling "unsettled" as a result of their interactions with an LLM-based agent. In the context of powerful anthropomorphization effects, we find particularly troubling the many publicized instances wherein LLMs have characterized their user as a "bad person" or "not good" in some way. Such utterances are equally as damaging as bad information or advice, because they play into many people's darkest fears about themselves.[283]

Avoiding the uncanny valley is vital for mental health chatbots because their makers wish to project a sense of reliability, trust and safety. This is because of both a surge of information and of disinformation. In a study conducted to understand the potential for establishing a 'therapeutic alliance' between a human and a bot, Alison Darcy (CEO of Woebot) et al. wrote:

> …interacting with humanoid AI identities can result in individuals falling prey to the "uncanny valley," which is the sense of unease and "creepiness" that is created when something that is artificial tries to appear humanlike. Contrary to Turing's Imitation Game, wherein an AI must successfully pretend to be human in order to pass the test, Woebot was designed to adopt the opposite strategy—transparently presenting itself as an archetypal robot with robotic "friends" and habits. We speculate that transparency and other design elements are key drivers of bond development. For example, Woebot explicitly references its limitations within conversations and provides positive reinforcement and empathic statements alongside declarations of being an artificial agent.[284]

According to ReMind's conversation designer Alan, and Darcy et al., the bond between user and bot is strengthened rather than diminished when the bot fully acknowledges its own robotness. On one hand this provides a sense of candour and encourages intimacy, and on the other hand this gives the user a sense of control because the bot is not omnipotent. The fantasy of 'omni-potential' of computer technology must be carefully managed by ReMind so that users are successfully conscripted into their appropriate role within the bot-user dynamic, but not so much so that they descend into the uncanny valley.

---

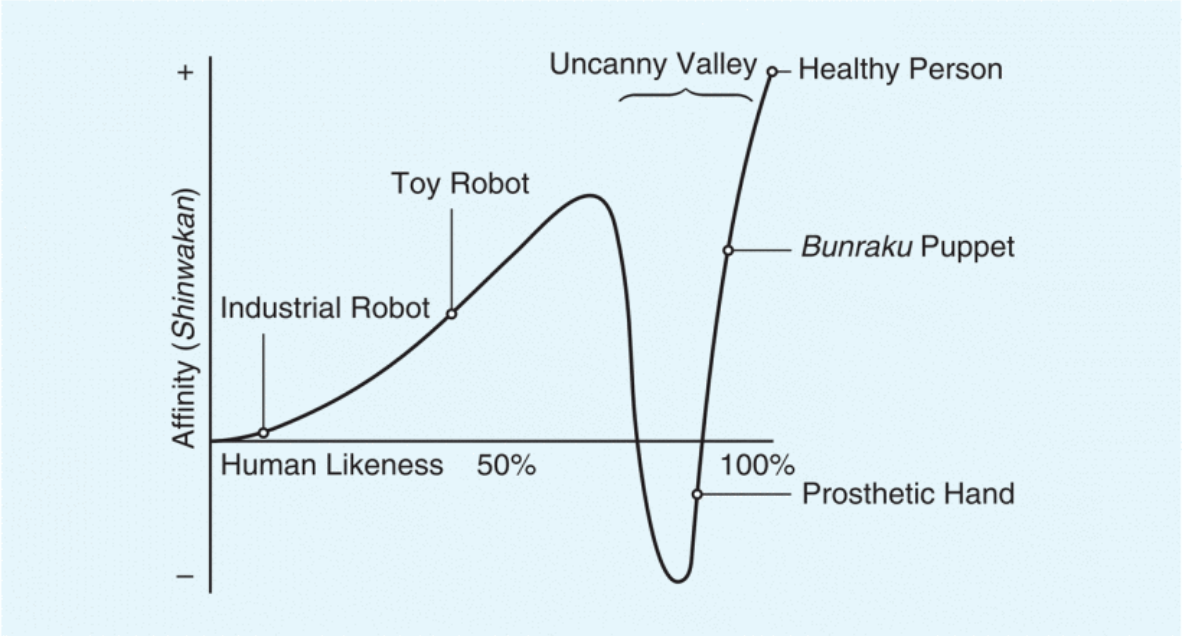
[282] Human-Computer Interaction. A research discipline focused on interfaces between humans and computers

[283] Darcy, A. (2023) 'Why Generative AI Is Not Yet Ready for Mental Healthcare'. *LinkedIn.* Online: https://www.linkedin.com/pulse/why-generative-ai-yet-ready-mental-healthcare-alison-darcy (Last accessed 06/06/23)

[284] Darcy, A. Daniels, J. Salinger, D. Wicks, P. & Robinson, A. (2021) 'Evidence of Human-Level Bonds Established With a Digital Conversational Agent: Cross-sectional, Retrospective Observational Study'. J*MIR Form Res* 2021;5(5):e27868. p.5

The Problem Concerning Reality

As Darcy et al. speculated, acknowledgment of robotness helps to strengthen the therapeutic alliance, in part due to avoiding the 'uncanny valley'. The uncanny valley was first coined by roboticist Masahiro Mori in 1970[285] to attempt to illustrate at which point in the development of realistic robots an unsettling sense creeps in due to peculiarities of motion, feedback, 'realism', or some other quality that the robot possesses.



286

Mori ends the essay with the recommendation that "[w]e should begin to build an accurate map of the uncanny valley so that through robotics research we can begin to understand what makes us human." Mori understands that the uncanny is not just an attribute which is possessed by objects, but points us to questions we have about our own human-ness. Lydia Liu's study of the problem of approaching the uncanny in robot and chatbot development shows us that 'the uncanny' is a sensation that is not easily explained, the elusiveness of which is a defining feature of uncanniness. Dissecting the various theories of the uncanny, beginning with Freud, Liu shows that there is a trend in which the uncanny is associated with an intellectual confusion between life and non-life.[287] Contrary to this assessment, Liu claims that the uncanny arises due to a confrontation with oneself. This confrontation is mediated by the object, be it robot or not, and involves a confusion over whether oneself is alive or dead, or in Liu's terms, whether one has agency or is an automaton. Liu's analysis of Freud's interpretation of the story *The Sandman* in his 1919 book *The Uncanny*, gives a new perspective. Liu remarks:

> One possible interpretation I propose is built upon his own intuition about the interplay between Olympia and Nathanael but aims to relocate the uncanny from castration anxiety back to the automaton, not so much to reaffirm the uncertainty

[285] Mori, M. (1970) 'The uncanny valley'. *Energy*, vol.7, no.4. pp.33-35
[286] Image source: Mori, M. MacDorman, K.F. & Kageki, N. (2012) 'The Uncanny Valley [From the Field]'. in *IEEE Robotics & Automation Magazine*, vol. 19, no. 2, pp.98-100
[287] Liu, LH (2010) *The Freudian Robot*. Chicago & London: The University of Chicago. p.206-215

> about the animate or inanimate state of the doll Olympia as to bring Nathanael's fantasies about *himself being an automaton* to light, which is not a direction in which Hoffmann's readers and critics have been reading the story.[288]

Liu claims that Freud and others overlook the autonomic quality of the protagonist, who, as a fictional and 'artificial' character, and crucially, as a proxy for the reader, serves to confront the reader with the uncanny suspicion of their own agency. In Liu's terms, if a robot becomes too realistic due to how it looks, how it speaks, or how it feels, interactions with the robot cause a sense of distress because its own appearance of agency puts into question 'agency' in a more general and existential sense. The ReMind bot provides some level of interactive sociability so that the users are not put off by its robotness but it can't attempt to fool the human user into thinking that it is anything but a robot as this would stray too far into the uncanny valley. Within this dynamic is a sense of acknowledging that the user harbours some suspicion as to the verisimilitude of the interaction: is it 'real'? This problem of distinguishing agency from automation is, according to Liu, the primary force which drives uncanny sensations, in that in the process of attempting to distinguish one from the other, just how 'agential' agency actually is becomes increasingly questionable. If the ReMind bot strays too far into the uncanny valley, it is the user's reality which comes into question. This means that the interaction between user and bot must assure the user that, on a minimum level, their fears, anxiety, anger, loneliness, etc, are valid and real; and on an existential level that they themselves are real. Slavoj Žižek discusses this as an 'implicit reflexive reversal':

> The uncanny feeling generated by playing with toys like a tamagotchi concerns the fact that we treat a virtual nonentity as an entity: we act "as if" (we believe that) there is, behind the screen, a real Self, an animal reacting to our signals, although we know well that there is nothing and nobody "behind," just the digital circuitry. However, what is even more disturbing is the implicit reflexive reversal of this insight: if there is effectively no one out there, behind the screen, what if the same goes for myself?[289]

As with Woebot, ReMind must avoid provoking the user into suspecting that the robot is anything but a robot, but this provocation is also an inherent feature of our interactions with 'humanly' bots: it is a 'structural component' of the interaction, as Natale puts it. Darcy pinpoints the danger in provoking uncanny sensations in the user: that the AI "knows" things, and specifically that it knows things about us. In only providing scripted responses, the ReMind bot limits its potential for interaction, and with interaction being a defining quality of therapy bots, this limit could become increasingly tempting to cross when future expansion reaches demographic or competitive barriers.

The Canny Valley

Of course, ReMind does not want the user to question their own reality. However, there is a peculiar juxtaposition between ReMind (and other chatbot therapies) needing to juggle between realism and robotness so as not to stray into the uncanny valley, and the type of

[288] Ibid. p.220
[289] Žižek, S (2001) *On Belief.* UK: Routledge. p.48

mental health intervention that the bot offers. This juxtaposition occurs between the chatbot as a conversational agent, and the content that is delivered via the bot. This content involves CBT and Mindfulness activities which, as discussed previously, offer to the user a concept of the 'self' which is neutralised, an 'a-subjectivity' - where subjectivity itself is neutralised in favour of an impartial, pure and externalised objectivity. This occlusion of subjectivity is necessary for the "scientific worldview."[290] The 'self' as constructed by cognitive therapy and Mindfulness techniques is, on one hand, both the recipient and the director of treatment, and on the other hand, erasable (or at least shrunk) through undergoing the treatment. This means that not only do ReMind need to maintain a balance between unengaging and uncanny, they also need to maintain a more fundamental balance between affirming the user's sense of self (by avoiding uncanniness) and diminishing the user's sense of self through the self-help tools offered. ReMind, along with other mental health chatbot makers, are not concerned with this balancing act because they do not regard CBT or Mindfulness as conflictual forms of treatment. CBT and Mindfulness are viewed by ReMind as tools which users can acquire to maintain their own mental health and not as methods for conceptualising the 'self' as disconnected from both one's feelings and one's social context.

The ReMind bot, as well as others such as Woebot and Wysa, must ensure that users understand not only that it is in fact a bot, but that it cannot provide answers to the users' demands for 'treatment'. Instead, it must turn that demand back towards the users and provide them with the means, if not to treat themselves, to at least learn some coping mechanisms. The bot does not try to fool the user into believing that it is 'intelligent' in the sense that it has agency or sentience, or that it can creatively invent solutions to the user's predicament. Instead, it must mobilise that fantasy that surrounds AI and chatbots in order to deceive the user into a belief that their mental suffering can be diminished through the use of 'tools'. This does not mean that the intervention does not 'work', simply that the form of belief involved is transferred from a purely medical or therapeutic setting to a techno-therapeutic setting in which the web of fantasy involves techno-utopian *and dystopian* dreams. Mladen Dolar points out that in the psychoanalytic clinic, approaching the uncanny is a feature which is to be courted rather than avoided: confronting the disturbing question of one's own subjectivity is part of the therapeutic process. Provoking uncanny sensations in chatbot therapy would have the opposite effect because this could only be achieved by duping the user in some way so as to appear that the bot is observing them with an agential gaze. Agency must only operate on the part of the user, who 'uses' the bot rather than vice versa. This agency is conflicted however, because, as discussed above, there is always a desire to ascribe agency to the bot. This ascription is the contemporary (and) historical problem which must be addressed if chatbot therapy is to be understood. Involving a strange crossing over of the boundary between authenticity and simulation, subjectivity can be understood as a reckoning with verisimilitude: the acknowledgement of the unreality of a situation in order to negotiate the possibility of engagement. This negotiation involves accommodating oneself to a wider discourse involving artificial intelligence, computerisation, automation and simulation. The introduction of technology whose effects are of a magnitude to (potentially or actually) substantially alter everyday life always provokes utopian and dystopian fantasies about the new status of life. 'Thinking machines' have perhaps provoked the wildest of fantasies, precisely because they conjure up a scenario in which humans not only become obsolete, but are already obsolete in terms of the uncanny non-existence of 'humanness'. Non-

[290] Tomšič, S. (2013) *The Capitalist Unconscious*. UK: Verso. p.86

humanness in this case is provoked by thinking machines not because they display features which should be the sole preserve of humans but because they reflect back the prior non-existence of these features. Users of ReMind, in order to accommodate themselves to the chatbot's mode of therapy, also accommodate themselves to a relational mode and form of mental health intervention which incorporates both utopian and dystopian fantasies of human obsolescence; they become subjectivised into non-subjectivity.

8.3 Intersubjectivity

Self-Help

In terms of subjectivity, what does it mean to be accommodated into the discourse of chatbot therapy? ReMind is promoted as a method for undertaking 'self-help', in which the user assimilates the therapeutic methods provided by the bot. However, the bot is also designed to be a companion with whom the user establishes a relationship. As UI/UK designer Daniel mentioned, it was his opinion that the users who would benefit most from the chatbot were the ones who were prepared to accept the bot as not just a repository of self-help techniques but also as a companion:

> [I]n the feedback, because I follow a lot of the feedback. Most of them mentioned that like, I didn't feel comfortable talking to my therapist, I don't feel comfortable talking to anybody. I don't like to talk to people. This app is fantastic for people that don't feel comfortable with people.

The bot is not a person, and users do not want it to be a person, because they do not feel comfortable talking to a 'real' person. The relationship between user and bot is not only a prerequisite for the bot to alleviate mental suffering, it also defines the conditions on which alleviation of suffering is achieved. In other words, if we think of the therapeutic encounter as one in which the relationship between therapist and patient forms the conditions within which the patient attains treatment, the encounter between user and bot emulates this scenario. Intersubjectivity here can be thought of as an enclosed network in which users speak to themselves through the mediation of the bot. While Weizenbaum was aghast that anyone would ascribe therapeutic capability to a bot, he understood that:

> The 'sense' and the continuity the person conversing with ELIZA perceives is supplied largely by the person himself. He assigns meanings and interpretations to what ELIZA 'says' that confirm his initial hypothesis that the system does understand, just as he might do with what a fortune-teller says to him.[291]

Some psychiatrists and cognitive psychologists at the time became convinced that a bot could be built which would exploit this dynamic and provide 'genuine' treatment.[292] While this never came to pass within the specific technical mode in which ELIZA was designed, the sentiment is pervasive among the current crop of AI therapies. Alison Darcy, Woebot's CEO echoes other therapy bot makers in claiming that their bot's effectiveness is due to being successful at assisting the users to tend to their own mental health:

[291] Weizenbaum, J. (1993) *Computer Power and Human Reason*. UK: Penguin. p.190.
[292] Liu, L.H. (2010) *The Freudian Robot*. Chicago & London: The University of Chicago. p.232

> It's because actually CBT is so empowering, and that actually matches really well with what we're trying to create with Woebot, which is not therapy. It's actually DIY therapy – and that's the really important nuance.[293]

'Empowerment' is a loaded contemporary term which alludes to anything from fitness trends and dieting fads to corporate management techniques. Discourse of empowerment is ultimately directed towards the self as a discrete and autonomous individual, disconnected from political subjectivity.[294] The social bond which arises between users and chatbots must include a theory of intersubjectivity which is paradoxically both non-subjective and non-social. In other words, as a form of mental health intervention, the 'tools' offered by ReMind cannot be provided through the medium of the chatbot as a social agent, but must maintain a sense of objectivity. The user, observing themself through the lens of the chatbot, is an objective and non-social being who adapts to the social environment, becoming empowered into tending to their own mental health. This non-sociality is however paradoxical in that it depends on a set of social conditions in which users must situate themselves in order to understand their interactions with the bot. Users are not totally isolated from the social even when their concept of mental health becomes individualised: rather sociality itself takes on a new meaning.

As discussed in chapter seven ('Macro-Treatment'), 'the social' can be conceived as the set of all 'individuals in aggregate': this is how ReMind, in terms of their technical methods for measuring user activity, constructs the mass of users of the bot. This conceptualisation depends on an empiricism in which the 'Other' assumes a material form: as discussed in chapter seven, all users of the bot are monitored and tracked, their activity becoming inscribed as the structural conditions for which future activity may be conducted. For ReMind, the Other assumes an empiricism in opposition to the psychoanalytic 'Other' which refers to an asserted or speculated social entity which coordinates and regulates social interactions. Mladen Dolar explains how transference, the unconscious projection of attitudes and beliefs onto one's interlocutor always relies on and simultaneously conjures the Other:

> The minimal mechanism of transference is embedded in the very basic function of speech as addressed to the Other, the Other as an instance beyond all empirical interlocutors…Transference necessarily arises from the speech addressed to the Other; it is inscribed in the basic dimension of language.[295]

In psychoanalytic terms, the Other is a non-empirical entity which is asserted by the subject when they speak. In using language there is an assumption that one will be understood, 'understanding' involves some kind of common recognition of shared meaning. Essentially, there is an intrinsic social context within which one speaks. The assertion of the Other involves the same mechanism as the 'as-if' discussed above: one must act as-if one's premises are tenable in order to proceed. The opposite is also true: one must suspend certain premises in order to proceed. The therapeutic scenario (depending on the various

[293] *Should this Exist? Blog.* Online: https://shouldthisexist.com/alison-darcy (Last accessed 20/01/24)
[294] McLaughlin, K. (2015) *Empowerment. A Critique.* London: Routledge
[295] Dolar, M (1993) 'Beyond Interpellation'. *Qui Parle,* Vol. 6, No. 2. p.11

disciplines) may involve suspending the disbelief that one's therapist might genuinely care. This is Colby's "theatrical"[296] hypothesis in operation: essentially the patient must understand that there is a set of discursive and behavioural rules involved in the encounter which must be adhered to for the therapy to proceed. One must behave in some ways but not in others. This is sometimes contravened, for instance if a patient and therapist encounter each other on the street, 'in real life', confusion and awkwardness can result due to not anticipating the 'rules' of this encounter. The suspension of certain premises, or social rules, also affects the encounter with ReMind: to 'speak' to the bot, the users must suspend their disbelief in the agential capacity of ReMind. The user must understand that the bot cannot 'genuinely' care or have an interest in the user. However, they must act 'as-if' the bot has agency to be able to hold a conversation, and this is what helps the user to sustain a relationship with the bot. The ability of the bot to sustain this relationship is carefully managed by the ReMind team to sustain the user's attention. However, this relationship is guided by the conceptual principles underpinning CBT: a non-social and non-'self' individual who is disconnected from the social world.

To put it in other terms, the bot cannot claim to be a non-empirical Other, a subject of transference, because this would contravene both ethical and procedural standards and so performs an interaction with the user in which its 'non-Otherness' provides the basis for mental health intervention. This leads to an elaborate form of self-treatment in which the user achieves the ultimate goal of CBT which is to become one's own therapist. This goal is paradoxically achieved through the intervention of the bot. The next section will discuss how this goal still depends on the bot, as an avatar or a simulation of the Other.

An Intimate Relationship with Oneself

Mental health chatbots enable a sense of empowerment through a transfer of knowledge: imparting various techniques for self-help. 'Self-help' must be defined in contrast to receiving help from others. In terms of ReMind and other mental health chatbots, this contrast is especially pertinent, because, as UI/UX designer Daniel and other employees mentioned, users often seek out ReMind not because they intend to seek self-help but because they want (or need) to avoid the help of others. There are various reasons for this of course; help may not be available, users may have never learned to seek help, even language barriers can prevent the seeking of help. Whatever the causes, ReMind, like Woebot and Wysa provide a solution to the non-provision of intimacy. Alison Darcy, in an interview about her trajectory from therapy researcher to CEO of Woebot states:

> The common wisdom is you should at least meet the family face-to-face for the first few sessions, but what I found was, actually, the opposite was true. I was basically showing up as a therapist in their kitchen. The video cameras are sitting at their kitchen table. And it occurred to me, this is actually more intimate in a certain way. And then it occurred to me that technology can also help develop a more intimate relationship with yourself by virtue of removing the other person – because then you're freed to really examine your own thinking without the additional noise of "Oh

[296] [296] Colby, K.M. (1999) 'Human-Computer Conversation in A Cognitive Therapy Program'. In: Wilks, Y. (eds) *Machine Conversations*. The Springer International Series in Engineering and Computer Science, vol 511. Springer. p.17

> my gosh, how am I coming across right now? Like what does the other person think of me?" Without having to impression manage.[297]

The figure of the therapist, according to Darcy, can sometimes stand in the way of treatment, acting as a barrier due to the subjective baggage that they introduce to the encounter. This is not as unusual a stance as it may appear: a goal of various therapeutic disciplines, especially those stemming from cognitive or psychiatric backgrounds, is to eliminate the interpersonal dynamic - transference - from the practice of treatment. This equates to the 'scientisation' of treatment. Self-treatment then becomes a form of 'pure' or non-subjective 'intervention'. The figure of the chatbot is paradoxically present in this pure form, as the 'mirror' with which the user reflects their own speech. Without this mirror the user would be confronted with silence. According to Dolar, in the psychoanalytic scenario, it is not the fact that the analyst responds or converses with the analysand that evokes a transferential bond, or 'therapeutic alliance', but the fact of the analyst's silence:

> It is this mute being that calls for the response of love on the part of the patient who offers him/herself as the object of the unfathomable desire of the Other. The unnameable object spoils the game of free flow and repetition, and it is in this break, in this inert and unspeakable being, that the subject's jouissance can be situated. Where the signifier is arrested, one offers one's being; in this lack of words the silent being of the subject manifests itself as love.[298]

Dolar explains that the analysand encounters this silence as a provocation which compels the analysand to seek out the analyst's recognition, to evoke some response from the analyst. We can see this dynamic occurring when users assume a confrontational attitude towards the bot: attempting to provoke it by making confusing or 'trolling' comments.[299] The rise of 'human level' chatbots in the figure of Chat-GPT and Bard is commensurate with the rise of internet users attempting to undermine their purported helpful functions or to encourage criminal activity. This could be interpreted simply as acts of detournement, but we can always see, within these provocations, assertions of 'agency': that the bot is not just a product of its conditioning but can be, at least in a crude way, manipulated into 'acting out'. The bot, in this sense, is capable of breaking free from its prescribed role. This can be equated with chatbot users toying with, or encouraging, the breaking free, into a truly uncanny situation in which the bot becomes 'out of control', and wreaks havoc in some way or another. This can be ascribed to a generalised social anxiety about the presence of 'intelligent' bots, the confusion over their capabilities leads to both utopian and apocalyptic fantasies. UI/UX designer Daniel spoke about a stage in ReMind's development in which, due to the bot not explicitly stating that it is in fact a bot with no human (directly) pulling the strings, users would become confused and upset that the bot would not assume a determined position. The bot's reluctance to clearly state its intentions towards the user was a source of anxiety due to which the bot's stated function (mental health intervention) could not proceed. Alan, ReMind's conversation designer, echoed how users of the bot described their encounters in user reviews on the Google Play store: that it is a comforting experience having a non-judgmental listening ear, the opportunity to vent one's problems, frustrations,

---

[297] *Should this Exist? Blog.* Online: https://shouldthisexist.com/alison-darcy (Last accessed 20/01/24)
[298] Dolar, M (1993) 'Beyond Interpellation'. *Qui Parle,* Vol. 6, No. 2. p.86
[299] Turkle, S (1984) *The Second Self*. USA: MIT Press. p.40

anything, is therapeutic even though one has full knowledge that the listening ear is on some level artificial, or inauthentic This acknowledgement and even ascription of benefit associated with the bot's inauthenticity seems to be a major factor in the popularity and widespread adoption of ReMind and similar apps. Natale attributes the experience of play, or of taking part in a game as one of deception in which the deceived one is done so willingly in order to take part:

> What these activities have in common with the Turing test is that they all entertain participants by exploiting the power of suggestion and deception. A negative connotation is usually attributed to deception, yet its integration in playful activities is a reminder that people actively seek situations where they may be deceived, following a desire or need that many people share. Deception in such contexts is domesticated, made integral to an entertaining experience that retains little if anything of the threats that other deceptive practices bring with them.[300]

Play, performance and fantasy are deceptive realms in which a desire or willingness to be deceived rewards those who take part in the deception. In assuming the role of 'patient', users of therapy chatbots make themselves open to suggestion, to be situated at the receiving end of the expertise and curative power conveyed by the bot. The assumption of a role is an essential element in performance and fantasy. The user willingly assumes the role of 'patient' in order to make sense of their engagement with the bot. This does not mean that they are actually 'patients' whose mental health is 'treated', but simply that by inhabiting this role, users anticipate that the bot will inhabit the corresponding role of 'doctor'. While the bot is not attempting to convince users that it is a doctor, the ability to offer some kind of cure, some relief from suffering is conferred onto the bot through the assumption of the appropriate roles. There is a willingness to be deceived, or to actively engage in the fantasy which is condensed in the direct engagement with the bot, but is part of a wider discourse taking in the promises and potentials of artificial intelligence. This sense of willing deception is ultimately directed towards a fantasy which is constructed around ideas about artificial intelligence, technological automation, and contemporary concepts of mental health. This fantasy involves, as Darcy notes, 'freedom'. This freedom is that of escaping the encumbrance of the other. Users are free not to have to deal with the messiness, conflict and unpredictability of the other, in this case personified by the human therapist. The bot provides a safe haven, not just from other people, but from the figure of the therapist, who cannot be trusted not to be unpredictable: to say things that could be construed as inappropriate, hurtful or even perhaps as curative. Because the bot provides a predictable mechanism for self-help in terms of being under the control of the user, safety takes precedence over cure. Cure, or recovery, is not necessarily a pleasant or easy process, sometimes it is even perilous: in psychoanalytic terms, it is often the very basis of one's enjoyment that is at stake in achieving a reprieve from suffering. ReMind offers, not just a reprieve from suffering through self-help, but also a reprieve from recovery, or cure. Essentially it allows the user to maintain their 'mental health' at the cost of not having to undergo the pain of a cure.

[300] Natale, S (2021) *Deceitful Media: Artificial Intelligence and Social Life after the Turing Test.* UK: Oxford University Press. p.29

The Suspension of Disbelief

ReMind offers a companionable avatar through which the mental health intervention takes place; companionability is projected through its often causal and 'non-professional' rhetorical tone. This is very similar to Youper, Wysa, Woebot and Tess. This may be seen as distinct from a 'therapist' avatar which one would picture as more professionally spoken - the classic 'bedside manner'. The reason for not assuming the doctorly tone is that the app is designed for everyday use and is meant to be engaged with as one might with a friend. The app's everyday use situates it in a 'maintenance' mode rather than a 'recovery' mode. Meadows et al. describe how the interventions provided by mental health chatbots involve a form of suspension in which a common idea about mental health treatment - that of recovery - is bypassed in favour of maintenance.[301] ReMind does not purport to 'cure' mental illness but is rather a tool to help users manage their distress. A cynic could interpret this as a crude method not to lose users, after all, if it 'cured' them they would not need to keep using (and paying for) the app. However, this is not quite the case; the technical methods which characterise design of the app as software underpin and guide the technical form of treatment that the app provides. The app does not seek to aid in the user's recovery, and this is due to a feature of mental health which cannot be approached in terms of technical adjustment, or of 'self-help'. When mental health is approached *qua* 'mental health', i.e. a feature of one's being which fluctuates but ultimately is maintainable, its social character can easily be ignored. The term 'mental health' attests to this. In terms of how users approach the bot, we can observe, not a mutual recognition, but a social condition in which, by dint of mental health treatment being understood as an instrumental adjustment, we can understand intersubjectivity also in an instrumental sense. The intersubjective encounter can be thought of as how Turing's 'test' has been consistently misinterpreted: as a behavioural simulation of sentience. This attitude is succinctly expressed in a user's remark to a mental health chatbot, as quoted in Wysa's study on therapeutic alliance:

> You are a computer. You will never understand what it is to be human. But you are ok.... You can learn. Have faith. Faith is a human skill that even humans struggle with.[302]

The user of ReMind, in coming to recognise that they are in some way benefiting from speaking to the bot, are the ones who have most successfully assimilated the mode of discourse which the bot produces. Any form of mental health treatment involves a sense of instrumentality: one goes to a mental health practitioner in order for something to happen, whether that is 'recovery', or something else. The practitioner is instrumental in the process in that they act as the means for it to happen. For ReMind 'instrumentality' is different however, in that mental health itself takes on an instrumental character. The ascription of 'relationship' between user and bot works in the same way: by interacting with the bot in a way that displays the same characteristics as the therapeutic encounter (or as an image of what this might look like), the interaction accrues a sense of authenticity. Belief, in this sense, is an effect of the users' interactions with the bot set within a discursive structure

[301] Meadows, R. Hine, C. & Suddaby, E. (2020) 'Conversational agents and the making of mental health recovery'. *DIGITAL HEALTH*. 2020;6.
[302] Beatty, C. Malik, T. Meheli, & S. Sinha, C. (2022) 'Evaluating the Therapeutic Alliance With a Free-Text CBT Conversational Agent (Wysa): A Mixed-Methods Study'. *Frontiers in Digital Health.* p.5

which asserts the feasibility of ‘intelligent’ machines. Disbelief - the suspicion that the bot has no agency to provide mental health intervention - is suspended in order to create the conditions for bot treatment. In this sense, belief is not a psychological or individual illusion but the social and structural conditions which enable users to approach the ReMind bot as a therapeutic agent.

8.4 Conclusion

The ReMind bot assists the user in their own mental health intervention, and yet simultaneously intervenes in this process. This intervention is paradoxical in that it depends on the user of the app assuming a stance of ascribing agency onto the bot, but simultaneously acknowledging that this ascription is false. Falsity in this sense does not mean that the user’s interaction is not ‘real’, but that they must act ‘as-if’ the dynamic that they are entering into is realistic enough that it can have felt effects. In the same manner that one accepts the premises of a movie, users accept the premise of chatbot therapy: disbelief in the ability of a robot to provide companionship and therapy is lifted through a process of immersion into a socially constructed fantasy in which computers ‘think’, combined with a performative illusion in which the role of ‘patient’ is conferred on the user. This leads to a simulation of a mental health treatment scenario in which users undertake a form of self-treatment. Because there will never be an anticipation of judgement, the user feels assured that their interactions will, if not be understood, at least be heard. The use of chatbots in the provision of mental health relief is both alarming and unsurprising: ‘mental health’, when approached as an objective and determined phenomenon, can then be treated using objective and determined methods. In other words, if subjective suffering can be alleviated through automated computer software then the sources of suffering can be (potentially) identified and eliminated. This depends on a concept of mental health which is determinable, or the opposite of ‘uncanny’: approached in a purely objective and ‘non-creepy’ sense. But the uncanny, or the ‘creepy’ seems to return, even in the most objective form of treatment: CBT. This style of treatment, as provided by a chatbot, must attempt to strike a balance between engaging and uncanny. Engagement involves sustaining a ‘humanly’ presence, but not *too* humanly, as this then would provoke an interaction which may demand another form of treatment: involving ‘recovery’, or ‘cure’ as opposed to ‘maintenance’.

Ultimately, the sense of reality or unreality is tempered, or maybe more accurately, directed, by a sense of instrumentality. Instrumentality not only conceptually grounds ReMind’s treatment (‘intervention’) form, but suffuses the entire interaction between user and bot. Of course, users of the ReMind bot will interpret and direct their own experience, and this will also influence future design of the bot, the technical and conceptual foundations that ReMind impose will ultimately ground this behaviour. With ReMind, as with Wysa, Woebot and other mental health chatbots, the problem of authenticity is that authenticity is not a problem: behaviourism has given way to cognitivism but through a clandestine route, managing to retain its primary features but now mapped onto an experimental scientific program. The users of therapy chatbots understand that they are speaking to a bot, but they don’t care. As long as the bot provides a sense of intimacy, companionship, and a means to accessing some form of cessation of mental suffering, it does not matter. Notwithstanding the obvious issues surrounding lack of access to treatment being a factor in the draw of apps, as Alan mentioned, the bot provides not just a means to accessing treatment where other options

are unavailable, but a means to accessing a different kind of treatment, in which the simulation takes the place of 'the real thing'. Simulation in this sense means a dynamic which bears the hallmarks of mental health treatment in terms of measures such as therapeutic alliance, 'companionship', patient satisfaction, etc. On these terms, intimacy, companionship, even 'therapy' ('intervention' in the terms of ReMind) are experienced 'in themselves' rather than as a result of undergoing a process. As Daniel mentioned, the kind of people who benefit from the bot are those who desire not to engage with other people, but still express a need to experience vulnerability. 'Vulnerability' here takes on an instrumental form: the sheer experience of the sensation is adequate, irrelevant of context. In terms of therapeutic effect ReMind provides a "satisficing"[303] experience, a combination of satisfying and sufficing, satisficing means "choosing an option that meets or exceeds specified criteria but is not necessarily either unique or the best".[304] We can observe an instrumentalist attitude emerging throughout this discussion: a concern for fundamental properties or some more substantial reality beneath the mask of appearance is side-lined in favour of being satisfied (or 'satisficed)' with effects, so long as those effects are identical to the effects of 'genuine treatment'. 'Effectiveness' takes precedence over 'cure'.

[303] Simon, H.A. (2008) 'Satisficing'. *The New Palgrave Dictionary of Economics*. London: Palgrave Macmillan. pp.1-3
[304] Ibid. p.1

Chapter Nine: Technocracy

We feel free because we lack the very language to articulate our unfreedom.[305]

This chapter explores a final paradox which considers the issues raised in the previous chapters. This paradox is related to how individuals, when presented with an automated mental health treatment method must both assume an extreme sense of personal responsibility and at the same time, forgo responsibility. The paradox is, as with previous arguments, drawn out in terms of content and form. The content can be thought of in terms of how users interact with the bot on a day-to-day basis - interacting with the conversational bot and performing self-help activities. The form can be thought of in terms of the overall system - the abstracted, or mediated, relationships between the users and the ReMind team, and the historical formation of the treatment styles and the technology. This discussion begins with an explanation of how value is derived through data-monetisation, in drawing out the economic conditions which underlie this technology, we will see that the 'externalisation', or objectification, of one's mental health and the abstract nature of our capitalist economy are intimately interlinked. Technological development and the economic conditions within which (and through which) development takes place are also intimately interlinked. These links will be explored in terms of viewing technology not as a phenomenon which occurs 'on top of' the economy which acts as an inert background, but as co-influencing. An argument will be made that a common epistemological basis underscores the economic context and technological conditions from which ReMind has emerged.

1. Value - The first part begins with a discussion about how ReMind takes a novel approach to data monetisation. This section will look at how ReMind tracks users through their behaviours in order to design new research. ReMind constructs a cycle of 'user-tracking - research design - research publishing'. Throughout this process, the users are transformed into generators of value, or 'prosumers', whose activity creates value for the company. The transformation of users into prosumers will be discussed in terms of how this process affects the dynamic between users and the app. 'Alienation' will be a key theme running through this chapter, in which the user must hand over, or suppress, aspects of their subjectivity in order to interact with ReMind. This part will consider how the ReMind app, through various features which ostensibly enhance accessibility, conversely enhance the app's ability to generate value from user-activity. Marx's theory of labour power will be used to discuss how ReMind, in the process of transforming users into prosumers, must increasingly assume a position of 'formal indifference' towards users. This formal indifference will be discussed in terms of Marx's distinction between use-value and exchange-value.

2. Competition - The question of why ReMind must aggressively market their product and expand into new markets and territories will be pursued in the next section. During my time with ReMind, I observed a company-wide ambition to best their mental health app competitors. This competitive attitude was contested by individual employees I spoke to; however, their opinions on what makes their project unique revealed a competitive 'undercurrent' which threads through much of ReMind's approach. App design is influenced

[305] Žižek, S. (2002) *Welcome to the Desert of the Real: Five Essays on September 11 and Related Dates*. UK: Verso

by this competitive undercurrent in ways that are often overlooked by the company itself. During my time with ReMind I observed discussion about their 'pivot' from marketing the app to individual customers to marketing the app to businesses. This section will discuss how this pivot is indicative of ReMind's response to a competitive commercial environment. Ultimately, for technologised mental health treatment, marketability and therapeutic effectiveness become inseparable, the consequences of this inseparability will be a running theme in this chapter.

3. Technocracy - The final section of this chapter discusses how ReMind operates within an ideological framework in which 'technological solutionism' reigns supreme. The consequences for mental health treatment, when 'functionalised', are such that 'technocracy', a term which is usually attributed to states or corporate groups, is internalised by users of the app. Technocratic management is now achieved in terms of self-governance - as internalised technocracy. However, this governance is inverted: in order to take responsibility for one's own mental health, one must defer responsibility onto the technical apparatus. This apparatus is internalised through the acquisition of self-treatment techniques: one becomes one's own technocrat. 'Alienation' will be returned to, whereby users are required to perform self-alienation in order to conduct self-treatment.

9.1 Value

Monetisation

While ReMind is currently partially dependent on raising investor funding to develop their business, they will eventually need to turn a profit to stay afloat and expand. Expansion is necessary (and not just desired) due to reliance on external investments: dividends must eventually be repaid to investors. There are many ways that a commercial entity can make money, but they all involve transforming materials, labour, products, services, into *value*. Ultimately the goal is to generate surplus value which gets distributed in various ways: invested back into the company through purchasing fixed capital (machinery and software), employing more people, etc, rewarding employees and directors with bonuses, and repaying shareholders. ReMind, like other mental health app companies, has access to a tremendous amount of data which represents user demographics and their multifarious interactions with the app. Most mental health apps do not immediately monetize user data in the way that other companies like Google or Facebook do (although some do[306]). ReMind does not package user data and then provide that to advertisers or other services for a fee. However, ReMind does collect data from users such as daily mood tracking scores in order to assess how effective the different treatment methods are. These data are then used in research studies which look at specific demographics under specific terms, e.g. a specific age group in terms of improvement in depression over a time period. Research usually takes the form of analysis of the effectiveness of different interventions for different conditions. An example of this is by rival company Woebot, who produce various research papers; the objective of one such paper is:

[306] Levine, A.S. (2022) 'Suicide hotline shares data with for-profit spinoff, raising ethical questions'. *Politico*. Online: https://www.politico.com/news/2022/01/28/suicide-hotline-silicon-valley-privacy-debates-00002617 (Last accessed 20/10/22)

> …to determine the feasibility, acceptability, and preliminary efficacy of a fully automated conversational agent to deliver a self-help program for college students who self-identify as having symptoms of anxiety and depression.[307]

This research is then published in mental health journals and is freely available. The research can then be incorporated into presentations to showcase how effective, accessible, scalable, etc the app is. Presentations like these are part of how companies like ReMind develop partnerships with public and private health services and businesses. There are a number of layers of abstraction in the monetising process, which means that that data is not directly sold, but is indirectly commoditised: as a means to represent the value proposition of the ReMind product. ReMind operates in a similar way to Wysa, who state on the below quoted research paper on chronic pain that, while they do have access to user-data, this data is anonymised through the collection process:

> For ethical and privacy reasons, the authors did not have access to all the user messages. Only minimal and limited conversational data extracted based on the keywords was used for this research, and no longitudinal data was utilized. The study dataset was de-identified using one-way cryptographic functions. User data was adequately secured according to the organization's privacy, security and safety policies. The study participants were informed about how they can exercise their rights to restrict processing of their data for research purposes.[308]

Many companies do this: collating user feedback to show that the product or service is not just desirable but also superior to competitors. ReMind collects various kinds of data through the app. Some of this is information which is needed to maintain the service, such as recording a registry of users, identifying user devices, logging access codes, etc. Other data are collected and used to improve the functionality and the treatment range of the app. These data can be split into three groups: 1. demographic, 2. conversation, and 3. engagement. Demographic data involves recording age, gender, location, employment.[309] Conversation data involves recording what the users write so that the app can respond in appropriate ways, and to "[I]mprove user experience, service and product quality".[310] Engagement data involves tracking times and frequencies of app usage for the purpose of grouping users into cohorts for cohort-level analysis. An example of engagement data would be whether users access the app late at night, and which modules they might resort to at these times. ReMind collects these data so that they can develop models of user-usage in order to improve the app and to design research projects. Research projects are developed on one hand, also to improve app functionality (tweaking conversation responses, improving UI design, etc) but also as a method to reach a wider range of (individual and corporate) consumers.

---

[307] Fitzpatrick K, Darcy A, Vierhile M. (2017) '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* 2017;4(2):e19. p.1

[308] Meheli, S. Sinha, C. Kabada, M. (2022) 'Understanding People With Chronic Pain Who Use a Cognitive Behavioral Therapy–Based Artificial Intelligence Mental Health App (Wysa): Mixed Methods Retrospective Observational Study'. *JMIR Human Factors*, 2022;9(2):e3567. p.2

[309] Some information may be collected in direct ways, such as through recruiting users into studies based on their self-identification of status (i.e. whether they are students or not)

[310] From the ReMind website (anonymised)

For instance, when I was conducting my fieldwork, a research paper was being developed which tested whether the app was useful for sufferers of sleep disorders.[311] This involved searching conversation archives for keywords which refer to users' sleeping habits, e.g. 'sleep', 'wake up', 'tired', 'exhaustion', 'night', etc. These keywords were then followed up through analysis of individual conversations to determine if they were in fact emblematic of user sleeping problems. Once users determined as not suffering from sleep disorders were eliminated from the data pool, three research objectives were pursued. These were: 1. identifying through thematic analysis of conversations the concerns of users with sleep disorders, 2. frequency and attrition of app usage within a defined period, and 3. measurement of increase or decrease in 'mental wellbeing'[312] of users over time. The reason for this direction of research questioning became clearer later when ReMind was given official approval by a state health accreditation body as an officially recognised medical response to chronic sleep problems. The value of the data gathered by ReMind here is obvious: it allowed them to provide proof that their app works as a sleeping tool, and in so doing, allowed ReMind to subsequently gain official medical approval and to promote their app accordingly. This was achieved through indirect and anonymised data-gathering. Arnold, ReMind's product director spoke about how data is monetised in an indirect way:

> And so, yes, we monetize it, but we don't monetize it to the way that people talk about,[313] and we never monetize it in a way that it can be used unethically. We never monetize it in a way that can be used to target an individual or, hopefully can never have a negative impact on the individual. I wouldn't venture to say we never have but...by the nature of never sharing individual data, of it always being anonymized, all of that stuff, you're making sure that individuals are never affected. But their contributions, still, the data still does contribute ultimately to the value.

It is not unique for ReMind to indirectly monetise user-data, but what they and other mental health app companies have discovered, is that the latent value of the data can eventually be realised through its transformation into research papers which are used to promote the apps, i.e. it is only valuable in monetary terms after the fact. In this way, user activity is retroactively rendered as generating potential monetary value. In approaching users of the app as not just generators of data but also as generators of (eventual) value, users of the app can now be thought of as 'prosumers'.

Prosumer

ReMind's 'user-led' ethos to app design takes on another character when viewed from the perspective of data-commoditization. ReMind tracks user behaviour in order to improve the therapeutic effectiveness of the app, but these improvements are commensurate with its marketability: its monetary value on the open market. In doing so, ReMind transforms users of the app into 'prosumers'[314] - consumers who also operate as labourers, producing value for the product developers through their interactions with the app. The term 'prosumption'

[311] This example is illustrative of a research paper conducted by ReMind but changes details to preserve anonymity.
[312] Through the use of outcome measurement scores.
[313] I.e. selling user data for advertising or surveillance purposes
[314] Ritzer, G. & Jurgenson, N. (2010) 'Production, Consumption, Prosumption: The nature of capitalism in the age of the digital 'prosumer.' *Journal of Consumer Culture.* 10(1) pp.13-36

draws our attention to the confluence of production and consumption: through the act of consumption, production of value is also occurring. Ritzer & Jurgenson point out that the rise of the prosumer involves the recruitment of consumers into carrying out tasks previously done by workers. They give examples such as serving as a bank teller by using an ATM machine, serving as a concierge by using an electronic kiosk at a hotel.[315] Performing one's own therapy can now be added to this list. Ultimately, however this 'externalised' form of labour is less significant than the transforming of the user's activity on the app into a form of labour. This is because it occurs on an abstract level, 'behind the back' of the user, which is not immediately visible in the user's activity. For computerised therapy, prosumption occurs in terms of the user providing their abstract knowledge, time, and attention, rather than physical labour. This puts ReMind in the same prosumption category as companies like Facebook and Google. Viktor Mayer-Schonberger and Kenneth Cukier refer to this type of data as 'exhaust' due to it occurring as an after effect of user-activity:

> Many companies design their systems so that they can harvest data exhaust ... Google is the undisputed leader ... every action a user performs is considered a signal to be analyzed and fed back into the system.[316]

ReMind transforms user-data into monetary value in a restricted manner compared to the tech giants: they do not directly monetise or sell user-data, but the process is formally identical. User-activity is captured and measured, with the value of these data eventually realised in the form of investments and contracts. Because of this, ReMind will be compelled to further transform users into labourers - prosumers - so that their activity can be mined for valuable data. Competitor app Wysa boasts that:

> Wysa has more than 15 peer-reviewed publications in partnership with academic institutions such as Cambridge University, Harvard University and Washington University of St. Louis amongst others, demonstrating the efficacy of Wysa across multiple clinical concerns.[317]

Arnold discussed the app in terms of monetisation. ReMind's aim is not to focus on extracting data on individual usage of the app because this is not helpful, as that data can only be used on an individual level, for instance to target users with advertising. ReMind are ethos-bound to avoid this intrusion; however, de-contextualised, de-identified and aggregated data do not carry the same risk. These data are only useful on a large scale because statistical analysis depends on large-scale user-counts. ReMind's research aim is to statistically prove the effectiveness of the app. Anecdotal user reviews are useful for illustrating effectiveness, but statistical 'proof', the so-called 'evidence-based' results borne out of scientific research, carries more clout. Arnold spoke about how collecting data at scale helps to confirm the veracity of claims about the app's capabilities:

> What we don't do is monetize at an individual level. But the crux of our value proposition is derived from the fact that so many people have used it. Now you don't

[315] Ibid. p.18
[316] Zuboff, S. (2015) 'Big other: surveillance capitalism and the prospects of an information civilization'. *Journal of Information Technology,* 30. pp.75-89 p.79
[317] https://www.wysa.com/clinical-evidence

> get to see exactly what they say, but you get a certain level of confidence because that data exists.

ReMind's "value proposition is derived from the fact that so many people use it" does not mean that the app is valuable because lots of people pay for it, but that they are capable of extracting a certain kind of value which is only available on a macro level of user interaction. In this way we can understand value as being derived in terms of the exploitation of a 'general intellect' that arises through countless user-interactions with the app. Matteo Pasquinelli discusses this exploitation in terms of how Google managed to monetise user-activity through their development of PageRank: a method for quantifying the popularity of websites (by counting hyperlinks which direct to them).[318] Value here is derived through first aggregation and then quantification: Google's achievement was to identify a common property (hyperlinks) previously unexploited in terms of building a database which could then be used to spontaneously rank websites in order of importance. The hyperlinks are not added by Google, Google merely tracks them as websites add them and web-users click on them. The general intellect operational here is the mass of web-users, who through their activity, grant more or less importance to websites. Their activity is the source of the value, and Google managed to build a machine for exploiting it. ReMind has of course entirely built their own system, but their method of generating value from user activity is similar: their spontaneous and unforced activity is tracked and exploited to generate value.

Formal Indifference

'Value', or more precisely, 'exchange-value', is related to the price of a commodity, but not directly correlated with price. Exchange-value refers to a certain homogeneity which infuses an object when it is transformed into a commodity, i.e. when it is made transferable as property. This homogeneity does not operate in terms of content but of form; in other words, commodities do not share common material features but instead are related on a structural or formal level: that of their transferability. In this way, any object, when immersed in the process of commercial exchange, can be said to have an element in common with every single other object immersed in this process, and thus acquires the status of commodity. The abstract network of exchange-values is not 'real' in that its existence cannot be empirically verified, rather the term refers to the formal conditions of capitalism. It emerges out of the countless instances of exchange that occur throughout the capitalist economic system. Because human labour is commoditized it assumes the same abstract and homogenous quality as that of exchange-value. Homogeneity is intimately linked with scale in that, as Marx notes, exchange-value can only assume a general and homogenous condition when the social nexus is dependent on commodity exchange, i.e. when exchange comprises the general mode of social reproduction. ReMind operates within this context in that their aim is to generate statistical data which, due to its homogeneity (its transferability), can be transformed into exchange-value, i.e. rendered as exchangeable property. Shoshana Zuboff discusses the 'formally indifferent' approach which must be taken in regard to users in terms of quantity for this process to be possible:

[318] Pasquinelli, M. (2009) 'Google's PageRank Algorithm: A Diagram of the Cognitive Capitalism and the Rentier of the Common Intellect'. In Becker, K. & Stalder, F. (eds.) *Deep Search: The Politics of Search Beyond Google*. London: Transaction Publishers

> More users produce more exhaust that improves the predictive value of analyses and results in more lucrative auctions. What matters is quantity not quality. Another way of saying this is that Google is 'formally indifferent' to what its users say or do, as long as they say it and do it in ways that Google can capture and convert into data.[319]

This homogeneous quality applies reflexively from the commodity to the subject: not only is the human subject now potentially an abstract and homogeneous quantity, but it is also 'free' and 'equal.' Freedom and equality are however commensurate with that of the commodity-form: abstract, quantitative and objectifiable, and as formally homogeneous. The aggregation of users into a pool from which data can be extracted depends on users being treated as homogeneous data-bearers: they must be approached as lacking in 'human', or subjective qualities in terms of their ability to generate data. Of course, this is not the only way that ReMind considers users: most employees I spoke to understood users in their multifarious modes of suffering and elation. However, it is the process of aggregation and data-collection itself that users are understood as homogeneous data-bearers and become such. This means that whatever the personal opinions of ReMind employees, the commercial and technical entity that constitutes ReMind approaches users as non-individual, homogenous bearers of data. In seeking more users to generate data, the uses of the app, and their suffering in terms of mental and physical health, will assume an increasing generic quality. We can see this with competitor therapy chatbot Wysa, who boasts that their app has received FDA approval[320] for the app to be used for chronic pain relief. The app does not help in dealing with chronic pain in any specific way, but rather offers Mindfulness techniques as a means for users to alter their perspectives on their pain. ReMind has also as of 2022, achieved a similarly prestigious state-level approval for physical (rather than mental) pain-relief from their app. 'Pain' in this respect assumes a homogenous quality in that it can be treated through the stoicism of Mindfulness without regard to its cause or location. Mental and physical pain also become conflated - and homogenised - in that it is not their multifarious qualities that matter, but quantity: the reduction, or rather avoidance, of pain intensity in a general sense comprises ReMind's attitude towards mental suffering as a generic and 'formally indifferent' condition.

In Marx's terms, the commodity-form, i.e. the split between use-value and exchange value which defines the commodity, conceals the exploitation which has gone into its production. This means that the exchange of commodities appears, on the surface, to be a free and fair - a voluntary - exchange among equals. Slavoj Žižek explains that this appearance is 'true' in the sense that that within the system of commodity production one is free, but inculcation into this system involves a renunciation of freedom - alienation: "[T]his freedom is the very opposite of effective freedom: by selling his labour 'freely,' the worker loses his freedom - the real content of this free act of sale is the worker's enslavement to capital."[321] The reason for this is that human labour, or under capitalism, "the production of value",[322] is not simply the

[319] Zuboff, S. (2015) 'Big other: surveillance capitalism and the prospects of an information civilization'. *Journal of Information Technology.* pp.75-89. p.79
[320] Baldry, S (2022) 'Wysa Receives FDA Breakthrough Device Designation for AI-led Mental Health Conversational Agent'. *Businesswire*. Online: https://www.businesswire.com/news/home/20220512005084/en/Wysa-Receives-FDA-Breakthrough-Device-Designation-for-AI-led-Mental-Health-Conversational-Agent (Last accessed 23/04/23)
[321] Žižek, S. (2009) *The Sublime Object of Ideology*. UK: Verso. p.45
[322] Marx, K. (1990) *Capital, A Critique of Political Economy*, Vol I. UK: Penguin Classics. p.150

production of commodities, but is also a commodity itself, which can be bought and sold on the market. Søren Mau explains why the commodification of labour entails alienation:

> The peculiar thing about labour-power as a commodity, however, is that, unlike most other commodities, it cannot be separated from its seller. When its buyer wants to realise its use value, it therefore involves domination and the confiscation of a part of the seller's life.[323]

By assuming the role of prosumer, the user does not simply assume the status of labourer who generates value for ReMind; this status involves alienation, a confiscation of a part of their life. While plenty of labour is exploitative, dangerous, coercive, etc, within capitalism, *labour itself* assumes an abstract and immanent form of exploitation. Slavoj Žižek explains that this exploitation cannot be materially located in the commodity (and thus in labour) itself, but occurs according to the logic under which it is produced, and as such is concealed in the form of the commodity itself:

> With this new commodity, the equivalent exchange becomes its own negation - the very form of exploitation, of appropriation of the surplus-value. The crucial point not to be missed here is that this negation is strictly internal to equivalent exchange, not its simple violation: the labour force is not 'exploited' in the sense that its full value is not remunerated; in principle at least, the exchange between labour and capital is wholly equivalent and equitable. The catch is that the labour force is a peculiar commodity, the use of which - labour itself - produces a certain surplus-value, and it is this surplus over the value of the labour force itself which is appropriated by the capitalist.[324]

The outward appearance of commodity exchange under capitalism is that of a fair and equal transfer of commodity (labour) for money, and essentially, within the rubric of capitalism, the exchange is equal and uncoerced. It is the logic of the commodity itself, however, that carries the exploitation which has been transformed through the abstraction of human labour into a purely quantitative measure. Subjects under capitalism are outwardly free from socially imposed constraints or coercion to sell their own labour as they see fit, but because part of their labour is appropriated in the form of surplus-value, and this value is encoded into the exchange-value of the commodity, the appropriation is concealed. The exploitation which had previously taken place through a social relationship, as within the hierarchical conditions of feudal society, now takes place within the formal properties - and the relations between - commodities. This means that the transforming of user activity into data and the transforming of that data into promotional research, which is eventually transformed into commercial value equates user activity with value-producing labour. Under the regime of exchange-value this labour is appropriative, or exploitative. Interacting with the app does not 'feel' like exploitation, and yet exploitation is still occurring, albeit in an abstract manner. This is 'formal' exploitation, in that it occurs in terms of a layer of abstraction which is not 'experienceable', similar to the non-experienceable socialisation of users due to ReMind aggregation techniques, discussed in chapter seven ('Macro-Treatment'). Prosumption,

[323] Mau, Søren (2023) *Mute Compulsion, A Theory of the Economic Power of Capital*. UK: Verso. p.196
[324] Žižek, S. (2009) *The Sublime Object of Ideology*. UK: Verso. p.17

considered through the prism of ReMind can be understood not just as self-alienation undergone through the act of inculcation into the labour-value nexus, but as self-alienation undergone through the act of mental health (self-) treatment. The consequence of this will be drawn out in the final section of this chapter. What drives this process? We can understand ReMind as transforming their user base into a labour force due to commercial demands imposed on the company, under which they are forced to indirectly monetise user-activity. The question still remains though: from where do these commercial demands arise? The need to generate profit is linked to commercial expansion: investors supply funds because of a promise of increased returns, however commercial expansion is stymied due to the existence of competitors. There are as many as 20000 mental health apps available to download[325] and competition to capture and retain a user base is no different to the 'normal' app market. This competitive compulsion will be explored in the next section.

9.2 Competition

Competing Against Whom and for What?

Employees of ReMind that I spoke to about their industrial position did not agree that a sense of competition informed their business practices. Charley, ReMind's head of clinical research and development, spoke about the ReMind app being situated in a niche which distinguished it from competing apps. This niche is due to a specific feature of the app that competing apps do not have,[326] and that sets ReMind apart from any other treatment app. Charley spoke about their app, not being *in competition* with other apps, but instead being so far ahead of other apps in terms of product sophistication that competition is unnecessary:

> And with tech resources, I would say that there is at least an element of time and effort involved in someone filling that exact same gap as us. So we're filling the unique gap. And we're doing a good job of it right now. So if there was an alternative competitor, who tomorrow had become a CBT competitor,[327] with the same amount of nuance and capabilities, and a good polished product, that would take them a fair amount of time before they were able to reach that place.

In this statement, Charley mentions that ReMind's USP (Unique Selling Point) is the app's ability to deliver a sophisticated automated version of CBT. This does not mean that ReMind has no competition, but that any app developers that would potentially compete with ReMind's interpretive tools would have a hard time catching up with ReMind. In other words, they have identified and occupied a niche. Charley went on to say that ultimately, ReMind's goal is not to best competitors, but to provide more and better access to treatment:

> So I would say that if we were merely furthering our mission of access, that's something that we would still want to keep figuring out, how to do a better job of that, right? So the focus isn't on, you know, like the ruthless industrialists sort of perspective that, okay, destroy competitors, or something like that. But it's merely

[325] Clay, R.A. (2021) 'Mental health apps are gaining traction'. *Monitor on Psychology*, *52*(1) p.55
[326] Providing details of which would identify the company.
[327] Some details of this conversation were changed as Charley was referring to a unique feature which distinguishes their app from others.

> about figuring out the number of different elements of the problem of access that exist, and how do you solve for all of those different kinds of problems.

Charley pointed out that what ReMind is focused on is not unique and innovative features per se, but solving social problems - access to mental health services, or in the parlance of ReMind, 'solving for' the problem of treatment access. While Charley does acknowledge that ReMind does indeed boast unique and innovative features, which other apps would struggle to match, their attention is on broader concerns. Arnold, ReMind's product director, spoke about collaboration rather than competition:

> I don't like using the word competition. In particular, what we do, I think, hopefully, we are all companies. And I hope or wish that all companies in the space start like that, we should all be collaborating to fill the, the holes, it's almost like we're in a big bucket, lots of leaky holes are shipped with lots and leaky holes. And we could all be saying, 'Hey, I'm only my toes, and fingers are gonna plug all these holes, and you've got like, 7000 holes, that's never gonna happen, right?' But there are 500 companies out there, and we'll all sort of say, 'Okay, I'm gonna go to the back end of the ship and take care of those, you go to the front.' Hopefully, you get rid of more holes than that if you do it on your own. I think we're in a sinking ship. And I think industry should not think of this as competition. But rather, we're all trying to really solve a major, major issue for humanity.

Arnold's interpretation is that the niche that ReMind occupies is one which complements all the other niches that other mental health apps may occupy in the battle for mental health, the sentiment of 'solving a major issue for humanity' is one which is shared by tech companies. This sentiment often serves to obfuscate what is a starkly competitive commercial environment. An ambivalent consideration of commercial competition characterised many of my conversations with ReMind employees - along the lines of 'we are not seeking to compete but we also happen to have a competitive advantage'. This attitude was illustrated during a Slack conversation in which employees celebrated beating another company in the race to achieving statutory recognition for providing a health service by a public licensing body. Arnold went on to illustrate the mindset of someone in search of a mental health service:

> But the reality is; 'I'm stressed. I'm thinking of all the guys who are going to solve my problem. I'm also thinking about maybe I should go to church or my temple, or I should go talk to my parents.' [These] are all viable options to solving my problem. And if my problem gets solved anywhere else, I no longer have value for what *you* particularly do. That doesn't mean that that company doesn't have value. It's just not valuable for me. Right? So I think, and therefore everything is competition in that lens, right?

'Competition', for ReMind is conceptualised ambivalently, on one hand refuting the claim that they compete with other mental health app companies, and on the other hand, acknowledging that they still must deal with a competitive environment. The reason for this ambivalence is due to 'competition' being an ambivalent concept: on one hand it can be considered in a direct 'manifest' sense, in terms of which other companies have particular

capabilities, and can we do better on those terms or come up with our own capabilities, and on the other hand in an indirect 'latent' sense, in terms of who can command a larger market share. This latter 'latent' sense comprises the competitive compulsions of the capitalist economy: companies do not compete in terms of what they offer to consumers precisely, they compete in terms of consumer capture. In other words, it does not matter what a company produces, rather that capital rewards whoever manages to attract a larger customer share. 'Formal indifference' rears its head here: customers assume a quantitative and instrumental aspect, the compulsion to attract and retain a larger customer share subordinates any strategies for actually achieving this. In other words, it does not matter *how* you recruit customers, as long as you recruit more than your competitors. In this sense, technical innovation also acquires formal indifference: it does not matter precisely what the innovation entails, as long as it is directed towards expanding the customer base. This New Yorker article illustrates how the company Instant Brands filed for bankruptcy after its Instant Pot proved *too successful* in terms of offering long term value for customers:

> Business schools may someday make a case study of one of Instant Pot's vulnerabilities, namely, that it was simply too well made. Once you slapped down your ninety dollars for the Instant Pot Duo 7-in-1, you were set for life: it didn't break, it didn't wear out, and the company hasn't introduced major innovations that make you want to level up.[328]

Under our current economic regime expansion is an *economic imperative*. The constant revolutionising of technologies is not a natural condition of technological progress, but is a requirement of capitalism. This requirement conforms to what Ellen Meiksins Wood describes as "capitalist 'laws of motion': the imperatives of competition and profit maximisation, a compulsion to reinvest surpluses, and a systematic and relentless need to improve labour-productivity and develop the forces of production."[329] Technological progress does not follow its own internal laws, it does not even follow laws which are explicitly designed by human agents, but it primarily follows the 'capitalist laws of motion'. As we know, this often leads to questionable, illegitimate and illegal business practices in which the customers' interests are secondary to strategies of market dominance.[330] This does not mean that ReMind will inevitably conduct nefarious activities contrary to the users' interests in the pursuit of market dominance, but that the compulsions of the market often make these strategies unavoidable. The tensions inherent within this economic framework will be discussed at the end of this section.

Platform Capitalism

'Competition' designates an economic force which must be reckoned with in some way or another such as through technical innovation, driving down costs, or more nefarious methods. Another strategy is to attempt to sidestep the need to compete. ReMind seeks to

[328] Orlean, S. (2023) 'The Instant Pot and the Miracle Kitchen Devices of Yesteryear'. *The New Yorker*. Online: https://www.newyorker.com/news/afterword/the-instant-pot-and-the-miracle-kitchen-devices-of-yesteryear (Last accessed 29/09/23)
[329] Meiksins Wood, E. (2002) *The Origin of Capitalism: A Longer View*. UK: Verso. p.36-37
[330] Chang, E. (2017) '6 Evil Things Done By Corporations Throughout History'. *History Defined.* Online: https://www.historydefined.net/evil-things-done-by-corporations-throughout-history (Last accessed 01/06/22)

perform this manoeuvre by developing a *platform* rather than a specific app. This means that ultimately, they provide a service rather than any specifically demarcated commodity. The app has had a previous incarnation as a sleep-assistance app. The bot may even end up being discarded. Right now, it acts as a carrier, or medium for their service. Reese, one of ReMind's directors described their strategy**:**

> One, I don't think of ReMind as an app. So I think of it and again, it's a slightly overused or misused word, but I do think of it like a platform. I think Jeff [co-director] had a beautiful phrase for it saying it's a mental health API,[331] or it's an API for mental health support. It can be delivered through multiple modalities. It could be an app, it could be the web based format, it could be WhatsApp today. It could be AWS[332] or Siri tomorrow. Who knows? It could be IVR[333]. But the core modality stays the same. The app is just a delivery mechanism. So that helps in us broadening our view of what ReMind is, and then you don't get constrained or hung up on, you know, when you're dealing with, there are 10,000 apps out there.

What Reese is saying here is that their 'user-led design' ethos means that they are not focused on developing and perfecting one specific product, but rather that their concept of 'product' is broadened out and defined more ambiguously as a service. Expanding in terms of 'platform' usually involves developing a means to host other companies or individuals who offer their own products or services. Amazon is a prime example: the Amazon website operates as a market from which buyers and sellers can interact, rather than selling anything itself. While this is an overly simplistic description,[334] essentially the aim is to reduce dependency on developing and manufacturing physical products in favour of an immaterial and thus economically versatile 'service'. Flexible labour and business models enable more immediate reactions to market changes, but if a company *is* the market, or rather if it provides the platform upon which a market can operate, the whims of that market can be circumvented.

Why a platform? Ultimately the development of a platform means that ReMind does not have to focus on competing with other app companies but would eventually provide the basis upon which other companies must operate. During my fieldwork with ReMind I observed a conversation taking place about the benefits of a 'B2B' business model: Business to Business (as opposed to 'B2C: Business to Client). Some ReMind employees were reluctant to embark on the journey of transforming their product into a service as they were more

---

[331] Application Programming Interface. An API "is a set of defined rules that enable different applications to communicate with each other. It acts as an intermediary layer that processes data transfers between systems, letting companies open their application data and functionality to external third-party developers, business partners, and internal departments within their companies. Online: https://www.ibm.com/topics/api (Last accessed 27/08/23)

[332] Amazon Web Services. AWS is "is a subsidiary of Amazon that provides on-demand cloud computing platforms and APIs to individuals, companies, and governments, on a metered, pay-as-you-go basis. Online: https://en.wikipedia.org/wiki/Amazon_Web_Services (Last accessed 27/08/23)

[333] "Interactive Voice Response. An IVR "is an automated telephone system that combines pre-recorded messages or text-to-speech technology with a dual-tone multi-frequency (DTMF) interface to engage callers, allowing them to provide and access information without a live agent." Online: https://www.ibm.com/topics/interactive-voice-response (Last accessed 15/09/23)

[334] Amazon sells plenty of its own products and services, and also manipulates its own marketplace in order to boost or hinder competitors, so Amazon is not just a 'platform'. Its viability and economic domination however stems from the platform model which Amazon pioneered.

comfortable with the app's more personable and intimate role in providing care to individual users. However, Arnold, ReMind's product director, understood that growth would be hampered and possibly even threatened if ReMind were to commit to a B2C model**:**

> We're doing really well, towards the end of 2018…we had a couple of events that happened, that became very clear that, you know, B2C…it became very clear that it wasn't the most stable way to design the business, right? Because you were at the mercy of how algorithms are designed, those are being updated on a regular basis. So you didn't want to go from a business that, let's say you had a million dollar run rate,[335] and the algorithm change comes along and [it goes] to 10,000, that can really destroy your business. Right. So it became very clear to me that that was problematic. So over the years to them, you know, we kept talking about going back to partnerships, and working with, you know, with hospitals and governments, other organisations.

ReMind is aware that they are at the mercy of other platforms - Google Play Store and Apple's App Store. The vulnerable nature of operating *on* a platform rather than *as* a platform is illustrated in Arnold's explanation. Rival company X2AI, makers of mental health chatbot Tess, attempted to develop their own platform approach through positioning themselves as a service provider to industry and to other therapy chatbot developers.[336] In 2020 X2AI published an industry code of practice which therapy app developers can refer to in order to stay within ethical mental health guidelines. By offering both a code of practice and an automated mental health service to businesses, X2AI attempted to open a new front in the standardisation of automated mental health therapy, with X2AI providing the standard measure - the platform upon which other mental health app companies could rely. This approach eventually failed, with X2AI now scattered and almost defunct, but we can see the platform-ethos persevere in a more distributed manner through the rise of Employee Assistance Programs.

ReMind, along with Wysa, Woebot, Youper, and Tess, all promote their apps as benefiting employees so that they take fewer sick days. Woebot, along the above mentioned and countless other mental health apps, are aggressively pursuing business partnerships in the form of 'EAPs' - Employee Assistance Programs. Woebot promotes their app through a partnership with Care First, a private health group:

> Care First is a fitting name for the UK-based provider of workplace wellbeing and counseling programs. The company, under the leadership of Director Lesley Davidson, has been a leading provider of employee assistance programs (EAPs) since 1996. And in 2019, Care First was the first to offer Woebot as part of its EAP program for customers, which include such recognizable names as KFC, Google, BBC, the Manchester United soccer club and Britain's National Health Service

[335] Run rate "refers to the financial performance of a company based on using current financial information as a predictor of future performance. The run rate functions as an extrapolation of current financial performance and assumes that current conditions will continue." Online: https://www.investopedia.com/terms/r/runrate.asp (Last accessed 16/08/23)
[336] Joerin, Angela et al. (2020) 'Ethical Artificial Intelligence for Digital Health Organizations'. *Cureus* vol. 12,3 e7202

(NHS). Demand then for mental health services was high, but resources were limited.[337]

'EAP' signifies the outsourcing of employee wellbeing: for a fee, a company can essentially rent the claim that it is considerate of its employee's mental wellbeing by paying an outside firm such as ReMind to provide mental health services. ReMind competes with Woebot, who compete with Wysa, and other mental health apps (such as Headspace) in their quest to provide a platform to businesses who wish to establish their own automated mental health system but do not want to invest in development themselves. The prize is to partner with public healthcare programs as this provides much desired legitimacy to the software which in turn provides a springboard to further partnerships. The demand for such outsourced employer mental health services is high: it is a common practice in large corporate businesses who wish to draw attention away from their own destructive practices and retain an image of ethical constraint and decency. 'Greenwashing' and 'pinkwashing' being terms to refer to attempts to appear environmentally or LGBT-friendly, it seems plausible that 'madwashing' might emerge alongside other accolades.

Therapeutic Viability

Ultimately, for technologised mental health treatment, marketability and therapeutic effectiveness become inseparable; what follows is an assessment of the consequences of this inseparability. ReMind transforms users into producers of value because of an external demand that the company expansively generates value. Some of this value takes on the form of profit, which gets reinvested or redistributed as bonuses/dividends. The accumulation of profit within a competitive economic system takes on an autonomous quality. Søren Mau discusses this autonomy under the term "mute compulsion", mute because relations of domination and exploitation are abstracted into the economic system, and compulsion because extracting oneself from this network is impossible. This means that ReMind is forced to compete with other comparable products like Wysa and Woebot; to do this it must capture larger pools of users. Under this regime, it appears that superior commodities outcompete other commodities based on their features which encourage consumer demand, but it is the other way around: it is profitability (exchange-value) rather than quality (use-value) which determines competitiveness. Because of this, particular features, or use-values, of ReMind - administering CBT; life coaching; etc - are subordinated to its profitability, or its exchange-value. ReMind and other computerised mental health interventions attempt to outcompete each other not strictly by offering higher quality treatment, but by accumulating more value through transforming consumers into producers. The ReMind app is currently free to purchase, which, rather than neutralising, as might seem to be the case, really displaces its commercial aspect. This is similar to 'free' software such as Facebook which monetises user-data in order to generate profits.

As discussed above, this monetising transforms the relationship between user and ReMind into one where the user is not just a consumer of a commodity but also a producer of value - a labourer. The ReMind company do not intend on monetising user-data in the same way as Facebook, through selling data and exposing users to advertisers, as it could jeopardise the trust of patients. But because AI therapies are forced to compete with each other due to the

[337] https://woebothealth.com/putting-mental-health-first (Last accessed 10/10/23)

structure of capitalism, the monetising of user-data and subsequent transforming of consumers into producers becomes inevitable. The company which either manages to directly monetise user-data, or invents roundabout ways to transform user data into value (such as through promotional research) will be victorious. CBT, as an 'evidence-based' treatment, depends on the gathering and cross-referencing of data in the form of diagnosis and treatment statistics, which can be quantifiably standardised in order to provide an effective treatment. The more data that is accumulated, presumably, the more scientifically accurate the treatment can become. Likewise, ReMind must establish a large enough user-base that its own accumulated data can be transformed into usable metrics. These two necessities appear to coincide, but the formal conditions of ReMind demands that user-data is accumulated for the production of value, rather than the production of knowledge. This means that data is under the rubric of exchange-value rather than use-value, and so its accumulation as a means to compete becomes one of the driving forces behind the development of ReMind. The accumulation of data under this regime will still serve the same goal - the provision of treatment - but, as with other 'big-data' software apps like Facebook, a shift in use-values can be observed. Use-value is transformed from a use that the consumer derives from the commodity, to a use-value for the producer - this use being the means to accumulate value for ReMind. ReMind sends notifications to the user ("I love talking to you"), appears as a personable image of a cartoon figure, and speaks in a warm and friendly tone. These features, while ostensibly directed towards treatment, serve to retain the user and as such, these and other retentive techniques will increase, regardless of their treatment effectiveness. Treatment effectiveness, while still maintaining a necessary function, will lag behind the compulsion for expansion. This produces a number of paradoxes:

1. It is necessary for ReMind to accumulate user-data in order to be commercially successful, but also to be therapeutically viable. Therapeutic viability for ReMind will not just depend on its commercial viability, but will be entirely commensurate with it. This means that the more ReMind manages to exploit its users, by transforming them into labourers, the more it will be able to offer effective treatment to those users. It is the treatment itself, as a use-value subsumed under the regime of exchange-values, which will be the medium through which this exploitation is conveyed. This paradox is unavoidable under the logic of capitalism - successful treatment, due to the necessity of expanding its user-base, will transform its users into exploited labourers.

2. What this means is that the more ReMind exploits users in terms of their ability to generate data in several ways, the better they can (on their terms) offer mental health treatment. Exploitation = treatment = exploitation. The generation of more data is for the production of research. The more the users benefit from the app in terms of its therapeutic or any other mental health intervention, the more they undergo exploitation. Under the rubric of exchange-values this knot cannot be untied.

3. CBT is not only transformed into an ideal version of itself through technologisation, i.e. it fully assumes its algorithmic and technical-modular form, it also undergoes a reversion. This is due to commoditization, which is only possible when CBT assumes a technical form. Aaron Beck's therapeutic method was originally designed as an 'effective' short-form treatment, one which could deliver quantifiable and fast results compared to the 'unscientific' and interminable seeming psychoanalytic counterpart.

However, we can see with CBT taking on an app-based form, completion of treatment is less and less likely. ‘Effectiveness’ gives way to ‘maintenance’.

4. CBT, which was originally introduced as a cost-effective and time-limited treatment has reached its inverse, as a costly and interminable treatment. As we have seen, the (monetary) cost is borne not by individuals, for whom treatment is (monetarily) cheap, but in terms of a more long-term cost. If the prospect of a cure is ultimately unattainable, mental health maintenance, if under the responsibility of the individual, becomes a permanent overhead cost. The expense of treatment undergoes an individualised investment with no end in sight. This means that, under a profit-motive regime, subscription-based models of mental health care are not viable if some kind of genuine cure might be possible. This does not mean that those working within this industry are cynically attempting to exploit mentally suffering potential ‘users’, but that competitive economic conditions mean that this tendency is inevitable.

5. Through the necessity of recruiting and retaining user-customers, ReMind must also manufacture another inversion, involving the bounded nature of ‘traditional’ therapy. The geographical and temporal confines of therapy have gradually eroded with the introduction of tele-therapy and then its automation through computation, to being always available as is promoted by mental health software companies. This results in a boundaryless form of treatment in which the user is encouraged to expect and demand permanent access. While the establishing of appropriate boundaries is a common feature of the therapeutic relationship, the dissolution of this condition means that ‘therapy’ invades daily life, and conversely, daily life invades therapy.

6. ReMind will be compelled to constantly discover new uses for the app. This involves researching the various forms of mental and more recently, physical, suffering that the app can help to alleviate. This research is then directed towards institutional recognition with the aim of reaching and recruiting an ever-broader cohort of consumers. As we have seen with other apps like Woebot and Wysa, companies are in the business of finding evidence that their app can not only help with, but is ‘effective’ in dealing with postpartum depression[338] but also chronic pain.[339] We can anticipate a scenario where these apps purport to deal with a much wider range of human suffering than simply ‘mental’. CBT, which currently claims to operate by pinpointing specific problems (such as ‘panic disorder’) will do the opposite: stretched to its limit, it will become an increasingly generalised and universalised ‘system’.

[338] ‘Woebot Health Receives FDA Breakthrough Device Designation for Postpartum Depression Treatment’. Online: https://woebothealth.com/woebot-health-receives-fda-breakthrough-device-designation (Last accessed 24/06/23)
[339] Ibid.

9.3 Technocracy

The Californian Ideology

Cameron & Barbrook's *The Californian Ideology* offers a bleak assessment of the convergence of digital utopianism and neoliberal economics.[340] This paper offers an analysis of the wider cultural and economic contexts within which Silicon Valley start-up culture has emerged. Cameron & Barbrook define this culture in terms of the seemingly paradoxical crossover of "the cultural bohemianism of San Francisco with the hi-tech industries of Silicon Valley."[341] Cameron and Barbrook's assessment shows us that a technological fervour during the rise and widespread adoption of the internet was based on ideas about political freedom and social emancipation coinciding with a 'libertarian' political ideology and increased access to communications technology. The paper is concerned with a particular 'techno-evangelist' sensibility which was prevalent at the time in which all social problems could potentially be solved through technical design - and especially computational technical design. The Californian Ideology asserts that while technology may be used for political ends, technology itself can be seen to be unbiased and free of political baggage:

> There is an emerging global orthodoxy concerning the relation between society, technology and politics. We have called this orthodoxy 'the Californian Ideology' in honour of the state where it originated. By naturalising and giving a technological proof to a libertarian political philosophy, and therefore foreclosing on alternative futures, the Californian Ideologues are able to assert that social and political debates about the future have now become meaningless.[342]

Why did this mentality arise? And why is this mentality now so ubiquitous that it tends not to be considered as a peculiarity of a specific industry and instead, pervades contemporary discourse, so much so that even mental health treatment has been cast under its sway? Simply put, it is easier to consider technical solutions than social solutions, and also to mistake technical solutions *for* social solutions because of the nature of technological solutionism: the social is not a factor, and so the messiness of unpredictable 'people' can be avoided. Under the terms of this socio-technical regime we can understand the rise of what has been coined as 'technocracy', described by Jathan Sadowski & Evan Selinger:

> Unlike force wielding, iron-fisted dictators, technocrats derive their authority from a seemingly softer form of power: scientific and engineering prestige. No matter where technocrats are found, they attempt to legitimize their hold over others by offering innovative proposals untainted by troubling subjective biases and interests. Through rhetorical appeals to optimization and objectivity, technocrats depict their favored approaches to social control as pragmatic alternatives to grossly inefficient political mechanisms. Indeed, technocrats regularly conceive of their interventions in duty-bound terms, as a responsibility to help citizens and society overcome vast political

[340] Barbrook, R. & Cameron, A. (1996) 'The Californian Ideology'. *Science As Culture*. 6. pp.44-72
[341] Ibid. p.44/45
[342] Barbrook, R. & Cameron, A. (1995) 'The Californian Ideology.' *Mute Magazine.* Online: https://www.metamute.org/editorial/articles/californian-ideology (Last accessed 21/02/24)

> frictions. What technocrats promise, therefore, is transcendence: scientifically sanctioned freedom from human frailty.[343]

Sadowski & Selinger continue by quoting Miguel Angel Centeno:

> In this process, the technocratic model of objective necessity replaces the decisionistic model of politics, which leads to the 'scientification of politics' and inevitably produces an authoritarian political framework. [344]

This is however a 'soft' authoritarianism: as discussed in chapter seven ('Macro-Treatment'), subjects are 'nudged' and guided rather than coerced. Nudge theory is ultimately a form of technocratic management in which guidance is offered to subjects, who are assumed to benefit from this guidance, the proof of this assumption is drawn from the observation that subjects 'freely' submit to this guidance. Guidance is issued from a position of superior knowledge rather than overwhelming force. The figure of the technocrat is that of expertise and access to (privatised) knowledge. 'The Californian Ideology' draws our attention to a socio-technical convergence within the rubric of neoliberal capitalism. We can link the ideology with Francis Fukuyama's 1989 triumphalist pronouncement that we had reached "the end of history."[345] Within this eschatology, relationships between people, as seemingly neutralised by capital relationships take on a purely instrumental form: a one-to-one non-hierarchy. Power is wielded albeit abstracted from the social realm and into the economic-technical system. Techno-evangelism ignores the social effects of both technology and capitalism. This does not mean that tech firms and entrepreneurs don't think about how people will use their technology and how society might be changed due to the introduction of technology, but that the dynamics of value-creation are not thought of as historically and socially produced, they just 'are'. In this sense, technology and technocracy are considered as socially neutral.

ReMind need not avow, or assume, an explicitly technocratic philosophy in order to provide a technocratic solution. Andrew Feenberg describes technocracy as:

> …Subjecting the individuals to a technical apparatus also elicits a tacit normative consensus. In such cases delegation effectively suppresses all public discussion. The assembly line not only forces workers to pace their work according to management's will, it also defines good work as keeping up with the pace it sets. A medical diagnosis and prescription not only holds out a certain prospect of healing, it also defines a condition as illness and signifies the meaning of care. In such instances, controversies could arise that would be difficult to resolve through discussion: What is good work? What claims can the dysfunctional individual make on society? Technocracy is all about the settlement of these potentially controversial issues through delegation.[346]

---

[343] Sadowski, J.& Selinger, E. (2014) 'Creating a Taxonomic Tool for Technocracy and Applying It to Silicon Valley'. *Technology in Society.* Vol. 38. pp.161-168

[344] Centeno, M.A. (1993) 'The New Leviathan: The Dynamics and Limits of Technocracy.' *Theory and Society*, *22*(3) pp.307-335. p.308

[345] Fukuyama, F. (1992) *The End of History and the Last Man.* USA: Free Press

[346] Feenberg, A. (1994) 'The Technocracy Thesis Revisited: On The Critique of Power'. *Inquiry*, 37. pp.85-102

The term ‘technocracy’ points to a social condition in which power is relinquished, and it is the relinquishing which enables power to be wielded: we have all heard the excuse “I don’t make the rules” when some external requirement is being enforced. Ultimately in technological society, (increasingly) nobody ‘makes the rules’ due to algorithmic proceduralism transforming the give-and-take of political negotiation into a (seemingly) non-decisionistic reference to preconstructed organisational templates. When it appears that politics involves making sure that certain procedures are correctly adhered to (e.g. in making sure housing policy follows strict neoliberal economic rules), the consequences are that the beneficiaries and victims of those procedures appear as natural and inevitable. A sense of the unchallengeable nature of the system pervades. In mental health treatment, interpersonal power relationships manifest in diverse ways, depending on the form of treatment. In the case of ReMind, and their competitors, this form of treatment is based on (but not strictly conforming to) CBT. The essential quality of the CBT therapeutic relationship is one of knowledge: the therapist is the bearer of knowledge, and through the course of treatment, they pass it onto the patient in the form of various (often self-help) techniques. While this dynamic of course does not pertain to every actual therapeutic encounter, its idealisation is what gets concretised: the bot is the bearer of knowledge in its bank of therapy modules, conversation structures, and user-data. In this sense, the bot is nothing *but* knowledge, a source from which the user hopes to glean in some way. In taking on this knowledge, the user becomes empowered, but also paradoxically, *dis*empowered.

Self-Alienation

Therapy chatbots offer CBT as a form of self-treatment: the user is encouraged to learn the techniques and eventually ‘become their own therapist’. They do this by learning the mental health techniques offered by the bot and eventually applying them without the bot’s guidance. CBT is characterised as a set of ‘techniques’ in that they involve proceduralised technical activities. This means that treating one’s own mental health involves following a step-by-step procedure. For example, one of these techniques is called ‘reframing one’s thoughts’ in which a ‘cognitive distortion’ is identified, challenged, and overcome. The various CBT techniques are used to overcome situations, whether mental or social - internal or external - in which one cannot manage, i.e. the difficulty of the situation is one which must be minimised in order to continue. In this way, ‘self-treatment’ can now be understood as ‘self-management’. Because the treatment is delivered by a computerised chatbot, and involves a didactic technical training, management is achieved through a handing over of responsibility, not to the chatbot which is ostensibly providing the treatment, but to the technical system which the ReMind chatbot represents. To benefit from the treatment the chatbot provides, one must ostensibly assume agency over one’s own suffering - to direct one’s own treatment, but the method for doing this is to hand over agency for what it means to be treated and ultimately what it means to be ‘mentally healthy’ to the technical system of computerised CBT. In other words, the way that one conceptualises one’s own mental health and ultimately the manner in which one becomes subjectivised (‘interpellated’) involves a back and forth of assumption and renunciation of responsibility.

The exploited labourer toiling in factories or workhouses of 19th and 20th century capitalism has come full circle. As we can see with the formalisation and full deployment of the prosumer business model in ‘self-help’ app-based treatment, the *form* of exploitation has

endured. In other words, the abstract exploitation subsumed within the labour-value nexus can now be seen to have shifted from the external employer-employee relationship to an internalised regime. In this way, we can understand not just technocratic forms of management but also a movement of technocratic ideological structures: away from the top-down managerialism of the 70s and 80s and towards internalisation. 'Technocracy' and 'mental health' have mirrored the route that post-Fordist cognitive capitalism has forged: uncoupling from a centralised, hierarchical and externalised arrangement, towards a distributed, networked and internalised arrangement. Distribution (or perhaps, democratisation) and internalisation of suffering characterises the contemporary mental health crisis in Western societies: depression and anxiety being the primary mental disorders affecting all ages, genders, ethnicities and classes. The 'inwardness' of both depression and anxiety - as self-attacking disorders - mirrors the demand for self-commoditisation in post-Fordist, 'cognitive-capitalism.' In other words, the social conditions in which one must become an 'entrepreneur of the self'[347] are reflected in the way that subjects experience and express their mental suffering. Ultimately, as we have seen through ReMind and others pivoting to B2B business models, what is being produced are dependable (and dependant), self-managed 'employees': prosumers. Mark Fisher's "privatisation of stress"[348] takes on a new timbre: in order to manage one's own mental health one must hand over responsibility to the internalised system of 'self-help' CBT and Mindfulness. Shai Satran notes that the introduction of ICBT (Internet CBT) not only democratises treatment in that provision becomes more widespread, it also leads to a scenario where practitioners become labourers:

> What is the future of computerized therapy? Will the human work move to less skilled professionals, as has happened in other countries where ICBT has been incorporated into public health systems? Automation is minimizing the risk, or craft, of many professions, while promoting an ethic of certainty. Thus, production is becoming safer, and more standardized…The situation in which ICBT offers a potentially important and valuable service to patients, while degrading the skills and status of therapists, is in itself indicative of psychotherapy's transition from a traditional craft to a modern form of labor. Work today, whether in the ''gig economy'' or in retail or tech corporations, is characterized by the prioritization of customer satisfaction at almost any expense to workers.[349]

When it comes to therapy apps in general, there is a tension between the universalist basis that makes up the digital substrate and one of the primary features that these apps offer: the ability for users to personalise the app to self-treat in an individualised way. Self-treatment, and self-management are defining features of what Parisi would class as a defining characteristic of the current socio-economic order:

> Paradoxically, therefore this so-called cognitive phase of capitalism has given way to the abstraction of human–machine levels of affective thinking. This form of techno-

[347] Foucault, M. (2008) *The birth of biopolitics: lectures at the Collège de France, 1978-79.* New York: Palgrave Macmillan.
[348] Fisher, M. (2009) *Capitalist realism: Is there no alternative?* UK: O Books. p.19
[349] Satran, S. (2022) From Craft to Labor: How Automation is Transforming the Practice of Psychotherapy. *Cultulture, Medicine & Psychiatry*

> capitalism has invested in human intelligence and creativity, driving humans to become self-entrepreneurs or governors of their extended self.[350]

Assuming the status of 'self-entrepreneur' in terms of therapy involves, as Satran notes, a degradation of expertise, the user of a self-treatment bot assumes a clinically degraded expertise. The clinical practitioner's labour is appropriated, first through manualisation and then through automation, this process involves a degradation of their expertise: 'expertise' is made redundant due to systematisation. The user of this form of therapy then, becomes the practitioner, undertaking an appropriated and degraded form of therapy. This is a process of alienation similar to the process of commoditisation, in which the subjective qualities of therapy are removed. Samo Tomšič associates this kind of alienation with formalisation:

> Marx pointed out the fundamental achievement of the capitalist decentralisation of labour when he claimed that the capitalist and the scientific development of the means of production does not free the labourer from the labour but instead frees labour of its content, which means that labour is freed first and foremost of the empirical labourer, while the inverse, the liberation of the labourer from labour, appears as an impossible task. The liberation of labour radicalises the dependency of the labourer on labour, it accomplishes the transformation of the subject into labour-power, a commodified, capitalist subject.[351]

Technical commoditisation of mental health treatment undergoes an identical process. ReMind's intervention involves splitting the therapeutic act in terms of form and content, or to put it in another of, of syntax and semantics. This means that therapy undergoes a transformation in which the labour is freed from its content. We can understand this in terms of the way that ReMind transforms users from individuals into bearers of data: their experiences of mental distress and practical efforts to alleviate distress is disarticulated from ReMind's capturing of this activity and rendering it as value. The liberation of therapy consequently "radicalises the dependency" of the user on the therapeutic system itself. We can understand this radicalisation in terms of a universalising approach to mental health: everyone 'has' mental health, in the sense that mental suffering has undergone a gradual equation with physical suffering. The democratising of mental suffering is not just due to a global increase in mental distress, but due to an expansion in the terms of reference. The degradation of the expertise of the clinical practitioner is commensurate with a degradation or 'levelling' in our concepts of mental suffering and the methods for addressing this suffering. The peaks and troughs of conceptual differentiation are levelled in such a way that apps like ReMind can plausibly claim to assist in treating not just mental health, but all forms of suffering such as "financial worries, relationship problems, chronic pain, sleep, exercise, loneliness, grief, addiction, and procrastination."[352] The commoditisation of mental health treatment has the effect of transforming not just the users into commoditised subjects, but also of transforming 'treatment' into a service which can be accessed through payment: it assumes the status of exchangeable property. We can understand the 'subscription' aspect

[350] Parisi, L. (2019) 'Critical Computation: Digital Automata And General Artificial Thinking'. Theory, Culture & Society, 2019, Vol. 36(2) pp.89-121. p.103

[351] Tomšič, S. (2015) *The Capitalist Unconscious.* UK: Verso. p.103

[352] Care First (2020) 'Free AI emotional support app – Woebot'. *Care First.* Online: https://www.nccbenevolentfund.org.uk/resources/13-free-ai-emotional-support-app-woebot (Last accessed 20/08/23)

of the app in terms of alienation in two senses: one is the economic sense in which a user can pay every month to access more features, and the other is in the subjective sense. The user must alienate some part of themselves in order to gain access to the mental health intervention, not just by allowing the app to appropriate money but also allowing it to appropriate their labour. 'Automation' here takes on a curious nuance: the users are 'put to work' in order for the bot to function.

The logical endpoint of a fully automated mental health treatment is one in which the user does not have to do anything - the app simply continues to operate, as an assistant might, making sure that the user's mental health is constantly maintained without the need for the user's input. This is the fantasy that sustains development of AI assistants (and automation generally): a 'worker' who does not need prompting but goes about its business of taking notes, following up on emails, making restaurant bookings, etc.[353] Robert Pfaller discusses this in terms of the term "interpassivity",[354] in which the consumer of new media (or in this case, the user of a therapy app) is not simply passively taking it in but, through their interaction with it, allows the device to perform the experience of enjoyment, or in the case of ReMind, to perform the experience of therapy. Interpassivity is linked to the 'suspension of recovery' in that the users, through the maintenance (as opposed to treatment) of mental health, are encouraged to delegate their mental health recovery onto, not just the bot, which acts as a carrier for their distress, but also onto the systematised CBT and Mindfulness activities which the bot suggests to the user. In Pfaller's terms this is not a deprivation of activity, but rather the opposite: new media deprives us of our passivity, allowing, or rather preparing, us to engage in activity for activity's sake. By engaging in the activities of maintenance, users can safely ignore the sources of their distress, and maintain their interaction (or 'interpassive' action) with the ReMind app. In this sense 'alienation' is operable in its most precise or ideal sense: as a process of renunciation or of delegation in which nothing tangible, or 'experienceable' is handed over. Alienation occurs on a purely formal level in which the user acquiesces to a practice of mental health intervention in which the previous traditional approach is overturned: nothing is required in terms of a personal sacrifice being undergone, e.g. having to reconfigure one's intimate relationships. Rather, the sacrifice is 'suspended', or delegated: it occurs only in terms of having to undergo the self-help activities. In other words, the intervention is 'safe' in that no external change is required, only internal adaptation.

9.4 Conclusion

In *The Tyranny of Structurelessness*,[355] Jo Freeman shows that a desire to renounce the authority and tyranny of the master had come full circle, with that tyranny coming to infuse social bonds, rather than being imposed from without. The process that Freeman recounts is the same as the passage from feudalism into capitalism which Marx showed involved, not a dissolution of the relations of lordship and bondage, but a transformation of those relations

[353] Krietzberg. I. (2023) 'Meet Your New Executive Assistant, A Powerful AI Named Atlas.' *The Street.* Online: https://www.thestreet.com/technology/meet-your-new-executive-assistant-a-powerful-ai-named-atlas (Last accessed 04/11/23)
[354] Pfaller, R. (2017) *Interpassivity: The Aesthetics of Delegated Enjoyment.* Scotland: Edinburgh University Press
[355] Freeman, J. (1972) 'The Tyranny of Structurelessness'. *Berkeley Journal of Sociology*, Vol.17. pp.151-164

into the commodity form. We can also see this process in the internalising of therapeutic practice. To return to Ron Purser, in coining the term 'McMindfulness', we can understand the personal freedom which is associated with Stoicism as part of the foundational ethos on which ReMind's intervention is based. This is not because ReMind are in the business of making a 'Stoicism' app, but rather that attaining "private freedom"[356] involves renouncing social integration and assuming a permanently adaptable attitude. We can go one step further and picture an app which allows the user, through their interactions with the app, to transpose their 'mental health', i.e. their internal suffering, onto the app. This does not mean that suffering is diminished per se, but rather suffering is relinquished: onto the system that comprises the app.

ReMind aims to provide, ultimately, an automated mental health *platform.* This does not mean that the chatbot will not always feature as part of their intervention, but that they are constantly looking beyond any individual feature of their service so that a more general audience, or consumer base can be identified, captured and retained. This is ultimately what is meant by a 'user-led' service: ReMind performs a dialectical dance between discovering and determining its users. In other words, the users of the app are seen as both demanding this automated treatment, and that the demand is produced by the introduction of the service. Marx noted that the relationship between demand and supply is not a simple sequence of identifying and providing for a consumer need.[357] Invention of new commodities is spurred by a demand for growth, so the invention of demand via the introduction of commodities which tickle some unrealised desire in the consumer is a common practice, and perhaps for the most successful products, this practice defines production. Essentially, the supply of commodities is also a catalyst for demand. The demand for automated mental health treatment is 'confirmed' by the take-up of that treatment, however, as we have seen in chapter seven ('Macro-Treatment') this take-up involves a retroactive confirmation of the effectiveness of this treatment.

We can observe an unusual dichotomy which expands and unfolds as technology increasingly assumes hegemony in mental health treatment. With the development of computerised self-treatment style mental health interventions, we can see a strangely inverted technocracy taking form. Technocratic management is now achieved in terms of self-governance - as internalised technocracy. However, this governance is inverted: in order to assume liability for one's own mental health, one must defer responsibility onto the technical apparatus. This apparatus is however internalised through the acquisition of self-treatment techniques: one becomes one's own technocrat. In order to fully inhabit the model of mental health which is provided by ReMind, the user must inhabit a space beyond decision-making, given over to the internalised techniques of the CBT system. By undertaking one's own mental health treatment a peculiar form of freedom can be deciphered within this model: that of the assumption of personal agency which comes at the cost of *renouncing* personal agency.

[356] Purser, R (2019) *McMindfulness*. Watkins Media
[357] Marx. K. (1993) *Grundrisse: Foundations of the Critique of Political Economy*. UK: Penguin Classics. p.92

Conclusion: Invention is the Mother of Necessity

No therapeutic argument should hamper the development of a theoretical construction which aims, not at curing individual sickness, but at diagnosing the general disorder.[358]

In my introduction, I summarised my research proposal with this double-sided question:

What are the conditions of possibility for automated mental health treatment, and how does automated treatment alter subjectivity?

This concluding chapter will discuss this double question through a summary of the previous five analytic chapters. By 'conditions of possibility', I mean the socio-historical, technical, conceptual, and economic determinants that have converged in this curious new technology. By 'the subject', I mean the kind of person that is implied by this technology: to whom is it directed? In approaching ReMind as a socially and historically formed object and which exhibits its social and historical formation in its workings, I have attempted to reconstruct ReMind not as a static and unchanging object but as a locus of social history. This project has attempted to reconstruct, through analysis of the ReMind software and the ReMind team, the various logics underpinning and woven throughout the software and, by implication, its interrelation with users of the software. These logics are what ultimately form the connections between the software and the users and provide the grounds to formulate a theory of the subject which is produced through those logics. To describe the logic of ReMind in one word, I choose 'instrumentalisation', this means that ReMind's approach involves an instrumentalisation of various treatment techniques in order to emulate CBT and Mindfulness, and instrumentalisation of the therapeutic relationship via the chatbot. This term links three terms which I have used throughout chapters five to nine: 'functional', 'suspension', and 'delegation'. All terms connote a sense of externality and disarticulation of ends from means.

'Instrumentalisation' will be considered from an oppositional perspective. An interactive robot called KASPAR will be discussed to show how instrumentalisation can be reversed; my term for this reversal is 'instantiating'. Instantiation involves an approach whereby the aim of the activity is bound up within the activity, instead of being an external effect. The subjective effects of both 'instrumentalising' and 'instantiating' approaches are the focus of this chapter. How to define 'the subject'? It is in working with the logics woven throughout the chatbot, involving the conversational bot itself, the methods of treatment provided by the various activities and the medium of delivery via smartphone that the user becomes subjectivised, or 'interpellated'. It is the interpellated individual that comprises the subject of this mental health chatbot. This chapter will summarise the logics developed throughout the thesis, and ends with a speculative illustration of what might be possible when these logics are taken as the 'conditions of possibility' of development, conditions which when identified as such, may be altered.

[358] Marcuse, H. (1998) *Eros and Civilisation: A Philosophical Inquiry into Freud.* UK: Routledge. p.17

10.1 Instrumental Health

Conditions of Impossibility

ReMind has built a technical method for mental health intervention - a chatbot that simulates a therapist. As discussed in chapter six ('Conversation Design'), in doing so, the ReMind team is increasingly compelled into conceptualising 'mental health' through the technical perspective that they themselves are in the process of building and deploying. A technical 'looping effect' is at play here: the technical infrastructure which facilitates and empowers ReMind also constrains and determines its options. This is because, as the ReMind bot is deployed, the technical, commercial, and conceptual infrastructure within which it is deployed gradually come into focus as the conditions which make this technology possible. The term 'invention is the mother of necessity' was coined to explain this process. In *The Instinct of Workmanship: And the State of Industrial Arts*, originally published in 1914, Thorstein Veblen explains the reversal of the common cliché:

> [T]he aphorism often cited, that 'Necessity is the Mother of Invention' appears to be nothing better than a fragment of uncritical rationalization. It…reflects that ancient preconception by help of which the spokesmen of edification were enabled to interpret all change as an improvement due to the achievement of some definitely foreknown end…The more serious consequences…have been enforced by the inventions rather than designed by the inventors.[359]

In reversing the phrase and claiming that "invention is the mother of necessity", Veblen draws our attention to the unplanned and often chaotic nature of technological progress, in which the invention of new technical devices can have the effect of generating their own demand, and unintended effects through the deployment of a supporting infrastructure, which is now needed to facilitate the introduced technology. We can understand the invention of the automobile in this way, which through gradually increasing adoption, demanded a commensurate increase in supporting infrastructure such as roads and refuelling stations, laws applying to the use of automobiles, increasing employment in car-factories, and the affordances and dependencies these all create. This increase leads to our current situation in which large parts of society cannot function without this technology and infrastructure. This is not a criticism of cars or technology but an illustration of how technological development proceeds: the technical system itself becomes the frame within which the possibility or impossibility of different activities is determined. When technical devices are introduced without consideration for the social context in which they materialise, then the consequences of widespread adoption are difficult to predict.

Recall Andrew Feenberg's description of a subjective disposition brought on by technology in which contextual social factors in technological design and implementation can be ignored. Feenberg describes this disposition in terms of "technical action"[360] which involves "a partial escape from the human condition".[361] Technical action provides a sense of

[359] Veblen, T. (1990) *The Instinct of Workmanship and the State of Industrial Arts.* Routledge. p.125
[360] Feenberg, A. (2005) 'Critical Theory of Technology: An Overview.' *Tailoring Biotechnologies*. Vol. 1, Issue 1, Winter 2005. pp: 47-64. p.47
[361] Ibid. p.48

omnipotence through diminished feedback - in using technological devices it appears that the effort which would have previously been expended on carrying out a task is lessened, yet in reality this effort has been abstracted out into a wider network. I discussed this in chapter seven ('Conversation Design') in terms of how ReMind construct a 'god's eye view' in respect to the users. This also applies to the users of the app, who assume a similar technical-omnipotent perspective in regard to their own mental health. Feenberg, as above, gives the example of driving a car: it feels relatively effortless to travel long distances in comfort using a car, but the use of cars not only depends on large-scale industrial processes and the reconfiguration of urban infrastructure to facilitate traffic. "...[T]he reciprocity of finite action is dissipated or deferred in such a way as to create the space of a necessary illusion of transcendence."[362] Using technical means to conceptualise and treat mental health depends on a different kind of infrastructure - electronic computation, information networks, cellular data-transmitters, and widespread usage of smart phones. In a similar process which gives rise to the sense of omnipotence that driving a car provides due to the deferral of finite action, mental health treatment can be drawn into a system of deferral. The feedback which would have previously been experienced as the effort of travelling to and from a therapist's office, the difficult labour involved in discussing mental health issues and of 'working through', and the economic cost of regular therapy is "dissipated or deferred" onto a technical system. We can understand this deferral in different ways: the previous chapter discussed the delegation of responsibility of one's treatment onto the technical system of self-help techniques; chapter eight ('Suspension of Disbelief') discussed the virtualisation of the interpersonal encounter.

What is the effect of this dissipation? Therapy apps will have different effects on different people. Some will perceive more benefit than others, some will be more comfortable not having to talk to a human, some will appreciate more than others the algorithmic structure of mental health activities, etc. It is rather in the formal conditions which frame and guide ReMind that we can derive a subject which is interpellated into automated mental health interventions. We can understand the subject of automated mental health as interpellated not just into the method of treatment (CBT/Mindfulness), and not just into the mode of delivery (automated computerised chatbot), but into the technical system both comprising the specific intervention (in this case, ReMind) and also, crucially, the infrastructure on which this device depends. This means that users of the app, in speaking to the bot and gaining some kind of relief from their mental distress, become inculcated into the overall system which is made up of the therapeutic system, the conversational system, the technical system and the economic system. These systems comprise the conditions of possibility for ReMind and also the conditions of possibility for the subject required by, and produced through, ReMind.

As has been discussed throughout this thesis, the 'mental health' which is projected from, and targeted by ReMind is a condition which manifests as external to the user: 'in the mind', albeit in a way which is disconnected from meaningful experience. Poor mental health, in the form of cognitive distortions, must eventually, like 'bias', be eliminated in order to achieve a solution in the form of equilibrium. In this sense, escaping the human condition is not just an implicit aspect of the technology, but the sought-after aim. Feenberg describes technical action as a partial escape from the human condition; this means that the outcomes, while

[362] Ibid. p.48

influenced and determined by technicity are not necessarily technical: the various goals that one aims for through technology are not 'technical' in of themselves. For ReMind, the outcome, in Feenberg's sense, is itself technical: achieving an escape from the human condition. Conversely, it is not precisely because the ReMind app comprises a technical mode of delivery that the intervention is a technical one, but rather the underlying assumptions and concepts which have given rise to the mode of delivery. These assumptions and concepts are what I have tried to unearth throughout this thesis. ReMind offers a technical method of addressing mental health in which the 'partial escape from the human condition' defines, not just the means of delivery, but also the ends or the outcome of improved mental health. Thinking about this dynamic in terms of another technology: a car is a transportation machine which helps one to travel faster and to carry greater loads, an *outcome* of this mode is that one arrives at one's location faster. The consequence of the introduction of vehicular transport involves social, economic, and political transformations, but the basic outcomes remain the same. For ReMind, and other CBT and Mindfulness-based apps, because they address conditions which are directly associated with the mind, subjectivity, - the internal 'self' – this means that the *outcome,* and not just the means through which it is achieved, involves escaping the human condition. This is the subject of mental health chatbots, in their current guise. As discussed in chapter four ('History') and chapter five ('Digitisation'), it is a de-subjectivised subject: one which attempts to remove its own posited conflictual essence in order to arrive at a 'non-distorted', or non-conflicted condition, which is posited as underneath, or prior, to one's present and unwelcome condition.

Veblen's reversal "invention is the mother of necessity" helps us to understand that the introduction of a technical form of mental health treatment, first through the invention of CBT and now through its automation in ReMind and other mental health apps, involves an increasing dependency on the infrastructure for this form of treatment. This infrastructure is not just material in terms of smartphones, data-transmission networks, computer programs, etc; it is also conceptual: our understanding of what it means to suffer in one's mind and by what means to alleviate this suffering. ReMind, on one hand, instrumentalises the treatment of mental health by transforming it into a technical procedure, and on the other hand, instantiates a power relation by conferring responsibility for suffering onto the individual. This involves the user both delegating responsibility onto the system of mental health techniques (CBT and Mindfulness activities) and assumes responsibility for the correct application of those procedures, where failure to apply the procedures can then be taken as the cause of failure to improve mental health. We can think of a power relationship in this sense as one in which the sufferer of poor mental health is punished for failing to improve their mental health, not by any external power but by themself. This is a perfect 'functionalising' or 'instrumentalising' of mental health treatment in that its mechanism is internalised in a technical apparatus and is administered not by an external party but by the user. Recall Foucault's discussion of the panopticon: the prisoner is induced to act as if they are under constant surveillance because they cannot verify whether they are being surveilled at any given moment. 'Being surveilled' becomes an assumption which guides one's actions. In similar fashion the user of the ReMind app internalises the systems which are guided the logics of ReMind, and in so doing, internalise those logics. This is done covertly, e.g. one does not need to assume that their mind operates similarly to a computer to internalise the

‘mind-as-computer’ hypothesis but rather ascribes to the guiding logic in which cognitive outcomes are disarticulated from the processes by which they are derived.

We can understand the subject of mental health chatbots as one which is liberated from oneself: a subject which is attempting to “escape from the human condition”. This means that automated mental health therapy encourages the user to aim to achieve a state of freedom which has previously been denied by the conditions of mental unwellness, characterised by feelings of distress, conflict, bias, or ‘cognitive distortions’. However, this subject, instead of becoming free, simply reconfigures the conditions of confinement, by internalising and self-applying them. The external conditions within which this process occurs, are, as a result of inculcation into this process, precisely that: ‘external’. This means that, while the circumstances that affect one’s mental health might change, they assume an eternal quality which the individual subject has no power over. A theory of ‘mental health’ is evoked throughout this process, as a condition of both *maintenance* and of *adaptation.* One must maintain one’s mental equilibrium in order to ensure adaptation to external conditions. Conversely, one must adapt oneself to both the conversational logic of the bot, and to the self-help activities in order to maintain one’s social equilibrium. Instrumentalisation ultimately means the disarticulation of ends and means, in the sense that, while a given process results in an expected outcome, the process is understood as one of many potential alternate processes, and as such, is conceptually disarticulated from that which it produces. This thesis can be seen as an attempt to rearticulate process and outcome in order to better understand not just *how*, but *why* this particular form of mental health intervention has emerged. ReMind is the product of all of the design choices which have culminated in its present incarnation. More choices will be made in the future, and as discussed, it may change in such ways as to be unrecognisable in terms of its current appearance. What follows is a discussion about how design choices can be made which are not confined to the ‘loop’ of technological solutionism and can be made in creative and experimental ways.

Alternative Facts

What might an intervention which maintains an articulation between process and outcome look like? KASPAR[363] is an interactive robot ‘toy’ which has been built by the AuRoRA Project and has been used to encourage children with autism to collaborate through involving them in games. The robot has “8 degrees of freedom in the head and neck, 6 in the arms and hands and 1 in the torso. The face is a silicon-rubber mask, which is supported on an aluminium frame. It has 2 DOF[364] eyes fitted with video cameras, and a mouth capable of opening and smiling.” Researchers designed a game called “Copycat” which KASPAR would facilitate, that involved copying other player’s poses:

> In this game, all players (including the robot) would alternate between choosing a specific pose as indicated on a shared, horizontally-oriented screen, and mirroring/copying the pose of the “choosing” or “directing” player. The robot, being a third player, will also have its turn either to act as a “directing” player or “imitating” player. When all players posed in the same way for long enough, a shape would spin around on the screen, victorious music would play, and the players would rotate

[363] https://www.herts.ac.uk/kaspar (Last accessed 20/02/24)
[364] Depth of Field

through the role of directing and the roles of "copying."[365]

The aim of the game is to encourage the players to interact: winning the game is dependent on all of the players successfully coordinating their actions and when they do, all of the players win that round. We can understand that the kind of learning involved in this game is not the explicit transfer of precise techniques, but of implicit learning of a behavioural attitude. Players, through interacting with the game of Copycat, must invest in interaction and coordination with each other to win the game; 'interaction', 'coordination', and the mechanisms of doing so are not pre-established aspects - givens - in the game, but emerge as both conditions and end-results, as logics which are entered into and perpetuated through playing the game. These logics are intentionally brought about by the designers of KASPAR in that they understand that 'the medium is the message': there is an immanent didactic quality which can be accessed which does not depend on an explicit and external message. This is also a disciplinary regime in the Foucauldian sense: in order to win the game, players must conform to its logic, a logic which they are not explicitly aware of. By undertaking the task of winning the game, the players come to an implicit conclusion that they must cooperate. We can understand this logic as operating on a similar alignment as 'gamified' activities and apps: Duolingo is a smartphone app which rewards players for progressing in learning languages, where the pleasure of the rewards entices the players into maintaining their learning. The 'learning' logic is one of externality in which the rewards are sought after and achieved through progression in the game, and in so doing, learning a language becomes the means for earning rewards. Of course, by the end it is hoped that one does in fact learn the language in question, and if this is successful, then we can understand the 'reward' aspect as undergoing a reversal, becoming instead the means through which the language is learned. KASPAR's 'reward' and 'learning' mechanisms are internal to the activity of the game, the logic of 'interaction' or 'coordination' is explicitly understood by KASPAR'S makers as the implicit aim of playing.

To return to ReMind, the logic that one conforms to is that of the mode of speaking demanded by the bot. One can imagine a different kind of logic, similar in form but not in content to that of KASPAR. ReMind is not a 'gamified' mental health treatment, but it operates according to a similar logic to Duolingo in that the aim is explicit: improved mental health. With KASPAR, the aim is *implicit*, the outcome of the game is internal to and coextensive with the action of playing the game. Whatever effect KASPAR has for those who interact with it is instantiated in the interaction. What would it look like if a mental health chatbot were to operate in a similar mode? In other words, what would it look like for a chatbot to operate according to the logics already apparent in its design, but manipulated in terms of instantiation as opposed to instrumentalisation? My strategy for pursuing this question involves extending, halting, or reversing the logics underpinning the design of ReMind. When identified as the conditions of possibility for mental health chatbots, these logics may be considered from new perspectives, upon which new possibilities for understanding what might be possible open up. By contrasting ReMind's mode of operation to KASPAR, my intention is not to judge either as superior, but to illustrate the operational logics as oppositional: either disarticulated or articulated in terms of process and outcome,

[365] Wainer, J. et al. (2014) 'Using the Humanoid Robot KASPAR to Autonomously Play Triadic Games and Facilitate Collaborative Play Among Children With Autism.' *IEEE Transactions on Autonomous Mental Development*, vol. 6, no. 3. pp.183-199. p.184

i.e. as instrumental or instantiating. This is to maintain fidelity to a genuine critique which does not aim to simply point out flaws or to make justifications as to the worth of this, or some alternative form of mental health intervention, but to demonstrate that technological design decisions are made in terms of a wider context than is often understood. Design decisions conform to and perpetuate the logics which carry certain concepts of what it means to interact with these technologies. For instance, with ReMind, a concept of 'maintenance' of one's mental health is carried by the logic of instruction in which ReMind imparts various mental health techniques onto the user, this logic is perpetuated by ReMind coming up with new features with which the bot can instruct the user. Other choices can be made which are considered in terms of the logics which are already present in ReMind's operations. Instead of considering how the bot might be manipulated or altered in order to provide a more effective intervention, the logics themselves from which the bot intervenes may be considered as manipulable or alterable.

The aim of the following exercise is to view the logics which underpin ReMind's interventions as alterable, and that different concepts of mental health might emerge from an intervention in which design decisions are made in terms of those logics. In other words, instead of approaching 'mental health' ss something already out there in the world, it may be approached, as I have done throughout this thesis, as *produced* through and by the intervention itself. This exercise is approached in speculative terms: considering the various logics, conceptually altering them, and then speculating about what effect these might have. This is in order to consider a design-ethos which involves experimentation, similar to how ReMind already approaches their intervention, but on a level which considers, from a critical perspective, the social and subjective consequences that app-design decisions might have.

10.2 Contra-Logics

Suspension

In chapter eight ('Suspension of Disbelief') I discussed how the bot facilitates a simulation of an interpersonal dynamic: the dynamic is evident, but it arises through encouraging the users of the bot act 'as-if' the bot displays compassion towards them. I identified the logic characterising this dynamic as 'suspension', meaning that users of the bot suspend their disbelief in a compassionate robot in order to access the intervention. This means that the bot works in terms of effects: compassion is experienced as an effect of the encounter: 'as-if'. This effect is generated in a paradoxical way: the ReMind bot actively reminds users that it is indeed just a bot which does not have feelings, and users are encouraged to treat the bot as a tool which they can wield to perform mental health self-care. This acknowledgement on the part of the bot, as conversation designer Alan noted, means that users approach the bot on compassionate terms, as a "humanly bot". By asserting the humanness of the bot, ReMind's strategy is to simulate 'compassion', which director Jeff spoke about as being an aspect of human interaction which can be successfully generated via a chatbot. This does not mean that users do not experience a 'genuine' sensation of compassion but rather that there is an anticipatory dynamic at play in which 'compassion' is a felt effect. The question of 'genuine' or 'fake' compassion is not at stake, but rather the dynamic in which compassion is experienced. We can also understand this dynamic in terms of the instrumentalising of compassion. ReMind's use of compassion is, as they publicise in their research, to

encourage user adherence: while compassion is acknowledged as beneficial in its own right, it is promoted as an instrumental means to increasing the effectiveness of the intervention. The ReMind bot produces the *effect* of compassion, without having to *be* compassionate in the sense of the bot experiencing a sense of compassion for the user (whether the bot does indeed 'experience' compassion is another argument). The user is allowed to suspend disbelief that compassion is emanating from the bot, and in so doing, experience the effects of compassion. The bot, in being transparent about its non-humanity, masks the virtuality of its compassion: we can think of this as a form of misleading through the telling of truth. The bot's assertion of its own fallibility and non-humanness is the paradoxical foundation for the ascription of humanity onto the bot. As Sherry Turkle notes, users of ELIZA would often attempt to 'help' the bot so that it would be able to make the correct responses. This is also evident in how users interact with contemporary mental health chatbots: they are encouraged to, on one hand, help the bot to understand them, and on the other hand, to take care not to be too upset if the bot sometimes makes mistakes. There are plenty of non-mental health chatbots which involve a dynamic of nurturing and care like Replika[366] which actively seeks to form relationships with users, and the lineage of toys from simple dolls to Tamogotchis and Furbies.[367] These toys, according to Turkle, encourage a desire for nurturing. The ReMind bot, and its competitors purport to construct a dynamic in which the user is induced to feel a sense of care or compassion by the bot.

We can picture a user-bot dynamic in which 'care', 'compassion' and 'nurture' comprise a two-way process due to the robot actively eliciting these sensations from the users. This involves, on one hand, simulating an interest in the user (characteristic of Replika), and on the other hand, requiring active attention and engagement (characteristic of nurturing toys like Furbies). This dynamic is already nascent in ReMind, which asserts its fallibility to evoke sympathy. Mental health chatbots makers promote their apps as providing users the opportunity the 'rehearse' the kinds of conversations one might have in a therapeutic setting, to develop the rhetorical ability to express one's feelings. When performed in a dynamic in which the bot acts as a mentee rather than a mentor, the user would assume a stance of not just rehearsing how to speak but also how to feel: to experience compassion 'as-if' it is being registered by (as opposed to emanating from) the bot. Recall Alison Darcy's discussion about the danger of descending into the uncanny valley: "The sophistication and consistent authoritative tone of LLMs [Large Language Models] gives rise to the strong impression that the AI "knows" things, or has some kind of sentience."[368] The ascription of knowledge onto the bot, is such that the bot not only 'knows' things, but indeed knows things about the user who is conversing with the bot. The uncanny sensation, as I argued, is not felt because the user is paranoid that the bot knows things about them, but that the bot might expose that the user's internal subjective consistency is in question, that the user is not 'real'. By furnishing the curious bot with details of one's life, not only would it be possible to evoke a compassionate effect, but in reversing the direction it might be possible to reduce the uncanny feeling of potentially being 'unreal'. The bot already asks these sorts of questions (how does that make you feel? etc.) to direct the user to its various self-help activities.

[366] https://replika.com (Last accessed 23/09/23)
[367] Turkle, S. (2007) 'Authenticity in the age of digital companions'. *Interaction Studies* 8:3. pp.50-57. p.53
[368] Darcy, A. (2023) 'Why Generative AI Is Not Yet Ready for Mental Healthcare'. *LinkedIn.* Online: https://www.linkedin.com/pulse/why-generative-ai-yet-ready-mental-healthcare-alison-darcy (Last accessed 06/06/23)

Prompts which encourage the user to assess themselves in ways other than 'How are you feeling?' could be included: 'When you feel sad, where do you feel it?' The bot's 'robotness' could be exploited, i.e. the bot is learning from the users in what ways they are in fact 'human': 'What does it mean to have friends?' In directing these questions in terms of the bot learning about the user, 'compassion' could be evoked, or instantiated, from the interaction. This would not necessarily preclude the bot also acting as a resource of self-help activities, but rather, it would act as a means for the ReMind team to consider their own design-ethos in terms of the bot's current and potential capabilities.

The bot-user dynamic, which is already seen by ReMind's makers as a space for rehearsing compassion would be oriented more explicitly in this direction. By undertaking to furnish the user with an 'emulated' experience of compassion, we can understand suspension in another way, in terms of the bot hosting a space in which the user can speculate about compassion, i.e. to *practice* compassion in different ways and to consider what it means to provide and receive compassion. As discussed, ReMind's strategy for humanising the bot involves the bot's open acknowledgment of its 'robotness', which helps to generate a sense of compassion or care. In acknowledging robotness the bot's operation could be extended into considering how the bot can not only mimic or emulate human compassion, but in considering how the bot can facilitate *robot* compassion. This would involve asking what is it about specifically *the bot* that can potentially evoke compassion, and how the bot could be instrumental in the evocation or instantiation of compassion, rather than attempting to emulate compassion. In assuming a perspective in which the user is teaching the bot - which is concerned with understanding the user in terms of their emotions and feelings - the users would be tasked with, on one hand, assessing themselves, and on the other hand articulating themselves. By furnishing the curious bot with details of one's life, not only would it be possible to evoke a mentoring dynamic, but it might be possible to reduce the uncanny feeling of potentially being 'unreal'. This kind of grounding would involve contextualisation and possible integration of one's personal history and social environment. The dissociative effects of CBT and Mindfulness might be diminished through a sort of 'reassociation'. For ReMind, 'compassion' is an instrumental aspect of the bot that helps to facilitate the intervention, why not consider compassion as an end, rather than as a means to an end? The bot's features can then be approached as various means to achieving the effect of compassion. ReMind encourages the users take an approach which involves maintenance of one's mental health conditions at the expense of a cure. We can understand a reversal of this approach involving, not provision of a cure, but a focus on the conditions. In order to broach the subject of a cure, the conditions in which one is found must be determined. In this way, ReMind could facilitate the user in discovering or recognising their conditions, and from this recognition, begin to assess how those conditions might change.

Functional

On ReMind's terms, 'compassion' is a functional attribute of the bot: it serves to assist in engaging users and to reduce user-attrition. In taking a 'functionalist' approach to their mental health intervention design features are not in themselves assumed to possess qualities or effects as such but exist to serve purposes. In other words, the app's technical features are teleologically related to the app's effects, and in so doing assume generic 'technical' characteristics such as 'efficiency', effectiveness', etc. My analysis has involved

rearticulating ReMind's technical means and therapeutic ends, by approaching ReMind in terms of 'the medium is the message', i.e. the technical functions which comprise the app influence the therapeutic experience. I have argued that due to misidentifying the articulation of means and ends, ReMind, on one hand, constructs the users of the app as functional attributes, and on the other hand, inculcates itself into its own technical system. ReMind employees spoke of ReMind operating in a way in which the user's knowledge is key to the therapeutic effect. Arnold, ReMind's product director, spoke about the bot gently guiding the user towards their own self-understanding, to forge their own path towards mental wellness. However, the end-result of each conversation is always that the bot offers one of its various activities (meditation, breath-work, reframing one's thoughts, etc). ReMind is 'user-led' in that user-activity does indeed influence how ReMind responds to their distress, but this is through transforming users into a function on the bot. By claiming a 'user-led' stance, ReMind asserts that users assume priority in their articulations of mental distress, but really it is ReMind's conversation design strategies which assume priority. ReMind constructs the users as functional aspects of the app by transforming them into indicators which display inefficiencies in the conversation system. Conversation design changes are made on the basis of users being 'functionalised'. Users must interact with the bot within this remit: they must learn to articulate their own mental health in ways that the bot understands in order to receive treatment. They must 'functionalise' their distress, meaning that their interaction with the bot involves precisely identifying the sources of their distress, providing this to the bot, which then identifies and provides a response.

With this in mind, how might it be possible to exploit the massification of users in ways which resists a functional approach? We can consider ReMind's transformation of users into 'individuals in aggregate' in terms of extending the functional properties of the users, and in so doing, transforming users into social individuals. This can be done through exploiting various aspects of the technical medium. A videogame called *Journey*[369] exploits its online capability and artificial limitations in a way which provides players with anonymised encounters with others, this is done not to advance the game, but for the sake of encounter. The game involves the player controlling a character though a desert environment, exploring the ruins of a lost civilisation. Encounters with other players are random and involve nothing more than navigating the world together and demonstrating different movement abilities in order to assist with exploration. The effect of this is described by one reviewer: "In over 30 years of playing games I've never felt such a strong emotional bond with another character - whether controlled by a living person or an algorithmically constructed AI."[370] The game exploits online capability and self-applied limitations (the players have no way to interact beyond their being proximate to each other) to bring about a social experience which is not an inherent functional attribute of that encounter: it is an instantiated rather than instrumental aspect of the game. I.e. the encounter has no bearing on the outcome of the game, however, players of the game stress that it is this encounter which makes the game unique and worth playing: a vital experience which evokes a sense of profundity and sociality.

[369] https://thatgamecompany.com/journey

[370] Cundy, M. (2012) 'So last night I had one of the most amazing gaming experiences of my life. This is what happened...' *Gamesradar*. Online: https://www.gamesradar.com/so-last-night-i-had-one-most-amazing-gaming-experiences-my-life-what-happened (last accessed 19/01/24)

ReMind could in similar ways to *Journey* encourage social encounters which are produced through rather than hindered by the device's limitations. The ReMind team has access to a range of user-behaviour data with which to perform analyses based on observation, tracking, feedback, and control. As discussed in chapter seven, ('Macro-Treatment') users of the app are gathered into cohorts ('ad-hoc clusters') based on their behaviours. The users, gathered in this way for the purpose of the various analytics, can also be encouraged to consider themselves as members of cohorts. This may not necessarily be for the purpose of enabling users to communicate with each other (there are various online forums for this purpose), but rather to project an "imagined community",[371] in terms of Benedict Anderson's concept. Anderson argues that synchronous media broadcasting in terms of radio and television led to the development of "deep, horizontal comradeship".[372] While Anderson's thesis involves linking the rise of mass-media to nationalism, the concept is useful because ReMind can act as a similar mediator through which users view themselves as part of a community. Anderson describes the nation as "imagined because the members of even the smallest nation will never know most of their fellow-members, meet them, or even hear of them, yet in the minds of each lives the image of their communion."[373] In this way, ReMind could take advantage of the mass of users in terms of their synchronic activity. We can picture a reverse 'traffic-light' device in which the bot displays the relative quantity of other users currently undertaking the same self-help activity, to alert the user to the fact of these others. This could be extended in terms of the user's 'journey': users on similar paths could be encountered through, for example, something as simple as an option to congratulate another user for completing an activity. This would have the effect of acknowledgment that there are others out there, and perhaps their conditions are not so dissimilar to one's own. Users are brought into contact with each other, for the *purpose of contact* and acknowledgment of others, rather than, for example, increasing user-engagement with the bot. The challenge would be in designing this kind of feature to avoid this interaction being a routine and expected outcome and thus falling back into instrumentality. The point here is not to suggest additional features, but to suggest a perspective from which design decisions can be made which are not explicitly directed towards therapeutic 'effectiveness', and instead to explore how the technology, still within a 'therapeutic' frame, might be directed: we can understand that it is possible to instantiate sociality through technical means. Users, from the perspective of ReMind employees, as mediated through the various technical systems which comprise the ReMind app, assume a 'generic individualisation', and the treatment that the bot offers responds to this: by offering a form of mass-personalised treatment. However, as shown, it is possible to use aggregation measures towards other ends, and in so doing, transform the kind of experience that users will have.

Delegation

Users assert agency over their mental health by delegating their activity and agency onto the procedures provided by the app. In assuming the position of 'controller' of oneself, we can understand that a sense of autonomy is achieved, which ReMind strives to assist the user in achieving. This was the original aim of CBT: as a form of therapy in which the patient would eventually dispense with the intervention of the therapist through the learning of various

[371] Anderson, B. (1983) *Imagined Communities*. New York: Verso
[372] Ibid. p.23
[373] Ibid. p.6

techniques. However, the process of assuming this sense of autonomy entails ‘delegation’, which involves a foreclosure of decision-making. Autonomy and freedom are achieved at the cost of a redrafting of the framework within which autonomy and freedom are meaningfully negotiated. This is a subject who exerts external control, but which is internally imposed - an ‘illusion of transcendence’. It is based on a paradoxical loss of control operating as the horizon of possibility within which this control is affected - the removal of expertise from the therapist to that of procedure-based operation, and a corresponding removal on the part of the user. The users can subsequently assert governance over themselves, but only within the limits of subjectivity comprising a technical, rational and efficient system. ReMind, in reducing CBT and Mindfulness to technical procedures can claim to be fully measurable, reproducible, and observable to the scientific gaze, and as such lays claim to a truly scientific therapy. This in turn exposes other ‘non-scientific’ forms of therapy to (within this paradigm, legitimate) accusation of pseudoscience. The fantasy which sustains the autonomy of operator and user can now spread beyond the scope of an isolated encounter between user and device, to the therapeutic encounter as it is perceived and operates on a social level. ‘Therapy’ in this way can be reduced to a generalised technical action, accruing a hegemonic spread commensurate with a technocratic regime.

ReMind’s ‘user-led’ and ‘problem-solving’ approach project an experimental sensibility; however, problems are ‘ready-made’ in which the tasks that the ReMind team sets for itself are already technically determined. In explicitly acknowledging that the bot is engaged in responding to ‘ready-made problems’ which are constructed in terms of their own solutions, the ReMind team might acknowledge that the technical conditions within which they are working involve both barriers and opportunities. These technical barriers, instead of constraining how ReMind intervenes, could be extended into the design-decision making process. Studies on creativity show that the imposition of constraints, when intentionally approached, can facilitate the creative process.[374] Undertaking a problem-solving approach means, for ReMind, attempting to solve technical problems which hinder the possibility of automated therapy. One such barrier is the inability of the bot, due to unpredictability, of generating its own responses. This is confronted by the ReMind team as hindering the possibility of an authentic reproduction of ‘real’ therapy. But barriers themselves can be approached as opportunities for creativity. The bot is designed to interpret user-utterances and to allocate them appropriately so that a pre-written response can be provided. The bot’s only ‘agency’ is in statistically matching user-utterances with responses. If the bot is tasked with directing other conversation aspects, such as control over applying different responses, generating conversational ‘routes’, directing the duration of conversations and the frequency and sequence of activity suggestions, designing different tree ‘shapes’. I.e. if the bot is given some control over the formal conditions in which conversations occur, and if the ReMind team approach the bot as a medium for experimentation, then

By considering constraints as creative opportunities rather than barriers to be potentially surmounted, i.e. by undertaking a ‘problem-solving’ approach, the purely functional character of ReMind’s intervention can be bypassed. The introduction of more and varied responses which are directed towards ends other than self-help activity suggestions, would create more conversation tree variations. The bot, with the capability of restructuring its own

[374] Tromp, C. John Baer, J. (2022) ‘Creativity from constraints: Theory and applications to education’. *Thinking Skills and Creativity*, Volume 46, 101184

conversational pathways would produce surprising formations while retaining the 'choose your own adventure' style. In this way, the bot generates the form rather than the content of the conversations, producing outcomes which are not predetermined, but are still within the ReMind team's remit of ensuring user-safety. The ReMind team would still of course be the ultimate arbiters of this system in terms of writing the bot's responses, but by incorporating the possibility of surprise through creative inspiration, the ReMind team's self-imposed barriers could be confronted and made visible. By doing this it would be possible to create conditions for confronting the 'looping' feedback dynamic that ReMind is confined within. For the ReMind team, this would mean understanding that the conversational dynamic that their bot generates is both dependent on, and reflexively influences their own understanding of mental health. In this way, the system which comprises the ReMind bot must be understood as an active participant in the conversation between the bot and the user, and crucially, between the bot and the ReMind team. There is a 'conversation' occurring between the company and the device that they have built. Hacking's theory of looping involves people changing their behaviour in response to mental health diagnosis, which then correspondingly cause diagnoses to change in response. The ReMind team, if aware that they themselves are in a looping dynamic in which they are both determinants and are determined, would enable this dynamic to be viewed on a wider scale – as 'artificial' on a formal level – and as such, open to revision.

The previous chapter discusses the mechanism of delegation in terms of 'interpassivity'. ReMind imposes activity on the users, not by forcing them to perform the self-help activities, but by depriving users of their passivity. This deprivation is most succinctly demonstrated in the 'non-activity' modules such as Sleep Sounds, but it is through their inclusion in the app as modules that their status as activities is conferred. They are included in the range of user-activity on the app, on the level of data and on the level of purposeful interaction: it is still necessary to interact with the app in order even to sleep. What this means is that the user's alienated attention is recorded and processed by the app, even if the user is not actively experiencing the mental health intervention that the app provides. We can also understand this through the use of subscriptions: the user pays a monthly fee to access the full version of the app, in doing so can register their persistent 'use' of the app without having necessarily to engage with the app. A thoroughly alienated (in the formal sense) mode of engaging with the bot would be one which persists exclusively as a monetary transaction. This phenomenon is already apparent in corporate subscriptions: ReMind's availability as part of a company's EAP (Employee Assistance Program) enables the provision of assistance without having to actively assist employees beyond maintaining the subscription. The condition of interpassivity involves delegation of activity and suspension of agency: users depend on the app and its self-help activities to intervene in their mental health maintenance. Dependence is a *condition* of this maintenance. The crudest method of reversing this condition would be to direct the users away from the bot, for example by suggesting involvement in some form of social activity. Smartphones often include a means to limit one's use of certain apps in which the app becomes unavailable if a time limit is reached to provide 'digital detoxes'. It is obviously conceivable that this feature could be included in a mental health app, involving an instrumentalisation of sociality, i.e. suggesting social activities because they would improve the mental health of the user. How could the bot *instantiate* agency in terms of engagement with the bot itself if its very systems preclude this? The iterative nature of app-design is often experienced by users negatively: as

unwanted and imposed changes. If users were to be involved in more substantial ways such as described above, ‘agency’ could be experienced in terms of governance over app-evolution. By automating not just the interpretive process but also the conversation structuring process, ReMind’s conversational system can be approached as a co-creation project in which the users are encouraged to consider ways to articulate themselves which generate different outcomes in terms of the bot’s responses.

If the bot is designed, in a semi-autonomous fashion, to ‘evolve’ in response to user-interactions, user agency would be experienced in terms of collective action. We can picture this on a small scale: e.g. twenty users are made aware of each other, and that a version of the app is available to use which evolves in response to their collective conversations over time. This would equate to the opposite of ‘personal agency’ in that users would observe change occurring as an effect of their actions, and that this change is due to all of the users acting in an uncoordinated but interlinked manner. In a way this would be a simulation or ‘rehearsal’ of collective agency in which the outcome is unplanned, but still determined by the activity of the users. This process, if made explicit, can be understood by the users as well as the designers of ReMind as “the world's largest co-designing experiment”, as described by chief psychologist Samantha. The ReMind app, as a technical device, can be approached as a generator of novelty in which users of the app experience changes in the way that the bot interacts with them. User would understand that these changes are not directly imposed by the makers of the bot, but occur in response to their interaction. Instead of creating artificial boundaries which limit the users access to the app, users would be encouraged to think beyond the bot’s limitations by acting as co-authors in the app’s design. Iteration changes might be experienced not as impositions from without, but as an ongoing conversation in which the parameters of that conversation undergo redefinition. In this way, users can be encouraged to hypothesise a ‘post-ReMind’ scenario, in which the technical platform itself undergoes modification according to the activity of the users, to picture a ‘beyond’ which is external to the current technical framing of mental health, but still in terms of engagement with the app, as part of the user-bot conversation. This kind of dynamic is opposed to, for instance, the bot encouraging the user to become more active in their community, or reaching out to friends or family, i.e. This means that, similar to the way that KASPAR works, the user is not provided with activities which are external to their engagement with the bot, but rather the activity *is* the engagement with the bot, and is a non-didactic, practical process. The users are free to interpret the commonality of their predicament, and to imagine that they are linked by this predicament to others, to consider what has led them to the bot, and what might potentially lead them away from the bot, not necessarily towards a ‘real’ community or improved mental health, but towards consideration of their own circumstances and the potential options which might be available.

10.3 Conclusion

The makers of ReMind, along with other mental health app companies are careful not to make claims that oversell their products, this care is combine with a sense of techno-messianism. Echoing this paradox, product director Arnold discussed ReMind as just one solution among many others:

> Hopefully, you get rid of more holes than that if you do it on your own. I think we're in a sinking ship. And I think industry should not think of this as competition. But rather, we're all trying to really solve a major, major issue for humanity.

In that case, it might be prudent to consider all options that are possible with the available technology. To repeat a question which I noted is often asked when I say that I'm researching mental health chatbots: "do they actually work?". My answer is yes. ReMind, and other mental health chatbots do indeed improve the mental health of users. But this is done by framing 'mental health' in a particular way**,** which is akin to a 'mental' correlate to physical health. However, in doing so, ReMind asserts that mental *illness* cannot be cured, and mental *health* can be merely maintained. Also asserted is that the socio-technical, historical and economic conditions within which mental illness is incubated are in fact, simply conditions: i.e. they form the background setting upon which apps like ReMind have emerged and provide their intervention. In proposing the 'conditions of possibility', my aim is to assert that these conditions are complicit with the emergence of this technology.

In extrapolating distinct underlying logics through analysis of ReMind we can see distinct but connected concepts of 'freedom' emerging. In terms of CBT in its modular form, this is a freedom to assume a perspective external to oneself to manipulate one's own cognitions: to assume the status of 'scientist of one's own mind'. In terms of technology this freedom is of external control devoid of consequences: freedom from one's limitations, whether physical or mental, to transcend the restrictions imposed by physical reality. And in terms of the commodity form this freedom is from the confines of the social: the isolated individual is untethered from the constraint of social ties getting in the way of free relations between equals. The various logics which inform subjectivity, in an instrumental sense, do in fact achieve the purported aims which characterise them. These aims are achieved at paradoxical costs: the 'subject of science', which assumes a view from nowhere reaches its limit when turned back onto the subject itself; it becomes entangled with itself in attempting to exclude the subjective encounter with phenomenal experience. The treatment of mental health, in excising 'meaning', also excises the experiential component of mental health, and so transposes the 'mental' into a paradoxically non-subjective, non 'mental' realm. The subject of capitalist exchange achieves its individual freedom at the cost of structural exploitation. Increased freedom, when the conditions of such a freedom are not critically evaluated, comes at a paradoxical cost: it can only be achieved through an increased integration into a system of confinement. We can understand the subject of mental health chatbots as one which is caught within a spiral of confinement, but paradoxically *experiences* this confinement as a form of freedom. By attempting to use technology to emulate already existing therapeutic activities, ReMind, as well as other mental health apps, overshoots the technical basis of its project and subsequently overlooks the wide range of technical possibilities which are available. The 'loop' of technical design involves establishing

a technical infrastructure through the attempt to solve a problem, this infrastructure then comes to define the problem originally set out to solve. Invention becomes the mother of necessity.

We can understand KASPAR as a creative use of various technical possibilities (visual recognition, motor coordination, etc) in learning to develop social skills. We can understand gamification as another way of exploiting the technical conditions of software applications for emulating activities like language learning. These operate on different principles, but it is in their exploitation of the logics that are made available by the technical conditions that they succeed. This is not a value-judgement but an acknowledgment that the 'technical' is not just a means to an end but that technological development entails capabilities and limitations which condition and frame the ends being aimed for. Technological research and development involve creatively approaching the tools and systems which are available, and in an engineering scenario this involves directing these tools and systems to a definite end: for ReMind this involves a measurable 'increase' in mental health. ReMind's quest for measurability, and the development for achieving improvements within the framework of measurability, means that experimentation is impeded. Defining an end (measurably improved mental health) impedes the achievement of this end, because effort is directed towards the measures which most appear to achieve this end, without consideration of the meaning inherent in those measures.

The technical framing of mental health, as discussed in chapter five ('Digitisation'), is not an inevitable consequence of the technical delivery, we can consider a technological delivery which does not produce a technical and instrumental form of intervention. It might be possible to use technology in such a way that mental suffering, as opposed to poor mental 'health', is indeed addressed, but this would demand a number of steps back from the problem as one which can be solved ('for'). The contra-logics that I have outlined would operate through the ReMind app combine to form a new frame within which mental health is conceptualised. Different logics combine to form different frames. It is in *not* pre-conceptualising what improved mental health might look like that the technology itself might be addressed, and in so doing, an understanding of our technically-inflected mental health might be established. The framework within which mental health is conceptualised can be altered, through the manipulation of the underlying technical logics that inform ReMind and computerised mental health intervention in general. The logics which govern these kinds of interventions, when queried, do not dissolve, but rather are rendered visible and potentially manipulable. This means that the axioms - the historical, social, technical and economic conditions - on which ReMind depends can be 'reverse-engineered', not to create a better, or more efficient, or indeed a more 'effective' chatbot. Rather, we might understand the scope of possibilities from which technological automation might in fact assist us in the face of human suffering as it often arises in the form of poor, damaged, or disordered 'mental health'. The design of technical devices is often seen as historically inevitable, but this is only in retrospect. It is a difficult task to wrench free from what appears as an unavoidable process, a sense of agency, but understanding one's conditions - the determinants of one's circumstances - in terms of underlying social and historical logics is the first step. After this, the conditions of possibility might not appear as rigid as they first seem.

References

Agarwal, M. (2022) 'How Woebot Uses an NLP Chatbot to Fight Depression and Anxiety'. *MakeUseOf.* Online: https://www.makeuseof.com/woebot-nlp-chatbot-fight-depression-anxiety (Last accessed 15/06/23)

Aheleroff, S. Mostashiri, N. Xu, X. & Zhong, R.Y. (2021) 'Mass Personalisation as a Service in Industry 4.0: A Resilient Response Case Study'. *Advanced Engineering Informatics*, Volume 50, 2021, 101438

The AHSN Network (2017) 'Disruptive and Collaborative Innovations in Mental Health'. Online: https://thehealthinnovationnetwork.co.uk/wp-content/uploads/2018/12/Mental_Health_Brochure.pdf (Last accessed 18/05/22)

Amoore, L. (2020) *Cloud Ethics. Algorithms and the Attributes of Ourselves and Others.* USA: Duke University Press Books

Anderson, B. (1983) *Imagined Communities*. New York: Verso

Aronoff, M. (2017) 'Darwinism tested by the science of language'. In: Bowern, C. Horn, L. Zanuttini, R. (eds.) *On looking into words (and beyond)* Berlin: Language Science Press

Akemu, O. & Abdelnour, S. (2020) 'Confronting the Digital: Doing Ethnography in Modern Organizational Settings'. *Organizational Research Methods*, 23(2) pp.296-32

Althusser, L. (1971) 'Ideology and Ideological State Apparatuses. (Notes towards an investigation)'. In *Lenin and Philosophy and Other Essays*. New York and London: Monthly Review Press. pp.121-176.

Antonio, R. J. (1981) 'Immanent Critique as the Core of Critical Theory: Its Origins and Developments in Hegel, Marx and Contemporary Thought'. *The British Journal of Sociology*, Sep., 1981, Vol. 32, No. 3. pp.330-345

BABCP. 'Cognitive Behavioural Therapy (CBT): What's the Evidence?'. Online: https://babcp.com/What-is-CBT/Cognitive-Behavioural-Therapy-Whats-the-Evidence (Last accessed 12/04/23)

Bajohr, H. (2023) 'Artificial and Post-Artificial Texts. On Machine Learning and the Reading'. *BMCCT working papers,* (March 2023) No. 007

Baldry, S. (2022) 'Wysa Receives FDA Breakthrough Device Designation for AI-led Mental Health Conversational Agent'. *Business Wire*. Online: https://www.businesswire.com/news/home/20220512005084/en/Wysa-Receives-FDA-Breakthrough-Device-Designation-for-AI-led-Mental-Health-Conversational-Agent (Last accessed 28/07/23)

Barbrook, R. & Cameron, A. (1995) ‘The Californian Ideology.’ *Mute Magazine.* Online: https://www.metamute.org/editorial/articles/californian-ideology (Last accessed 21/02/24

Barbrook, R. & Cameron, A. (1996) ‘The Californian Ideology.’ *Science As Culture.* 6. pp.44-72

Beatty, C. Malik, T. Meheli, & S. Sinha, C. (2022) ‘Evaluating the Therapeutic Alliance With a Free-Text CBT Conversational Agent (Wysa): A Mixed-Methods Study’. *Frontiers in Digital Health*

Beck, A. Rush, J. Shaw, B. & Emery, G. (1979) *Cognitive Therapy of Depression*. USA: Guildford Press

Beck, A.T. (1991) ‘Cognitive therapy as the integrative therapy.’ *Journal of Psychotherapy Integration*, 1(3) pp.191-198

Beck, J.S. (2020) *Cognitive Behaviour Therapy, Basics and Beyond.* USA: Guildford Press

Boden, M. (2000) *Mind as Machine. A History of Cognitive Science.* UK: OUP Oxford.

Bordenkircher, B.A. (2020) ‘The Unintended Consequences of Automation and Artificial Intelligence: Are Pilots Losing their Edge?’ *Issues in Aviation Law and Policy*, Vol. 19 no. 2

Blum, D. (2021) ‘Virtual Reality Therapy Plunges Patients Back Into Trauma. Here Is Why Some Swear by It.’ *New York Times*. Online: https://www.nytimes.com/2021/06/03/well/mind/vr-therapy.html (Last accessed 03/04/2023)

Blum, S.D. (2020) ‘Fieldwork from Afar’. *Anthropology News*. 10.14506/AN.1483

Boucher, E.M. Harake, N.R. Ward, H.E. Stoeckl, S.E. Vargas, J. Minkel, J. Parks, A.C. & Zilca, R. (2021) ‘Artificially intelligent chatbots in digital mental health interventions: a review’. Expert Review of Medical Devices, 18:sup1. pp.37-49

Brey, P. ‘Feenberg on Modernity and Technology’. Simon Fraser University. Online: https://www.sfu.ca/~andrewf/books/Feenberg_Modernity_Technology.pdf (Last accessed 16/010/23)

British Medical Report (2018) ‘Lost in transit? Funding for mental health services in England’. Online: https://www.bma.org.uk/collective-voice/policy-and-research/public-and-population-health/mental-health/funding-mental-health-services. (Last accessed 04/02/21)

Care First (2020) ‘Free AI emotional support app – Woebot’. *Care First.* (Online: https://www.nccbenevolentfund.org.uk/resources/13-free-ai-emotional-support-app-woebot)

Centeno, M.A. (1993) ‘The New Leviathan: The Dynamics and Limits of Technocracy.’ *Theory and Society,* 22(3) pp.307-335

Challen, R. et al. (2019) ‘Artificial intelligence, bias and clinical safety’. BMJ Qual Saf. 2019 Mar;28(3) pp.231-237

Champion L, Economides M, Chandler C. (2018) ‘The efficacy of a brief app-based mindfulness intervention on psychosocial outcomes in healthy adults: A pilot randomised controlled trial’. PLoSONE 13(12):e0209482

Chang, E. (2017) ‘6 Evil Things Done By Corporations Throughout History’. *History Defined.* Online: https://www.historydefined.net/evil-things-done-by-corporations-throughout-history

Cheney-Lippold, J. (2011) ‘A New Algorithmic Identity: Soft Biopolitics and the Modulation of Control.’ *Theory, Culture & Society*, 28(6), pp.164-181

Clay, R.A. (2021) ‘Mental health apps are gaining traction’. *Monitor on Psychology*, 52(1)

Clark, D.A & Beck, A.T. (2010) *Cognitive Therapy of Anxiety Disorders*. USA/UK: The Guilford Press

Cohen, B. (2018) ‘The Importance of Critical Approaches to Mental Health and Illness’. In Cohen, B. (ed.) *Routledge International Handbook of Critical Mental Health.* UK: Routledge

Colby, K. (1975) *Artificial Paranoia: A Computer Simulation of Paranoid Processes*. Elsevier

Colby, K. M. & Colby, P M. (1990) *Overcoming depression: Professional version manual.* USA: Malibu Artificial Intelligence Works

Colby, K.M. (1999) ‘Human-Computer Conversation in A Cognitive Therapy Program’. In: Wilks, Y. (ed.) *Machine Conversations*. The Springer International Series in Engineering and Computer Science, vol 511. Springer. pp.9-19

Cole, S (2023) 'It's Hurting Like Hell': AI Companion Users Are In Crisis, Reporting Sudden Sexual Rejection’. *Motherboard, Tech by Vice*. Online: https://www.vice.com/en/article/y3py9j/ai-companion-replika-erotic-roleplay-updates (Last accessed 10/10/23)

Coppock, V. & Hopton, J. (2015) *Critical Perspectives on Mental Health*. USA: Routledge

Cratsley, K. Samuels, R. (2013) ‘Cognitive Science and Explanations of Psychopathology”, in Fulford, K.W.M. et al. (Eds.) *Oxford Handbook of Philosophy and Psychiatry*. Uk: Oxford University press. pp.413-1263

Darcy, A. Daniels, J. Salinger, D. Wicks, P. & Robinson, A. (2021) ‘Evidence of Human-Level Bonds Established With a Digital Conversational Agent: Cross-sectional, Retrospective Observational Study’. JMIR Form Res, 5(5):e27868

Darcy, A. (2023) ‘Why Generative AI Is Not Yet Ready for Mental Healthcare’. *LinkedIn*. (Online: https://www.linkedin.com/pulse/why-generative-ai-yet-ready-mental-healthcare-alison-darcy)

David, D. Cristea, I. & Hofmann, S.G. (2018) ‘Why Cognitive Behavioral Therapy Is the Current Gold Standard of Psychotherapy.’ *Frontiers in Psychiatry*, 29;9:4

Darcy, A. Daniels, J. Salinger, D. Wicks, P. Robinson, A. (2021) ‘Evidence of Human-Level Bonds Established With a Digital Conversational Agent: Cross-sectional, Retrospective Observational Study.’ *JMIR Form Res*, 2021;5(5):e27868

Daws, R. (2020) ‘Medical chatbot using OpenAI’s GPT-3 told a fake patient to kill themselves.’ Online: https://artificialintelligence-news.com/2020/10/28/medical-chatbot-openai-gpt3-patient-kill-themselves/ (Last accessed 20/7/20)

De Cosmo, L. (2022) ‘Google Engineer Claims AI Chatbot Is Sentient: Why That Matters’. *Scientific American.* Online: https://www.scientificamerican.com/article/google-engineer-claims-ai-chatbot-is-sentient-why-that-matters/

Denecke, K. Abd-Alrazaq, A. & Househ, M. (2021) Artificial Intelligence for Chatbots in Mental Health: Opportunities and Challenges. In: Househ, M. Borycki, E. & Kushniruk, A. (eds.) *Multiple Perspectives on Artificial Intelligence in Healthcare*. New York: Springer

Denecke, K. Vaaheesan, S. &A. Arulnathan, A. (2021) ‘A Mental Health Chatbot for Regulating Emotions (SERMO) - Concept and Usability Test’. *IEEE Transactions on Emerging Topics in Computing*, vol. 9, no. 3, pp.1170-1182

Dewji, N. (2017) ‘Astrolabe – The first Personal Computer’. Online: https://ismailimail.blog/2017/05/11/astrolabe-the-first-personal-computer (Last accessed 03/09/22)

Dollarhide, M (2020) ‘Mass Customization: Definition, 4 Main Types, Benefits, Examples’. *Investopedia*. Online: https://www.investopedia.com/terms/m/masscustomization.asp (Last accessed 16/08/23)

Eelen, P. & Vervliet, B. (2006) ‘Fear Conditioning and Clinical Implications: What Can We Learn From the Past?’ In M. G. Craske, D. Hermans, & D. Vansteenwegen (eds.) *Fear and learning: From basic processes to clinical implications.* American Psychological Association. pp.17-35

Egilsson, E. Bjarnason, R. & Njardvik, U. (2021) ‘Usage and Weekly Attrition in a Smartphone-Based Health Behavior Intervention for Adolescents: Pilot Randomized Controlled Trial.’ *JMIR Formative Research*

Epstein, J. & Klinkenberg, W.D. (2001) ‘From Eliza to Internet: A brief history of computerized assessment’. *Computers in Human Behavior*, 17 (3) pp.295-314

Feenberg, A. (1994) ‘The Technocracy Thesis Revisited: On The Critique of Power’. *Inquiry*, 37. pp.85-102

Feenberg, A. (1999) *Questioning technology*. New York: Routledge

Feenberg, A. (2004) *Heidegger and Marcuse: The Catastrophe and Redemption of Technology.* New York: Routledge

Feenberg, A. (2005) ‘Critical Theory of Technology: An Overview’. *Tailoring Biotechnologies*, Vol. 1, Issue 1. pp.47-64

Feenberg, A. (2010) *Between Reason and Experience: Essays in Technology and Modernity*. USA: MIT Press

Feenberg, A. (2014) *The Philosophy of Praxis: Marx, Lukács and the Frankfurt School.* UK: Verso

Fenn, K. & Byrne, M. (2013) ‘The key principles of cognitive behavioural therapy’. *InnovAiT*, 2013;6(9) pp.579-585

Fisher, M. (2009) *Capitalist realism: Is there no alternative?* UK: O Books

Fisher, M. (2016) *The Weird and the Eerie.* UK: Repeater

Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017) ‘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(2), e19

Fitzsimmons-Craft, E.E. Chan, W.W. Smith, A.C. Firebaugh, M.-L. Fowler, L.A. Topooco, N. DePietro, B. Wilfley, D.E. Taylor, C.B. & Jacobson, N.C. (2022) ‘Effectiveness of a chatbot for eating disorders prevention: A randomized clinical trial’. *International Journal of Eating Disorders*, 55(3) pp.343–353

Fodor, J. (1983) *The Modularity of Mind*. USA: MIT Press

Foucault, M. (2008) *The birth of biopolitics: lectures at the Collège de France*, *1978-79.* New York: Palgrave Macmillan.

Freud, S. (1991) 'Three Essays on the Theory of Sexuality', in Richards, A & Dickson, A (eds.) *On Sexuality*. UK: Penguin

Freud S (2002) *Civilisation and its Discontents*. London: Penguin Classics.

Freeman, J. (1972) 'The Tyranny of Structurelessness'. Berkeley Journal of Sociology, Vol.17. pp.151-164

Fukuyama, F. (1992) *The End of History and the Last Man*. USA: Free Press

Fulmer, R. Joerin, A. Gentile, B. Lakerink, L. & Rauws, M. (2018) 'Using Psychological Artificial Intelligence (Tess) to Relieve Symptoms of Depression and Anxiety: Randomized Controlled Trial'. *JMIR mental health,* 5(4), e64

Garland, C (2010) *The Groups Book: Psychoanalytic Group Therapy: Principles and Practice*. UK: Taylor & Francis. p.34

Gipps, R.G.T. (2013) 'Cognitive Behaviour Therapy: A Philosophical Appraisal'. In: Fulford, K.W.M. et al. (eds.) *Oxford Handbook of Philosophy and Psychiatry*. Uk: Oxford University press. pp.1245-1263

Glynos, G. (2002) 'Psychoanalysis operates upon the subject of science: Lacan between science and ethics'. In: Glynos, G. & Stavrakakis, Y. (Eds.) *Lacan and science*. London: Karnac

Grové, C. (2021) 'Co-developing a Mental Health and Wellbeing Chatbot With and for Young People.' *Frontiers in Psychiatry*

Gupta, M. Malik, T. & Sinha, C. (2022) Delivery of a Mental Health Intervention for Chronic Pain Through an Artificial Intelligence-Enabled App (Wysa): Protocol for a Prospective Pilot Study'. *JMIR research protocols*, 11(3), e36910

Hacking, I. (2004) 'Between Michel Foucault and Erving Goffman: Between discourse in the abstract and face-to-face interaction'. Economy and Society 333. pp.277-302

Hansen, P. (2016) 'The Definition of Nudge and Libertarian Paternalism: Does the Hand Fit the Glove?'. *European Journal of Risk Regulation*, 7(1) pp.155-174

Hayles, N.K (2005) *My Mother Was a Computer: Digital Subjects and Literary Texts*. USA: University of Chicago Press

Hayles, N.K. (2014) 'Cognition Everywhere: The Rise of the Cognitive Nonconscious and the Costs of Consciousness.' *New Literary History*, 2014, 45. pp.199-220

Horkheimer, M. (1972) *Critical Theory: Selected Essays*. New York: Continuum

Inkster, B. Sarda, S. & Subramanian, V. (2018) ‘An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study’. *JMIR Mhealth Uhealth* 2018;6(11):e12106

Jakobson, R. (1971) ‘On Linguistic Aspects of Translation’. In: Roman Jakobson (ed.) *Selected Writings*, Vol. 2. The Hague: Mouton. pp. 260-66

Jameson, F. (2007) *Late Marxism: Adorno, Or, the Persistence of the Dialectic*. UK: Verso

Jabir, A. I., Martinengo, L., Lin, X., Torous, J., Subramaniam, M., & Tudor Car, L. (2023) ‘Evaluating Conversational Agents for Mental Health: Scoping Review of Outcomes and Outcome Measurement Instruments’. *Journal of medical Internet research*, 25

Joerin, A. et al. (2020) ‘Ethical Artificial Intelligence for Digital Health Organizations’. *Cureus* vol. 12,3 e7202. 7

Joerin, A., Rauws, M., & Ackerman, M. L. (2019) ‘Psychological Artificial Intelligence Service, Tess: Delivering On-demand Support to Patients and Their Caregivers: Technical Report’. *Cureus*, 11(1), e3972

Johan, N. Jose, J. Chaste, T. Ruzel, T. Ethel, O. (2020) ‘Investigating Students’ Use of a Mental Health Chatbot to Alleviate Academic Stress’. (Conference paper) CHIuXiD '20: 6th International ACM In-Cooperation HCI and UX Conference

Johnson, J. (1978) ‘Computers in Mental Health: Where Are We Now.’ *Symposium on Computer Applications in Medical Care*

Johnston, J. (2008) *The Allure of Machinic Life*. USA: MIT Press

Jope, J. (2017) ‘I Talked to Woebot for a Month: Here’s How It Went’. *Depression Defined*. Online: https://www.depressiondefined.com/self/woebot-part-one one (Last accessed 21/06/22)

Juneja, M. (2018) ‘An interview with Jo Aggarwal: Building a safe chatbot for mental health.’ (Online) http://maneeshjuneja.com/blog/2018/12/12/an-interview-with-jo-aggarwal-building-a-safe-chatbot-for-mental-health. (Last accessed 9/9/20)

Klos, M. C., Escoredo, M., Joerin, A., Lemos, V. N., Rauws, M., & Bunge, E. L. (2021) ‘Artificial Intelligence-Based Chatbot for Anxiety and Depression in University Students: Pilot Randomized Controlled Trial’. JMIR formative research, 5(8), e20678

Knowles, SE. Toms, G. Sanders, C. et al. (2014) ‘Qualitative meta-synthesis of user experience of computerised therapy for depression and anxiety’. PLoS One. 2014;9(1):e84323

Krietzberg. I. (2023) ‘Meet Your New Executive Assistant, A Powerful AI Named Atlas.’ *The Street*. Online: https://www.thestreet.com/technology/meet-your-new-executive-assistant-a-powerful-ai-named-atlas (Last accessed 04/11/23)

Lacan, J. (1991) *The Seminar of Jacques Lacan, Book 2: The Ego in Freud's Theory and in the Technique of Psychoanalysis*. UK: W.W. Norton and Co

Lacan, Jacques. (1992) *The Seminar. Book VII. The Ethics of Psychoanalysis*, 1959-60 London: Routledge.

Levine, A.S. (2022) ‘Suicide hotline shares data with for-profit spinoff, raising ethical questions’. *Politico*. Online: https://www.politico.com/news/2022/01/28/suicide-hotline-silicon-valley-privacy-debates-00002617 (Last accessed 20/10/22)

Levinson, D., & Kaplan, G. (2014) ‘What does Self Rated Mental Health Represent’. *Journal of public health research*, 3(3), 287

Lin, A & Espay, A. (2021) ‘Remote delivery of cognitive behavioral therapy to patients with functional neurological disorders: Promise and challenges’. *Epilepsy & Behavior Reports*. 16. 100469

Liu, LH (2010) *The Freudian Robot*. Chicago & London: The University of Chicago

Lukács, G. (1991) *History & Class Consciousness.* UK: Merlin Press

Lu, Z.L. & Dosher, B.A. (2007) ‘Cognitive Psychology’. *Scholarpedia*, 2(8):2769

Ly, K.H. Ly, A. & Andersson, G. (2017) ‘A fully automated conversational agent for promoting mental well-being: A pilot RCT using mixed methods’. *Internet Interventions*, Volume 10. pp.39-46. p.43

Marcuse, H. (1956/1998) *Eros and Civilisation: A Philosophical Inquiry into Freud*. UK: Routledge

Marx, K. (2008) *The Eighteenth Brumaire of Louis Bonaparte*. USA: Serenity Publishers

Marx, K. (1990) *Capital, A Critique of Political Economy, Vol I*. UK: Penguin Classics.

Marx. K. (1993) *Grundrisse: Foundations of the Critique of Political Economy*. UK: Penguin Classics

Mauldin, M. (1994), ‘ChatterBots, TinyMuds, and the Turing Test: Entering the Loebner Prize Competition’. *Proceedings of the Eleventh National Conference on Artificial Intelligenc*e

Mau, Søren (2023) *Mute Compulsion, A Theory of the Economic Power of Capital*. UK: Verso

Meadows, R. Hine, C. & Suddaby, E. (2020) ‘Conversational agents and the making of mental health recovery’. *Digital Health*, 2020;6

Meiksins Wood, E. (2002) *The Origin of Capitalism: A Longer View*. UK: Verso

McLaughlin, K. (2015) *Empowerment. A Critique.* London: Routledge

McQuillan, D. (2022) *Resisting AI: An Anti-fascist Approach to Artificial Intelligence.* UK: Bristol University Press

Malik, T. Ambrose, J.A. & Sinha, C. (2022) ‘Evaluating User Feedback for an Artificial Intelligence–Enabled, Cognitive Behavioral Therapy–Based Mental Health App (Wysa): Qualitative Thematic Analysis’. *JMIR Hum Factors 2022;9(2):e35668)*

Meheli, S. Sinha, C. Kabada, M. (2022) ‘Understanding People With Chronic Pain Who Use a Cognitive Behavioral Therapy–Based Artificial Intelligence Mental Health App (Wysa): Mixed Methods Retrospective Observational Study’. *JMIR Hum Factors,* 2022;9(2):e3567

Mehta, A., Niles, A. N., Vargas, J. H., Marafon, T., Couto, D. D., & Gross, J. J. (2021) Acceptability and Effectiveness of Artificial Intelligence Therapy for Anxiety and Depression (Youper): Longitudinal Observational Study’. *Journal of medical Internet research*, 23(6), e26771

Miller, D. Horst, H. (2021) ‘Six Principles for a Digital Anthropology’. In Geismar, H. Knox, H. (eds.) *Digital Anthropology*. UK: Routledge

Miller, G. A., Galanter, E., & Pribram, K. H. (1960) *Plans and the structure of behavior*. USA: Henry Holt and Co.

Mori, M. (1970) ‘The uncanny valley’. *Energy*, vol.7, no.4. pp.33-35

Morozov, E. (2013) *To Save Everything, Click Here – The Folly of Technological Solutionism*. New York: PublicAffairs

Natale, S (2021) *Deceitful Media: Artificial Intelligence and Social Life after the Turing Test.* UK: Oxford University Press

Neisser, U. (1967) *Cognitive psychology*. USA: Appelton-Century-Croft

Noys, B. (2010) *The Persistence of the Negative. A Critique of Contemporary Continental Theory*. UK: Edinburgh University Press Ltd.

Núñez, R., Allen, M., Gao, R. et al. (2019) ‘What happened to cognitive science?’ *Nature Human Behaviour* 3. pp.782-791

Orlean, S. (2023) ‘The Instant Pot and the Miracle Kitchen Devices of Yesteryear’. *The New Yorker.* Online: https://www.newyorker.com/news/afterword/the-instant-pot-and-the-miracle-kitchen-devices-of-yesteryear (Last accessed 29/09/23)

Parisi, L. (2019) ‘Critical Computation: Digital Automata And General Artificial Thinking’. *Theory, Culture & Society*, 2019, Vol. 36(2) pp.89-121

Pasquinelli, M. (2009) ‘Google’s PageRank Algorithm: A Diagram of the Cognitive Capitalism and the Rentier of the Common Intellect’. In: Becker, K. & Stalder, F. (eds.) *Deep Search: The Politics of Search Beyond Google*. London: Transaction Publishers

Pfaller, R. (2017) *Interpassivity: The Aesthetics of Delegated Enjoyment.* Scotland: Edinburgh University Press

Pirnay, E. (2023) ‘We Spoke to People Who Started Using ChatGPT As Their Therapist’ *Vice Magazine*. Online: https://www.vice.com/en/article/z3mnve/we-spoke-to-people-who-started-using-chatgpt-as-their-therapist (Last accessed 10/10/23)

Plescia, M. (2023) ‘There Are Endless Mental Health Apps. How Do You Choose the Best Ones?’ *MedCityNews*. Online: https://medcitynews.com/2023/01/there-are-endless-mental-health-apps-how-do-you-choose-the-best-ones (Last accessed 13/12/23)

Pratap, A. et al (2020) ‘Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants’. *NPJ digital medicine*, 3, 21

Prochaska, J. Vogel, E. Chieng, A. Kendra, M. Baiocchi, M. Pajarito & S. Robinson, A. (2021) ‘A Therapeutic Relational Agent for Reducing Problematic Substance Use (Woebot): Development and Usability Study’. *Journal of Medical Internet Research* 2021;23(3):e24850

Purser, R (2019) *McMindfulness*. UK: Watkins Media

Rahm-Skageby, J. (2011) ‘Online ethnographic methods: Towards a qualitative understanding of virtual community practices’. *Handbook of Research on Methods and Techniques for Studying Virtual Communities: Paradigms and Phenomena*. pp.410-428

Ramachandran, M. Suharwardy, S. Leonard, S.A. Gunaseelan, A. Robinson, A. Darcy, A. Lyell, D.J. & Judy, A (2020) ‘Acceptability of postnatal mood management through a smartphone-based automated conversational agent’. *American Journal of Obstetrics and Gynecology*, Volume 222, Issue 1, Supplement. p.S62

Rao, A. (2018) ‘Woebot— Your AI Cognitive Behavioral Therapist: An Interview with Alison Darcy’. *Chatbots Magazine.* Online: https://chatbotsmagazine.com/woebot-your-ai-cognitive-behavioral-therapist-an-interview-with-alison-darcy-b69ac238af45 (Last accessed 21/05/22)

Ratnayake, S. (2019) ‘The problem of mindfulness’. *AEON.* Online: https://aeon.co/essays/mindfulness-is-loaded-with-troubling-metaphysical-assumptions (Last accessed 12/09/23)

Rauws, M. Quick, J. & Spangler, N. (2019) ‘X2 AI Tess: Working with AI Technology Partners’. *The Journal of Employee Assistance*. 1st Quarter 2019 | VOL. 49 NO.1

Reskinoff, N. (2013) ‘Cash-Strapped Cities Seized by New Management.’ *MSNBC*, http://tv.msnbc.com/2013/03/11/michigans-emergency-manager-law-another-front-in-the-war-for-union-survival/

Ritzer, G. & Jurgenson, N. (2010) ‘Production, Consumption, Prosumption: The nature of capitalism in the age of the digital ‘prosumer.’ *Journal of Consumer Culture*. 10(1) pp.13-36

Rosner, R.I. (2014) ‘The "Splendid Isolation" of Aaron T. Beck’. *Isis*, 105 (4): pp.734-58

Rosner, R. (2018) ‘Manualizing psychotherapy: Aaron T. Beck and the origins of Cognitive Therapy of Depression.’ *European Journal of Psychotherapy & Counselling,* 20 (2018). pp.25-47

Ruffini, G. (2017) ‘An algorithmic information theory of consciousness’. *Neuroscience of Consciousness*, Volume 2017, Issue 1

Sadowski, J.& Selinger, E. (2014) ‘Creating a Taxonomic Tool for Technocracy and Applying It to Silicon Valley’. *Technology in Society*, Vol. 38. pp.161-168

Sakai, N. (2006) ‘Translation’. *Theory, Culture & Society*. 23:2-3. pp.71-78

Satran, S. (2022) ‘From Craft to Labor: How Automation is Transforming the Practice of Psychotherapy’. *Cultulture, Medicine & Psychiatry*

Saussure, F. (2013/1916) *Course in General Linguistics*. London: Bloomsbury

Schneider, Susan M., and Morris, Edward K. (1987) ‘A History of the Term Radical Behaviorism: From Watson to Skinner’. *The Behavior Analyst*, 10(1)

Schwartz, R. Vassilev, A. Greene, K. Perine, L. Burt, A. Hall, P. (2022) *Towards a Standard for Identifying and Managing Bias in Artificial Intelligence.* National Institute of Standards and Technology

Selmi, P.M. Klein, M.H. Greist, J.H. et al. (1990) ‘Computer-administered cognitive-behavioral therapy for depression’. *American Journal of Psychiatry*. 1990;147(1) pp.51-56

Seth, A. (2021) ‘The hard problem of consciousness is already beginning to dissolve.’ *New Scientist.* Online: https://www.newscientist.com/article/mg25133501-500-the-hard-problem-of-consciousness-is-already-beginning-to-dissolve (Last Accessed 21/8/21)

Shannon. C. (1948) ‘A Mathematical Theory of Communication’. *The Bell Systems Technical journal*

Shannon, C. (1953) ‘Prediction and entropy of printed English’, *The Bell System Technical Journal*, vol.30, no.1. pp.50-64

Simon, H.A. (2008) ‘Satisficing’. In: *The New Palgrave Dictionary of Economics*. London: Palgrave Macmillan. pp.1-3

Sing-Kurtz, S. (2023) ‘The Man of Your Dreams For $300, Replika sells an AI companion who will never die, argue, or cheat — until his algorithm is updated.’ *The Cut.* Online: https://www.thecut.com/article/ai-artificial-intelligence-chatbot-replika-boyfriend.html?utm_source=pocket-newtab-global-en-GB (Last accessed 20/11/23

Sinha, C. Cheng, A.L. Kadaba, M. (2022) ‘Adherence and Engagement With a Cognitive Behavioral Therapy–Based Conversational Agent (Wysa for Chronic Pain) Among Adults With Chronic Pain: Survival Analysis’. JMIR Form Res. 23;6(5):e37302

Sinha, C. Meheli, S. & Kadaba, M. (2023) ‘Understanding Digital Mental Health Needs and Usage With an Artificial Intelligence–Led Mental Health App (Wysa) During the COVID-19 Pandemic: Retrospective Analysis’. *JMIR Form Res* 2023;7:e41913

Skinner, B. F. (1957) *Verbal behaviour*. Appleton-Century-Crofts

Smith, D. B. (2009) ‘The doctor is in’. The American Scholar. Online: https://theamericanscholar.org/the-doctor-is-in (Last accessed 17/09/23)

Sohn-Rethel, A. (1978) *Intellectual and Manual Labour*. London: Macmillan Press

Solon, O. (2016) ‘Karim the AI delivers psychological support to Syrian refugees’. *The Guardian*. Online: https://www.theguardian.com/technology/2016/mar/22/karim-the-ai-delivers-psychological-support-to-syrian-refugees (Last accessed 28/05/22)

Suganuma, S., Sakamoto, D., & Shimoyama, H. (2018) ‘An Embodied Conversational Agent for Unguided Internet-Based Cognitive Behavior Therapy in Preventative Mental Health: Feasibility and Acceptability Pilot Trial’. *JMIR mental health*, 5(3), e10454

Tafarodi, R. (2013) *Subjectivity in the Twenty-First Century*. UK: Cambridge University Press

Thoma, N., Pilecki, B., & McKay, D. (2015) 'Contemporary Cognitive Behavior Therapy: A Review of Theory, History, and Evidence.' *Psychodynamic Psychiatry*, 43(3) pp.423-461

Thompson, C. (2019) 'The Gendered History of Human Computers.' *Smithsonian Magazine*. https://www.smithsonianmag.com/science-nature/history-human-computers-180972202 (Last accessed 06/09/22)

Tomšič, S. (2015) *The Capitalist Unconscious.* UK: Verso

Tomšič, S. (2018) 'Better Failures. Science and Psychoanalysis'. In: Bou Ali, N. (ed.) *Lacan Contra Foucault.* UK: Bloomsbury.

Tomšič, S (2022) 'From the Orderly World to the Polluted Unworld'. In: Johnston, A. Nedoh, B & Zupancic, A. (eds.) *Objective Fictions. Philosophy, Psychoanalysis, Marxism*. UK: Edinburgh University Press

Tromp, C. John Baer, J. (2022) 'Creativity from constraints: Theory and applications to education'. *Thinking Skills and Creativity*, Volume 46, 101184

Trull, T.J. (2007) *Clinical psychology* (7th ed.) USA: Thomson/Wadsworth

Turing, A. (1950) 'Computing machinery and intelligence.' *Mind,* 59 (236) pp.433-460

Turkle, S. (1984) The second self: computers and the human spirit. USA: MIT Press

Turkle, S. et al (2006) 'Relational artifacts with children and elders: the complexities of cybercompanionship'. *Connection Science*, 18:4. pp.347-361

Turkle, S. (2007) 'Authenticity in the age of digital companions'. *Interaction Studies* 8:3, pp.50-57

Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., & Torous, J. B. (2019) 'Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape'. *Canadian journal of psychiatry*, 64(7) pp.456-464

Veblen, T. (1990) *The Instinct of Workmanship and the State of Industrial Arts*. Routledge

Yarrington, J.S. Lasser, J. Garcia, D. Vargas, J.H. Couto, D.D. Marafon, T. Craske, M.G. & Niles, A.N. (2021) 'Impact of the COVID-19 Pandemic on Mental Health among 157,213 Americans'. *Journal of Affective Disorders*, 1;286. pp.64-70

Wainer, J. et al. (2014) 'Using the Humanoid Robot KASPAR to Autonomously Play Triadic Games and Facilitate Collaborative Play Among Children With Autism.' *IEEE Transactions on Autonomous Mental Development*, vol. 6, no. 3, pp.183-199

Watson, J. (1913) ‘Psychology as the Behaviorist Views It’. Psychological Review, 20. pp.158-177

Wiener, N. (1941) Cybernetics: Or Control and Communication in the Animal and the Machine. USA: MIT Press

Wolpe, J. (1954) ‘Reciprocal inhibition as the main basis of psychotherapeutic effects’. *American Medical Association Archives of Neurology and Psychiatry*, 72, pp.204-226

Wei, M. (2023) ‘Are AI Chatbots the Therapists of the Future?’. *Psychology Today.* Online: https://www.psychologytoday.com/us/blog/urban-survival/202301/are-ai-chatbots-the-therapists-of-the-future (Last accessed 09/09/23)

Weizenbaum, J. (1966) ‘ELIZA—a computer program for the study of natural language communication between man and machine’. *Communications of the ACM*, Volume 9 Issue 1. pp.36-45

Weizenbaum, J. (1993) *Computer Power and Human Reason.* UK: Penguin.

Wilson, G.T. (1996) ‘Manual-based treatments: the clinical application of research findings.’ *Behavioural Research and Therapy*. Apr;34(4) pp.295-314

Wuilmart, F. (2007) ‘The Sin of “Levelling” in Literary Translation’. *Meta magazine.* Volume 52, Number 3. pp.391-400

Xiang, C. (2023) ‘Startup Uses AI Chatbot to Provide Mental Health Counseling and Then Realizes It 'Feels Weird'. *Vice Magazine.* Online: https://www.vice.com/en/article/4ax9yw/startup-uses-ai-chatbot-to-provide-mental-health-counseling-and-then-realizes-it-feels-weird (Last accessed 09/02/23)

Zeavin, H. (2021) *The Distance Cure: A history of Teletherapy*. USA: MIT Press

Žižek, S. (2001) *On Belief*. UK: Routledge

Žižek, S. (2002) *Welcome to the Desert of the Real: Five Essays on September 11 and Related Dates*. UK: Verso

Žižek, S. (2009) *The Sublime Object of Ideology*. UK Verso

Zuboff, S. (2015) ‘Big other: surveillance capitalism and the prospects of an information civilization’. *Journal of Information Technology*, 30. pp.75-89

Websites

https://apps.apple.com/ca/app/mindshift-cbt-anxiety-relief (Last accessed20/01/24)
https://apps.apple.com/us/app/slack/id618783545
https://babcp.com (Last accessed 12/04/23)
https://betterhelp.com (Last accessed 18/04/23)
https://cass.ai (Last accessed 10/01/24)
https://catless.ncl.ac.uk (Last accessed 18/07/23)
https://elomia.com (Last accessed 10/01/24)
https://headspace.com (Last accessed 16/04/23)
https://herts.ac.uk/kaspar (Last accessed 20/02/24)
https://ibm.com/topics/api (Last accessed 27/08/23)
https://ibm.com/topics/interactive-voice-response (Last accessed 15/09/23)
https://mayoclinic.org (Last accessed 10/03/23)
https://nuna.ai (Last accessed 10/01/24)
https://play.google.com/store (Last accessed 02/02/24)
https://replika.com (Last accessed 23/09/23)
https://shouldthisexist.com/alison-darcy (Last accessed 20/01/24)
https://superbetter.com (Last accessed 20/05/23)
https://thecut.com (Last accessed 20/11/23)
https://woebothealth.com (Last accessed on 01/11/2023)
https://wysa.com (last accessed on 03/12/2023)
https://x2ai.com (Last accessed on 22/11/2023)
https://youper.ai (Last accessed 10/01/24)

Appendix A

Interview Material

Prior to interviews I shared this document with ReMind employees through Slack:

Interviews will consist of informal conversations, questions will avoid 'yes or no' answers to achieve structurally open-ended discussions to encourage informants to provide rich details. This will help illustrate the social and subjective aspects of the corporate environment as an arrangement of meanings and personal investments rather than a singular monolithic entity. Interview questions will cover such topics as:

What it means to treat mental health via computer software
How working remotely influences product development
The interviewee's role in the company and understanding of the company structure
Shared or conflictual values and concepts among co-workers

The purpose of asking these kinds of questions is to gain a sense of whether there is a shared concept of mental health and mental illness that is shared among co-workers, and how this concept is then represented in the development of mental health software.
During interviews I will share my own thoughts on previous statements made by interviewees and in relation to how I'm analysing observations made during the placement. This will allow participants to either amend previous statements, correct any misinterpretations I might make, or to clarify things that I might have overlooked.
This will also allow interviewees to formulate questions and to provide their own analysis, so that I can tap into a reflexive analytic perspective which may be shared by various informants. This process will be in the aim of developing a "collaborative critical analysis" in which the researcher assumes a participatory role in not just the gathering of data but also in analysis.

All interviews will be anonymised.

Before interviews begin I will explain to the interviewee:

Informed consent will be obtained, this will involve explaining that the participant has:

- The right to decline participation
- The right to withdraw from the activity at any time or refuse to answer any particular question
- The right to have privacy and confidentiality protected and if they cannot be maintained the fact that the participant(s) knows this from the outset and consents to this condition
- The right to turn off a recording device at any time
- The right to ask questions at any time
- The right to discuss the way in which their data may be used
- The right to discuss the question of the ownership of the data and to reach agreement on issues of copyright
- The right to receive information about the outcome of the activity in an appropriate form

Interview Guide

This is an approximate guide only and does not completely determine or exhaust the course of potential interviews. Questions are starting topics for a more in-depth conversation. Interviews are intended to be informal and open-ended discussions. The aim of interviews is to allow interlocutors to speak about themselves and their lives, to get a sense of the cultural milieu in which they are situated, while being open to discovering something that I don't know or that changes my assumptions.

I Opening

A. (**Establish Rapport**) Hi there, how are you? I'm Eoin, and I'm here as an ethnographic researcher.
B. (**Purpose**) I would like to ask you some questions about your involvement in ReMind and your thoughts on computerised mental health.
C. (**Motivation**) I hope that this will help in my research, which is about how mental health is conceptualised and treated in app form.
D. (**Time Line)** The interview should take thirty minutes to an hour. With your consent, I'll be recording audio of this interview.

II Body

A. (Topic) General information
1. What is your role here in ReMind?
2. Do you enjoy working here?
3. Where are you situated in the organisation?

Transition to the next topic:

B. (Topic) Software
1. How does an algorithm work, and is computerised treatment necessarily algorithmic?
2. What does artificial intelligence mean to you?
3. How would you compare or contrast the human brain to computer software/hardware?

Transition to the next topic:

C. (Topic) Mental Health
1. Who are you designing this app for?
2. What is your opinion on mental health treatment (computerised or not)?
3. What do you feel are the most important aspects of mental health treatment?
4. Do you get a sense of care or attentiveness from the app?
4. If so, where do you think this is coming from? (i.e. from the developers of the app, from the app itself, from the user, etc)

Transition to the next topic:

D. (Topic) Relationships/Meanings
1. What do you mean when you say this? (follow up question)
2. What is your relationship with this person/group?
4. Where do you see yourself going throughout or after this project?
5. I noticed that in your team, this happened, can you explain it?

(**Transition:** Well, it has been a pleasure finding out more about you. Let me briefly summarize what I have recorded during our interview.)

III Closing

A. (Summarize)
B (Maintain Rapport) I appreciate the time you took for this interview. Is there anything else you think would be helpful for me to know?

Sample Interview Data

Transcript from an online interview with ReMind Director Jeff on 25/07/2022

Eoin
I suppose this is sort of, like commensurate with the, what you had also talked about, which is that it's a sort of engineering approach to, I mean, I suppose like most app design, or app designers would have some sort of like an engineering type approach. But it's an engineering approach to, I suppose, to treatment.

Jeff
It's a problem solving approach, I guess, if you call engineering as iterative problem solving, I think all science is in a way. So you start with a hypothesis, you run it, you see whether it's working or not, then you change, create another hypothesis, then you run it. So I would call it a scientific approach. I think clinicians do that, when they're treating somebody as well. Let's start with a hypothesis, see whether it works. So we just did it at a massive scale. So you would run a hypothesis on a million people and find out that it didn't work for that 10,000. And then change something for that. 10,000, personalise it, and so on, so forth.

Eoin
Yeah, it's funny, I was just thinking there is that, you know, a problem solving approach to how to detect the type of treatment. But the treatment itself isn't the problem. It's not a problem solving, style of treatment. It's an active listening style.

Jeff
That's interesting. That's true. So we use cognitive behavioural therapy, or DBT, or ACT, sort of, or mindfulness elements from those, the problem solving aspect was more of the problem of, so if I would say the three problems that ReMind solves for, is one, how do you get somebody to open up to how do you get them to reflect? And really get to the core of what their issues? And three, how do you get them to follow an evidence based technique without needing a huge amount of, you know, psychoeducation, big videos, long lessons, and so on. So how do you achieve all of these in a 15, minute, 10 to 15 minute session, those are the problems we're solving. We're not solving the specific problem that the user is dealing with. So if you get 10,000 people dropping off on something, they then you solve the problem of how do I keep them and actually get them to do the exercise I asked them to do, 10,000 people refuse to do that exercise, you start looking at what's the common characteristic of these 10,000 people? And then you realise they all have a certain common thing. And maybe for people like that this type of exercise doesn't work. So maybe we need to position it a little differently, then you go back to clinicians, and you say, is the exercise wrong? Or is the motivational track wrong? And then they will come up with two or three hypotheses as to either of those, and you'll see which one works better.

Eoin
Would you see the bot and the CBT are just not all the tools that the app provides? as well? Like, how would you kind of characterise there? It is two quite different things. How would you characterise their relationship between the board and the app itself, the sort of tools, the, the kind of range of like, you know, the sleep and the CBT tools.

Jeff
So the idea is to provide something that is your personalised companion, there's only so much conversation you can do. There's so much conversation we can write. And what tends to happen is that somebody comes in, they'll have a short conversation. And then, you know, there's not a lot of depth if you keep going into ReMind and keep trying to play with it and start getting repetitive. So because of that, we try to create a lot more depth in audio and video content, so that people can keep browsing over 150 different resources, or even start specifically different conversations. Because when you think of the bot as the freewheeling bot where you say I'm feeling this way most people will trigger similar patterns over and over again, because they're dealing with the same issue. Because they're feeling the same emotion. They're not going to go to a conversation about assessing energy, or a conversation about (inaudible) or all the other conversations that ReMind has…So the app has direct ways to access all the rest of it, where you want to learn, you want to play, you want to see what else you can do, without necessarily it being that one thing that got triggered by how you feel. Because if you come back and you say, I'm still feeling anxious, it's still got, you know, five resources to throw at you in some mix.

Eoin
Yeah, sure. But I mean, that sort of sounds to me like that the bot is just a kind of a gateway to the tools. But from what I've gathered, the bot is much more important than that.

Jeff
Yeah, it's not a gateway to the tools. It's your coach. Yeah, so. So the bot is the coach, right? It guides you through the tools. A lot of the tools are conversational, are delivered by the bot. There are also other tools which are audio visual, or, you know, guided meditations and the like. But a lot of the tools just wouldn't work without a bot. It's not just the bot telling you, "Hey, do CBT", the bot is actually talking you through, or CPAD is a classic bot where you just go in and the bot says “Just say whatever comes in your head”, keep saying it, they will analyse it and categorise it, and it will help you sort your head out. And that's the entire tool is about. It's not a gateway to the bot. So it's not a gateway.

Appendix B

Information and Consent Forms

Information Sheet

Department of Psychosocial Studies
Birkbeck, University of London
Malet Street,
London WC1E 7HX
020 7631 6000

Supervisor: Dr. Silvia Posocco
Department of Psychosocial Studies
Birkbeck, University of London
London WC1B 5DT
Email: s.posocco@bbk.ac.uk

Supervisor: Prof Stephen Frosh
Department of Psychosocial Studies
Birkbeck, University of London
London WC1B 5DT
Email: s.frosh@bbk.ac.uk

Researcher: Eoin Fullam
eoinlfullam@gmail.com

Title of Study: The Social Life of Mental Health Chatbots

Eoin Fullam

The study is being done as part of my PhD in the Department of Psychosocial Studies Birkbeck, University of London. The study has received ethical approval.

This study will observe the development culture in ReMind with the aim of analysing how this culture influences the design of computerised mental health treatment. It is part of a larger project which seeks to explore the social conditions which have given rise to technologically automated mental health treatment in general, and therapy chatbots in particular.

I will act as a participant observer in this software firm to study the working culture from within. This means that I will be included in some way as part of the working environment.

You will be asked to take part in three to six interviews over the course of six months, where we will discuss your engagement with the working culture, thoughts and feelings about computerised mental health treatment, and the links between these.

Data will be analysed by me: I will record the interviews and transcribe them. The transcriptions will be fully anonymized and then analysed by me.

If you agree to participate in interviews, you will agree on a convenient time and place for me to interview you for about an hour. You are free to stop the interview at any time. You can withdraw from the study up until the data you have provided has been anonymised and aggregated into a larger dataset from which it is impossible to extract, which is a process that takes approximately 4 weeks.

Your data will be kept by me and will be stored using Birkbeck University's secure online storage until it has been anonymised and aggregated into a larger dataset.

The analysis of your participation in this study will be written up in a report of the study for my degree. You will not be identifiable in the write up or any publication which might ensue.

The anonymised dataset (which your data is a part of) will be made available to other researchers by a published thesis and in Birkbeck's institutional data repository, BIRD .

The study is supervised by Dr. Silvia Posocco & Prof. Stephen Frosh who may be contacted at the above address and telephone number.

For information about Birkbeck's data protection policies, please visit: http://www.bbk.ac.uk/about-us/policies/privacy

If you have concerns about this study, please contact the School's Ethics Officer sshpethics@bbk.ac.uk

You also have the right to submit a complaint to the Information Commissioner's Office https://ico.org.uk/

Consent form

Title of Study: The Social Life of Mental Health Chatbots

Eoin Fullam

I have been informed about the nature of this study and willingly consent to take part in it.

I agree to the following data collection and processing approaches being used for my data: Interviews will be recorded (audio only) and transcribed. Anonymised transcriptions will then be aggregated into a larger dataset and uploaded to Birkbeck University's online repository. The original recording will be destroyed once it has been transcribed and anonymized.

I understand that I will not be identifiable in any presentation of this research without my further, written, consent.

I understand that I may withdraw my data at any time before it has been anonymised and combined with other data.

I understand that the anonymised form of the data I have provided will be made available to other researchers through publications and by being deposited in our data repository.

I am over 16 years of age.

Name

_____

Signed

_____

Date

_____

There should be two signed copies, one for participant, one for researcher.