Conversational agents and the making of mental health recovery

Robert Meadows ⓘD, Christine Hine ⓘD and Eleanor Suddaby

Abstract

Background: Artificial intelligence (AI) is said to be "transforming mental health". AI-based technologies and technique are now considered to have uses in almost every domain of mental health care: including decision-making, assessment and healthcare management. What remains underexplored is whether/how mental health *recovery* is situated within these discussions and practices.

Method: Taking conversational agents as our point of departure, we explore the ways official online materials explain and make sense of chatbots, their imagined functionality and value for (potential) users. We focus on three chatbots for mental health: Woebot, Wysa and Tess.

Findings: "Recovery" is largely missing as an overt focus across materials. However, analysis does reveal themes that speak to the struggles over practice, expertise and evidence that the concept of recovery articulates. We discuss these under the headings "troubled clinical responsibility", "extended virtue of (technological) self-care" and "altered ontologies and psychopathologies of time".

Conclusions: Ultimately, we argue that alongside more traditional forms of recovery, chatbots may be shaped by, and shaping, an increasingly individualised form of a "personal recovery imperative".

Keywords

Chatbots, mental health, recovery, artificial intelligence, expertise

Submission date: 20 February 2020; Acceptance date: 22 September 2020

Introduction

"Recovery" has been iterated as the new "paradigm in mental health policy and practice."[1] In "No health without mental health: A cross-government mental health outcomes strategy for people of all ages'[2] the coalition government set six mental health objectives, one of which was that more people with mental health problems will recover. This recovery-guided approach is continued in "Closing the gap: Priorities for essential change in mental health",[3] where a commitment is made to commission services with an emphasis on recovery. More recently the "Five year forward view for Mental Health'[4] document was published that stresses that commissioners should prioritise early intervention, choice/personalisation and recovery.

Yet, despite its increased visibility and importance within the mental health arena, recovery remains a vague, "polyvalent concept'[5] that defies definitional consensus. Within the clinical literature, recovery is often situated as the amelioration of symptoms so that a person can resume activities within what is considered a normal range. A second form of recovery has its origins in the Independent Living and Civil Rights Movement of the 1960s and 1970s. This sense of

Department of Sociology, University of Surrey, Guildford, Surrey, UK

Corresponding author:
Robert Meadows, Department of Sociology, University of Surrey, Guildford, Surrey GU2 7XH, UK.
Email: R.Meadows@surrey.ac.uk
Twitter: @RobertMeadows16

recovery "does not require remission of symptoms or other deficits, nor does it constitute a return to normal functioning. Rather, it views mental illness as only one aspect of an otherwise whole person".[6,7] Whilst the multiple meanings of recovery are often situated as dualisms or either/ors, Pilgrim and McCranie[8] suggests that there are four "different shades" of the meanings of recovery: (i) recovery as a personal journey; (ii) recovery as a critique of services that emphasises choice, empowerment, reform; (iii) recovery as therapeutic optimism; and (iv) recovery and the social model of disability. Whilst there is a conceptual separation in their emergence from different social groups there is a practical co-presence in their everyday application that can create "working misunderstandings".[8]

Recovery is also (bio)political. Firstly, dominant discourses of recovery are said to individualise what are social problems.[9] There has been a neoliberal intrusion on the word and it has been taken over by marketisation, language, techniques and outcomes.[10] Within this neoliberal, individualising arena recovery has become the "struggle for recognition" and "it is only when the collective, structural experiences of inequality and injustice are explicitly linked to processes of emotional distress that recovery will be possible."[9]

Secondly, "definitional debates about recovery reflect wider ideological debates about the nature of mental health."[9] Policy documents talk about recovery within the context of the "global burden" of mental illness[7,11] whilst often sidestepping controversy surrounding diagnosis in psychiatry. The *Campaign for the Abolition of Schizophrenia Label*, for example, suggests that "the idea that schizophrenia can be viewed as a specific, genetically determined, biologically driven, brain disease has been based on bad science and social control since its inception".[12] As Lasalvia and Ruggeri (2018) note, "even within the psychiatric field there have been calls to change the term schizophrenia as not scientifically valid and unhelpfully focusing on a biological explanation of what is a heterogeneous and context-influenced disorder".[13]

In short, recovery remains a vague and political concept that sits at the core of current mental health policy despite significant critical voices. In this context, it is timely to explore whether, and in what forms, recovery is situated within conversational agents (chatbots) - an emerging technology viewed as having potential to transform mental health treatment. The Topol Review on the digital future of mental healthcare and its workforce concluded that there is already a strong evidence base for these kinds of interventions and that by 2021 chatbot systems could be offering advanced automated or semi-automated diagnostic and therapeutic tools.[14] However, whilst considerable research has been carried out on the efficacy, feasibility and ease of use of mental health chatbots,[15] very little attention is being paid to how chatbots relate to ideas of mental health recovery. We do not know, for example, the extent to which chatbots are designed to work with – and encourage – particular versions of recovery. The paper therefore asks: to what extent are chatbots shaped by, and shaping, the concept of mental health recovery?

We begin with a critical review of the existing literature on chatbots and mental health. Exploring how chatbots shape, and are shaped by, the concept of mental health recovery requires us to look both at overt reference to recovery as well as possible sites of socio-political struggles that may come to conceptualize recovery.[1] To operationalise this, we derive analytic directives by bringing contradiction, contention and debates within the literature on chatbots into dialogue with the existing literature on the meanings of mental health recovery. We then move to outline the methodological and analytical focus. We interrogate a range of online sources to explore how chatbots are being positioned and the claims being made of them.[16–18] Analysis starts with the foreshadowed ideas from the critical literature on recovery and asks: how do chatbots relate to recovery? How are chatbots positioned in terms of expertise? How do chatbots relate to time? How do chatbots relate to everyday experience and clinical subjectivities? We conclude with a discussion of how recovery may be being troubled by chatbots and the implications for the future of mental health recovery.

Chatbots and mental health recovery: A view from the literature

Chatbots - "systems that are able to converse and interact with human users using spoken, written, and visual languages"[19] – are increasingly being used in the field of mental health. A scoping review conducted by Abdalrazaq et al. (2020) identified 53 studies that involved 41 different chatbots. In 17 studies, chatbots were used for therapeutic purposes and of these 10 were based on cognitive behavioural therapy and just under half were focused on people with depression and anxiety.[19] *Woebot*, for example, treats "depression and anxiety using digital version of time-tested cognitive behavior therapy",[20] whilst *#BeALifeline Direct Message (DM) Chatbot* aims to help start conversations about suicide and self-harm.[21] They are also being tailored to populations, with *HARR-E* being launched to tackle the male mental health problem and *Wysa* being originally tested and co-designed with an adolescent user group. As well as therapy, chatbots also aim to provide training, improve social skills and act as a screening tool for

such things as depression, dementia and posttraumatic stress disorder.[19]

Recovery has an ambiguous presence within this literature. Whilst there is minimal overt mention of recovery, there is a clear focus on efficacy and symptom reduction with Fulmer et al.,[22] for example, examining scores on the Patient Health Questionnaire (PHQ-9), Generalized Anxiety Disorder Scale (GAD-7), and Positive and Negative Affect Scale (PANAS) at baseline and 2 to 4 weeks later. Other studies investigate differences in effect index across groups and time,[23] or report on the development of a chatbot to deliver support for caregiving professionals; also, as measured by the PHQ-9 and GAD-7.[24] Recovery, then, appears to be tacitly situated as a (clinical) outcome and the amelioration of symptoms.

With that said, there are areas of contradiction, contention and debate within the literature that point towards emerging social, political and technological tensions around the conceptualization of recovery. Several of these are made apparent in the following clinical abstract:

"A World Health Organization 2017 report stated that major depression affects almost 5% of the human population. Major depression is associated with impaired psychosocial functioning and reduced quality of life. Challenges such as shortage of mental health personnel, long waiting times, perceived stigma, and lower government spends pose barriers to the alleviation of mental health problems. Face-to-face psychotherapy alone provides only point-in-time support and cannot scale quickly enough to address this growing global public health challenge. Artificial intelligence (AI)-enabled, empathetic, and evidence-driven conversational mobile app technologies could play an active role in filling this gap by increasing adoption and enabling reach. Although such a technology can help manage these barriers, they should never replace time with a health care professional for more severe mental health problems. However, app technologies could act as a supplementary or intermediate support system. Mobile mental well-being apps need to uphold privacy and foster both short- and long-term positive outcomes."[25]

When brought into dialogue with the wider literature, the above draws attention to the overlap between (i) struggles over expertise; (ii) the role of AI technologies; and (iii) the appropriate temporal scale to use. Debates surrounding notions of expertise are common across variants of recovery[26] and this is perhaps best exemplified by current controversies surrounding the "recovery star". The recovery star contains ten areas covering aspects of an individual's life: such as social networks, skills, identity and self-esteem, addictive behaviour and relationships. Service users set their goal for each one (on a scale of 1 to 10) and then progress is monitored over time. In opposition to this "redundant, unhelpful and blunt tool" the user-led group "Recovery in the bin" developed the UnRecovery star to reflect a non-individualised notion of mental health that acknowledges structural and contextual issues. Dimensions seen as important for the UnRecovery star include loss of welfare state, racism, unstable housing, sexism, economic inequality and discrimination. Despite an emphasis within policy on the importance of experts by experience,[27] service users often feel overlooked in discussions of recovery.

Struggles over expertise manifest in the literature where conversational agents are situated as offering low-cost, on-demand, highly adaptable, self-help which will help address a growing global challenge. However, at one and the same time, it is made clear that conversational agents should supplement rather than replace traditional therapeutic options. In many respects this is a variant on an old theme. As telepsychiatry developed it was suggested that this could increase access.[17] However, clinicians were also presented as being resistant to its advancement because of the deleterious effect on the doctor-patient relationship deemed central to therapy. For example, and as Pickersgill notes,[17] in his 1952 address as President of the American Psychiatric Association, Leo Bartemeier expressed the need to be cautious about the rise of technological approaches to clinical practice and alert to the possible impact on clinical and therapeutic interpersonal relationships. The dual emphasis on reach and tradition creates complex biomedical virtues – which "refers to the (profession-defined) praxis of goodness within the laboratory and the clinic"[17] – where telepsychiatry and direct clinical care need to be carefully situated as extending clinical care rather than replacing it. Pickersgill[17] uses the example of Chill Panda, an app that captures biometric data – such as heart rate and blood flow – and suggests playful tasks that suit your current state of mind. Users are encouraged to learn how to manage their stress and feel better. However, this is offered with provisos and "clinical expertise thus remains salient, even as (potential) patients are encouraged to contribute to their care".[17]

Within the literature on chatbots, these tensions between tradition and reach intersect with ambiguities surrounding artificial intelligence. Vaidyam et al.[15] define chatbots as "digital tools existing either as hardware (such as an Amazon Echo running the Alexa digital assistant software) or software (such as Google Assistant running on Android devices or Siri running on Apple devices) that use machine learning and artificial intelligence methods to mimic humanlike behaviours and provide a task-oriented framework with

evolving dialogue able to participate in conversation". Seen in this way, AI becomes an essential part of chatbots and conversational agents are situated as part of the general AI revolution said to be "transforming mental health".[28] The Alan Turing Institute is also attempting to drive forward research in to AI-based precision mental health. There is also a move towards a language of "prediction". A 2018 paper[30] for example, suggested that AI could be used to identify how well people at risk of psychosis will function in the future. The news forum "Scimex" reported this under the banner – "AI could be used to predict mental health recovery". Yet, the majority of the chatbots identified in a review (92.5%) depended on decision trees or pre-defined rules to generate their responses. Only 7.5% used machine learning approaches.[19] The nature of artificial intelligence within the chatbots remains ambiguous and it is therefore unclear whether struggles over expertise involve users gaining more control and autonomy, or the chatbot itself becoming the dominant agent: further individualising "treatment" and ultimately alienating both service users and clinicians from it.

There are also temporal ambiguities surrounding the ways in which AI is discussed; including a mix up of tenses between what AI could do and what AI actually does presently. This temporal ambiguity complicates the role being claimed for AI in recovery whilst also complicating what recovery is. Temporality has long been central to discussion of recovery. For example, literature framed by the biomedical/disease model suggests that disturbances in the experience of time are a commonly reported feature of mental disorders. Also working within a biomedical/disease model, Eugene Minkowski, noted that with those diagnosed with schizophrenia "experienced time is altered in its flow, being experienced as frozen, immobilized, without "elon vital". Experience of time is also affected by specialization – that is time is felt as divided in juxtaposed elements – and temporal fragmentation: "With the fracturing of time flow, we observe an itemization of now-movements in consciousness so that each now-movement in a person's stream of consciousness will be experienced as detached from the previous one and from the following, hence as extraneous to one's stream of consciousness and sense of selfhood."[31] Time continued to be a methodological tool and pathological focus of clinical psychiatry and psychology throughout the first half of the 20th century – and Mnkowski played a key role in shaping subsequent understandings of abnormal time and depression and schizophrenia.[32]

Amongst other things, this literature also notes a "subjective decrease in the experienced velocity of the flow of time in depressive disorders"[29] and that "depressed patients" showed "general changes in their experience of the passage of time. The most common was that of time having slowed down."[29] Perhaps in contradiction, Vogel et al.[29] also found some indication of an increase in the experienced velocity of the passage of time. As noted further below, it therefore becomes important to consider how chatbots frame and portray the experience of time. Words like relapse, recurrence, deterioration, periods of improvement remain common within biomedical discussions of depression. As Fava and Visani[33] suggest, for example: "clinicians working with depressed patients are often confronted with the unsatisfactory degree of remission that current therapeutic strategies yield, and with the vexing problems of relapse and recurrence". Yet, as illustrated in the abstract above, the literature on chatbots makes redress to "short and long term" positive outcomes. In doing so, it appears to move us away from expressing and experiencing recovery through a language of relapse, recurrence and deterioration. Chronicity remains in uncertain ways and the language of "outcomes" sidesteps, rather than resolves, tensions that inhabit notions of recovery.

Research questions

In essence, whilst the academic literature is clear in situating chatbots as an emerging hope for treatment, therapy and training there is both (i) little overt mention of recovery and where it does occur it appears rendered as "outcome"; and (ii) potential for significant impacts on understandings of recovery, particularly in connection with expertise, technology and temporality. Within this paper we build on this by looking in more depth at the sociotechnical assemblage of the chatbot. Technologies are sociotechnical objects that have been co-produced in a particular historical and cultural context.[34,35] Values and politics are incorporated into the design of technologies and they are likely coded with ideologies that reflect and refract contemporary knowledge claims to mental health and recovery. At one and the same time, "technologies have both *makers* and *users*; the expectations of innovators, the experience of users and the adaptation of technology to meet their needs, wants and desires in everyday life all influence the cultural spaces that the technology eventually comes to occupy."[18] Within these cultural spaces, the meaning of chatbots may be inextricably bound up with the meanings of recovery (and vice versa).

Reflecting the discussion above, the paper asks: how is recovery situated by chatbots? What models are being used and what models are excluded? How are chatbot users situated in terms of "need"? How are expertise and the capacities of the technology framed and discussed? How is temporality situated?

Methods

This paper focuses on the ways official online materials explain and make sense of chatbots, their imagined functionality and value for (potential) users.[16] Our interest is in how chatbots are positioned and the claims being made for them, rather than the experience of using the chatbot in itself. As apps are interpreted by surrounding materials that give them meaning,[36] empirical data reported on in this paper is comprised of publicly available online materials. Materials are sourced from official company pages and include blog posts, promotional videos, marketing materials and user guides.

As noted, chatbots have been developed for a range of tasks including therapy, training, screening and improving social skills. It has also been proposed that voice signal extraction can be used to detect mental illness.[37] Within the present study, we adopted a purposive, homogeneous, sampling strategy[38] and focused on those offering forms of therapy. Sampling proceeded as follows: First, a decision was made to explore those apps which are situated as offering assistance for depression/anxiety. This is because reviews suggest that this is the most common focus of chatbots.[19] Second, three chatbots for mental health were selected for analysis: Woebot, Wysa and Tess. These are three of the most popular chatbots; for example, appearing at the top of Clarke's[39] list of "AI Bots and Apps for Depression". Woebot launched in the Summer of 2017 and is "designed to offer convenient care to those struggling with depression by mimicking human conversation, offering self-help related guidance and companionship to its users". Woebot uses chat-based functions to mimic online human interactions. It is built around principles from Cognitive Behavioural Therapy (CBT) and uses natural language processing to "get to know" users over time and "more accurately detect and meet your emotional needs at a given time, offering personalized resources, self-help guidance, information and support related to your concerns". It won the google play award for "standout wellbeing app" in 2019. Wysa is reported to have around 1,500,000 users. Like Woebot, both Wysa and Tess incorporate principles of CBT. However, Wysa is also built around dialectical behavioural therapy, meditation practices and motivational interviewing. Tess moves away from app based approaches and uses text-based messaging to track user goals and provide guidance and interventions.[39]

These chatbots are developed outside of mainstream mental health services and can be expected to be driven, to some extent, by commercial interests and the desires of potential customers, including mass healthcare purchasers such as in the UK the National Health Service. Yet, at one and the same, whilst it might be assumed that chatbots will draw from a biomedical framework because of this, developers also stress the important role that users play in their evolution. For example, the Wysa website states that "Over 60 psychologists and 15,000 users have provided specific inputs to shape how Wysa helps them. 65 users have volunteered to help us translate Wysa into their language." Materials analysed are thus formulated within a complex space that includes developers, psychologists, "users", consumer concerns and healthcare commissioning rules. Materials were systematically collected. For each chatbot, we began with the official company webpage. These pages were downloaded, along with anything that was directly linked. For example, the Woebot website lists two scientific references on its "how it works page". These were also downloaded and analysed. The Tess website also includes a series of references as well as embedded videos from both clinicians and users. As Berg[16] notes in his analysis of self-tracking devices, "Since all of these materials are part of the companies" marketing strategies and social media presence, they are regarded as not only official presentations of the products, but also as naturally occurring empirical materials." Online Appendix 1 lists the sources analysed and accompanying urls. All sources were analysed between August 2019 and February 2020. Links are accurate as of then.

Analysis proceeded in three steps. First, materials were examined for explicit reference to "recovery". Where the word recovery was present, we reflected on the context and framing and the discursive resources being deployed. Second, a wider reading was conducted foreshadowed by the ideas and questions from the critical literature review and materials were read sensitized to possible struggles over expertise, technology and temporality that implicitly invoked recovery. Finally, we drew these two sets of observations into the set of cross-cutting themes presented below. This discussion is not an exhaustive set of findings but focuses on sets of tensions we consider particularly pertinent to the development of future notions of recovery.

Findings

Across Woebot, Wysa and Tess there is largely an absence of overt reference to "recovery". Exceptions include the statement on the Tess website that "92% of people moved towards recovery" (Tess_2) and Woebot's claim that "struggle is actually a necessary part of recovery" (Woebot_18). There is no further elaboration on the statement and it is instructive that it appears on a page titled "outcomes". Across the three chatbots, outcomes dominate as an idea. The "science" page of the Woebot website, for example, has very little information beyond a graph showing

"reduction in depressive symptoms" following use of Woebot (Wobebot_2) Readers can follow a link to a paper by Fitzpatrick et al.[40] which reports that "Intent-to-treat univariate analysis of covariance revealed a significant group difference on depression such that those in the Woebot group significantly reduced their symptoms of depression over the study period as measured by the PHQ-9 (F=6.47; P=.01) while those in the information control group did not". Somewhat similarly, Wysa's webpage points the user to a paper by Inkster et al.[25] that evaluates the Wysa app using ideas of impact and engagement. Impact was measured by capturing the average improvement (pre-PHQ-9 minus post-PHQ-9) between a group of high and low users (Wysa_2 & Wysa_6)

There is a prima facie case, then, for suggesting the conversational agents are invoking an idea of recovery that involves smooth and linear reduction in symptoms over time and locates the agency for bringing about that recovery in the app. Debates surrounding diagnosis are made redundant by claims that the chatbots are "agnostic when it comes to diagnosis" (Woebot_18) and very little critical attention is paid to who is deciding which outcomes are core or to the standard questionnaires and measurement devices employed in the clinic to capture these outcomes. Screening tools arguably *constitute* the phenomena they purport to track rather than simply reflect them.[41] – especially through techniques such as "reduction", "expression", "quantification" and "normalisation".

Further analysis does suggest complex and competing ideas which coexist with this narrow view of recovery. In making these apparent, we firstly discuss three themes that map on to the struggles over practice, expertise and evidence[1] foreshadowed by the literature review: "troubled clinical responsibility", "extended virtue of (technological) self-care" and "altered ontologies and psychopathologies of time". Within the concluding discussion we explicitly connect these back to recovery.

Troubled clinical responsibility

A video on the Tess site focuses on Lloyd Werk MD, MPH; a person deliberately identified through their professional status and medical qualifications (Tess_9). Dr Werk outlines how the clinical encounter usually involves face-to-face meetings to engage patients and identify goals. Following this a coach may meet with them to monitor progress. The problem with this model, according to Dr Werk, is the wait between meetings and an extensive wait list. Further to this, it misses much of the therapeutic work that happens at home. Dr Werk emphasises the traditional nature of the therapy by noting how Tess enables the patient to stay in touch and hold many of the same conversations that a coach and therapist would have with the client. Dr Werk notes how he gets the analytics for his patients and can reflect on frequency and type of conversation at the next face-to-face consultation. Somewhat similarly, Woebot emphasises reach and enhancement of access when it stresses that they are "on a mission to make high-quality tools for mental health radically accessible to everyone" (Woebot_3). Woebot also claims it "will never replace therapy or therapists, and it's not trying to do" but instead they "will exist side-by-side" (Woebot_15).

Reflecting Pickersgill's[17] discussions of biomedical virtue and clinical responsibility, these materials situate chatbots as extending clinical care but not replacing it. Traditional forms of clinical responsibility are upheld, even where the relationship between therapy and technologies operates in complex ways. For example, Wysa documentation makes clear links to AI and outlines how the chatbot was designed by a wide group of people – which includes therapists, coaches, users and AI folk. In doing so, Wysa also appears to be situated as individual, personal, and different from traditional forms of therapy. It is about your own *private* reflective space for when *you* need to get your head straight. Whilst this may appear to trouble the dominance of traditional models of therapy, pages contain disclaimers that "Wysa cannot and will not offer medical advice" and the following quote from an academic review can be found under the idea of "safety": "It is worth noting that Wysa offers the option of getting support in a more traditional sense in conjunction with or as an alternative to the automated service. Its users can receive support from real-life mental health professional ('Wysa Coach') for a fee. This can prove particularly helpful in contexts where there is a lack of trained mental health professionals (see Wysa_ 2; Wysa_3)." In this formulation, it is ultimately traditional therapy that protects from risk. The juxtaposition of Wysa with "tradition" and ethical discourse surrounding the use of the automated service, also reflects processes of "performative nominalism", which have helped to reify mHealth as an object of (cautious) fascination.[17]

Technology is also discussed in a way that reaffirms the salience of the clinician and clinical encounter. The official Woebot Facebook page hashtags "AI" in numerous posts (Woebot_22). Technical magazines describe Woebot as "An AI-based therapy chatbot to help with depression and anxiety"[42] and Andrew Ng announced his appointment to the Woebot board of directors under the title "Woebot: AI for mental health".[43] Woebot also announced on the 26th September that it had been named a finalist in the category of "Best application of AI". Yet, on the official

pages little note is made of the AI behind Woebot (although see Woebot_19). Perhaps reflecting the historical concern that expert level cognition is distinctly human[44,45] there is also the claim that "chatbots are not about AI, they're about a more human interface" (Woebot_12). Altering the way in which the technology is framed ensures that the main "agents" of expertise included in the Woebot materials are clinicians and the user remains absent as an agent of expertise.

This is not to say that traditional forms of clinical responsibility remain entirely untroubled. For example, across the materials analysed, Woebot is situated as being different from traditional therapy. In an article linked to the company twitter feed (Woebot_22), Woebot is described as "fundamentally different from any other form of therapy." "The Woebot experience doesn't map onto what we know to be a human-to-computer relationship, and it doesn't map onto what we know to be a human-to-human relationship either . . . "It seems to be something in the middle.[46] At one and the same time, discursive resources are deployed to position Woebot as the same as traditional therapy. Official pages state that "the popular opinion about therapy is that it holds a kind of special magic that can only be delivered by individuals who are highly trained in this mysterious art form" (Woebot_15). Instead "the truth is that modern approaches to mental health revolve around practical information gathering and problem solving" (Woebot_15)– something which, although not explicitly stated, computers are renowned for. Woebot's traditional forms of expertise are continually championed, with other pages stressing that the creators are "psychologists who built programs in the clinic" and "worked at Stanford for over 10 years" (Woebot_2). They ask, "Don't you need a therapist?" as a response they claim that there are "20+ years of rigorous research to show that DIY CBT works" and that CBT delivered via the internet (and even video games) can be as effective as therapist-delivered CBT for both anxiety and depression (Woebot_2). At the same time as being framed as both "different" from, and the "same as" traditional therapy, official pages further appear to position Woebot as "better" than traditional therapy. For example, the landing page of the company website links to an article in Wired magazine: "Woebot's creators believe it has the potential to actually *improve* on human therapists. "It's almost borderline illegal to say this in my profession, but there's a lot of noise in human relationships," says Alison Darcy, one of the psychologists behind Woebot, and the company's CEO. "Noise is the fear of being judged. That's what stigma really is." There's nothing like venting to an anonymous algorithm to lift that fear of judgement"[47](Woebot_27)

Perhaps ironically, the most significant threat to traditional virtues of clinical responsibility is a continued, uncritical, emphasis on Cognitive Behavioural Therapy. Cognitive Behavioural Therapy – (CBT) a dominant mode of treatment used in contemporary mental health practice[48] – is almost hegemonic in its status: "our Enlightenment heritage calls for a rationalist ordering of the therapies, in accordance with narrow and pre-constructed values that correspond with those of society's most powerful institutions. It is in this context that we should understand CBT's overwhelming emergence as the therapy of choice. The risk of its institutional success is the establishment and legitimisation of a therapeutic hegemony, and the gradual diminishment of a once rich landscape of therapeutic possibilities".[49] In stressing the idea that CBT delivered via the internet (and even video games) can be as effective as therapist-delivered CBT, Woebot plays down the innovative nature of chatbots and instead inserts continuity with the established authority of CBT. In doing so, it problematizes the "expertise" of any human actor whilst implicitly borrowing CBT's embedded view of recovery.

Extended virtue of (technological) self-care

The troubling of clinical responsibility – or not – operates alongside an extended virtue of self-care. As we argue in this section, this becomes most apparent when we examine the ways in which the clinical and everyday are situated. Woebot, Wysa and Tess normalise the need for conversational agents as everyday objects. For Wysa, "sometimes we all get stuck inside our heads" (Wysa_1). For Woebot, "Everybody could use someone like me" (Woebot_18); whereas Tess states "When you say something in a certain way, a good friend will know how you actually feel, it's the same thing with Tess" (Tess_16/Tess_14) and "When I feel stuck, Tess helps me look at the situation in a different way" (Tess_14).

Attempts at everyday normalisation can be seen elsewhere in the materials analysed. For example, Tess is described as an "entirely new form of care" which is affordable and scalable (Tess_10). The site states that most people prefer Chat with Tess over "traditional therapy" (Tess_13). Under the "for individuals" page, the Tess website contains a short video testimonial from a female user (Tess_14). The user narrates how she was looking for an app that would help her manage stress and build self-awareness. Tess is compared to a Tamagotchi toy from her childhood. 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."[50] 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 seen in the literature review above. It also suggests a different orientation to recovery to the one invoked by reference to CBT. Ideas of recovery are almost side-lined as the focus remains on the present and chatbots become the care that we all need to deal with our everyday emotions.

Altered ontologies and psychopathologies of time

As noted above, clinical discourses are included across the materials. Woebot, Wysa and Tess, for example, highlight research on outcomes such as PHQ-9 Scores and depressive symptoms. However, these clinical discourses are largely left to exist in silos to the everyday normalisation also noted above. This leaves the ontology of mental illness troubled. Labels such as anxiety and depression become ambiguous and simultaneously situated – cf. Szasz's myth of mental illness[51] – as general, pervasive, problems of living, and as "real" clinical concerns. This ontological troubling is accompanied by ambiguity surrounding medicalisation. For example, when the user in the Tess website references bereavement, they call upon something that for many years was used as an exclusion in the diagnosis of major depression.[52] When DSM-5 removed this exclusion, critics were concerned that this would "medicalize" ordinary grief and encourage over-prescription of antidepressants.[52] At the same time as troubling the ontology of mental illness, then, chatbots have the potential to extend its medicalisation where it situates mental illness as a "real", clinical concern, but one that is part of an epidemic and general modern malaise.

Temporality plays a key role in these antagonisms. The materials analysed contain little overt mention of relapse, recurrence, deterioration or improvement. There is also no clear statement as to when one might consider stopping using the chatbot. Within the user video embedded within the Tess website, the main protagonist discusses how she does not visit the chatbot as much now, but it is good to know that it is still there when/if she does. This narrative of continued use illustrates further the extension of the virtue of self-care. Where temporality does appear, it is because others might not be awake when you need them.

"On an individual level, however, there are so many reasons why people find it hard to reach out. We often say that when you are feeling low, "you should talk to someone". But insisting that this is the only way to get help leaves behind all of those for whom that is not an option. What if it's 3am?" [Woebot_15]

"Wysa is your 4 am friend" [Wysa_1]

"patients seem engaged and even initiate chats 24/7" [Tess_9]

In other situations, being awake at 3am or 4am would be the target of treatment. For example, the biomedical literature reports a link between sleep and depression with about three quarters of those experiencing depression also experiencing insomnia symptoms. Hypersomnia is also present in around 40% of young adults given a diagnosis of depression and 10% of older adults. Anomalies in sleep architecture are also linked to treatment outcomes: for example, predicting worse response to CBT.[53] Yet, within the materials analysed the problem is not that that the user is awake at 4am. Advice is not offered on circadian rhythms, warnings are not given on using blue light-based devices at 4am in the morning. There is no suggestion that you might want to visit a GP. The pathology is simply that no-one else is awake at the time. In this instance, being awake is normalised and the arrhythmic state becomes the fact that the rest of the world is sleeping.

Temporality creates its own tensions across the three chatbots. On the one hand, there is an emphasis on outcomes with Tess, for example, making clear that chatting with Tess leads to reduced symptoms of depression (–28%) and anxiety (–18%). Yet, ideas of temporality are most prominent away from these discussions of outcomes and are focused more on the timing of delivery rather than anticipating the upshot of the intervention. As McWade,[1] notes, a key aspect of the socio-political struggle which surrounds recovery is temporal and within the section below we turn all this back to the overall research question and ask: to what extent are chatbots shaped by, and shaping, the concept of mental health recovery?

Concluding thoughts – What then of recovery?

Following Williams et al.[18] firm conclusions may be unwise, given the embryonic nature of these developments. However, a number of tentative conclusions may be drawn. For one, multiple versions of recovery appear to be approximated within the materials. On the one hand, there appears to be a discursive move towards a recovery *from* model. As Davison and Roe note,

"Recovery from serious mental illnesses involves the amelioration of symptoms and the person's returning to a healthy state following onset of the illness. This definition is based on explicit criteria of levels of signs, symptoms, and deficits associated with the illness and identifies a point at which remission may be said to have occurred. This definition thus has many advantages from clinical and research perspectives, as it is clear, reliable, and relatively easy to define, measure, and link to dysfunctions or wellbeing in other areas of life. People who enjoy this sense of full recovery could be considered to have recovered from psychosis in the same way that other people may recover from an infection, a broken leg, or, in the case of recovery over a longer period of time, asthma".[26]

Recovery *from* is overtly referenced using a repertoire of "outcomes" and "reduction in symptoms". The use of "science" pages, and redress to clinical studies/scientific method also foregrounds recovery *from* as dominant and "expert". Whilst the dominance of this biomedical frame is perhaps to be expected a second form of recovery is also suggested. Narratives of continued use suggest recovery *in* and something more akin to "A deeply personal, unique process of changing one's attitudes, values, feelings, goals, skills and/or roles. It is a way of living a satisfying, hopeful and contributing life, even with limitations caused by the illness. Recovery involves the development of new meaning and purpose in one's life."[54]

Notwithstanding this, a third form of recovery is also made apparent across the contradictions and areas of debate within the materials. Both the recovery *from* and the recovery *in* model are challenged by the vague ontological status of mental illness across the materials analysed. Where the disease model is questioned – and mental illness is situated as a "problem of living" - recovery can no longer be about recovering from, or learning to live with limitations caused by, illness. In the absence of overt discussion, users appear to be implicitly invited to bring together knowledge from recovery and positive psychology and, in the process, see recovery as "personal": "Personal recovery involves working towards better mental health, regardless of the presence of mental illness."[55] Across the materials, "personal recovery" appears accompanied by an ever-increasing imperative built on the normalisation of apps as a form of self-care. Conversational agents should be used by all; not least because shifting temporalities mean we do not always find others awake when we may need them. AI becomes somewhat diminished in status as simply a way to deliver more accessible care.

"Personal recovery" is increasingly individualised and the experiences of the user may become further alienated from their own recovery. When we talk of recovery from versus recovery in, scientific versus consumer models of recovery, clinical versus personal recovery or versus social recovery[56] we are often talking about whether it is clinical or personal expertise which is brought to the fore. The conversational agents explored here potentially trouble this dichotomy by situating biomedical virtue within Cognitive Behavioural Therapy itself rather than the individual user or therapist. It may be more appropriate then to suggest that chatbots are being shaped by, and shaping, (im)personal recovery imperatives.

Further research is needed here. The limited scope of the current study means there remains a need to delve deeper into the positionings embedded within use of the apps. "Walkthroughs" offer the possibilities for further exploration of digital applications" "sociocultural representations as much as its technological features or data outputs, which also have social and cultural influences."[57] The "walkthrough" method requires researchers to go through individual chatbots paying attention to such things as (I) the app's vision; (ii) its operating model; (iii) its governance; (iv) mediator characteristics, such as how the app guides users through activities via menus and buttons; (v) registration and entry; (vi) and app suspension and closure.[57] Such approaches can provide data on intended purpose, embedded cultural meanings, and implied ideal users.[57] Applying this method to a range of chatbots can shed light on how different design contexts impact on the ways that recovery is embedded within these technologies. Capturing stakeholder meanings and motivations is also important moving forward.

As well as further exploring how chatbots are shaped by, and shaping, recovery and asking what models are excluded and what are the processes of truth-making, future research needs to consider how social scientists might work with key stakeholders to challenge enduring hegemonic forms. It is also important that research engage with the question of whether we *should* focus on recovery. As Pilgrim and McCranie[8] note, whatever sociological frame we employ: "it is clear that recovery as rhetoric in the policy and services field is neither self-evidently desirable or worthy".

ORCID iDs: Robert Meadows ⓘ https://orcid.org/0000-0002-9165-1305
Christine Hine ⓘ https://orcid.org/0000-0002-1172-0252

References

1. McWade B. Temporalities of mental health recovery. *Subjectivity* 2015; 8: 243–260.
2. Department of Health. *No health without mental health: a cross-government mental health outcomes strategy for people of all ages. Supporting document – the economic case for improving efficiency and quality in mental health*. London: Department of Health, 2011.
3. Department of Health. *Closing the gap: priorities for essential change in mental health*. London: Department of Health, 2014.
4. Taggart H. *The five year forward view for mental health*. London: Department of Health, 2016.
5. Pilgrim D. Recovery and current mental health policy. *Chronic Illness* 2008; 4: 295–304.
6. Davidson L, O'Connell MJ, Tondora J, et al. Recovery in serious mental illness: A new wine or just a new bottle? *Professional Psychology: Research and Practice* 2005; 36: 480.
7. Vigo D, Thornicroft G and Atun R. Estimating the true global burden of mental illness. *Lancet Psychiatry* 2016; 3: 171–178.
8. Pilgrim D and McCranie A. *Recovery and mental health: a critical sociological account*. Basingstoke: Macmillan International Higher Education, 2013.
9. Harper D and Speed E. Uncovering recovery: the resistible rise of recovery and resilience. In *De-medicalizing misery II*. London: Palgrave Macmillan, 2014, pp. 40–57.
10. *Recovery in the bin*, https://recoveryinthebin.org/ (accessed 1 October 2020).
11. Caffier J. I tried to treat my depression with AI therapy, *VICE*, https://www.vice.com/en_uk/article/vby8ma/i-tried-to-treat-my-depression-with-ai-therapy (2018, accessed 1 October 2020).
12. Hammersley P and McLaughlan T. Campaign for the abolition of the schizophrenia label. *Asylum Magazine*, https://asylummagazine.org/2017/09/campaign-for-the-abolition-of-schizophrenia-label-by-paul-hammersley-and-terence-mclaughlin/ (accessed 1 October 2020).
13. Lasalvia A and Ruggeri M. Renaming schizophrenia: benefits, challenges and barriers. *Epidemiol Psychiatr Sci* 2019; 28: 251–253.
14. Foley T and Woollard J. *The digital future of mental healthcare and its workforce: a report on a mental health stakeholder engagement to inform the Topol Review*. NHS, https://topol.hee.nhs.uk/wp-content/uploads/HEE-Topol-Review-Mental-health-paper.pdf (2019, accessed 1 July 2020).
15. Vaidyam AN, Wisniewski H, Halamka JD, et al. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. *Can J Psychiatry* 2019;64:456–464.
16. Berg M. Making sense with sensors: self-tracking and the temporalities of wellbeing. *Digital Health* 2017; 3: 2055207617699767.
17. Pickersgill M. Digitising psychiatry? Sociotechnical expectations, performative nominalism and biomedical virtue in (digital) psychiatric praxis. *Sociol Health Illness* 2019; 41(Suppl 1): 16.
18. Williams SJ, Coveney C and Meadows R. "M-apping'-sleep? Trends and transformations in the digital age. *Sociol Health Illness* 2015; 37:1039–1054.
19. Abd-Alrazaq AA, Alajlani M, Alalwan AA, et al. An overview of the features of chatbots in mental health: a scoping review. *Int J Med Inform* 2020; 132: 103978
20. Singh O. Chatbots in psychiatry: can treatment gap be lessened for psychiatric disorders in India. *Indian J Psychiatry* 2019; 61: 225.
21. Robinson J and Thorn P. Do chatbots have a role to play in suicide prevention? *The Conversation*, https://theconversation.com/do-chatbots-have-a-role-to-play-in-suicide-prevention-105291 (2019, accessed 1 October 2020).
22. Fulmer R, Joerin A, Gentile B, et al. Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: randomized controlled trial. *JMIR Mental Health* 2018; 5: e64.
23. Suganuma S, Sakamoto D and Shimoyama H. An embodied conversational agent for unguided internet-based cognitive behavior therapy in preventative mental health: feasibility and acceptability pilot trial. *JMIR Mental Health* 2018; 5: e10454.
24. Joerin A, Rauws M and Ackerman ML. Psychological artificial intelligence service, Tess: delivering on-demand support to patients and their caregivers: technical report. *Cureus* 2019; 11: 1–7.
25. Inkster B, Sarda S and Subramanian V. An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. *JMIR mHealth and uHealth* 2018; 6: e12106.
26. Davidson L and Roe D. Recovery from versus recovery in serious mental illness: one strategy for lessening confusion plaguing recovery. *J Mental Health* 2007; 16: 459–470.
27. Fox J. The importance of expertise by experience in mental health services. *Int J Leadership Public Serv* 2008; 4:39–43.
28. Narayanan S. *How AI is transforming mental health*, https://becominghuman.ai/how-ai-is-transforming-mental-health-c0804ff075af (2019, accessed 1 October 2020).

29. Vogel DH, Krämer K, Schoofs T, et al. Disturbed experience of time in depression—evidence from content analysis. *Front Human Neurosci* 2018; 12: 66.

30. Koutsouleris N, Kambeitz-Ilankovic L, Ruhrmann S, et al. Prediction models of functional outcomes for individuals in the clinical high-risk state for psychosis or with recent-onset depression: a multimodal, multisite machine learning analysis. *JAMA Psychiatry* 2018; 75: 1156–1172.

31. Stanghellini G, Ballerini M, Presenza S, et al. Psychopathology of lived time: abnormal time experience in persons with schizophrenia. *Schizophr Bull* 2015; 42: 45–55.

32. Fryxell AR. Psychopathologies of time: defining mental illness in early 20th-century psychiatry. *Hist Human Sci* 2019; 32: 3–31.

33. Fava GA and Visani D. Psychosocial determinants of recovery in depression. *Dialog Clin Neurosci* 2008; 10: 461.

34. Jasanoff S (ed.) *The idiom of co-production in states of knowledge: the co-production of science and social order*. London: Routledge, 2004.

35. Winner L. Do artefacts have politics? *Daedalus* 1980; 109: 121–135.

36. Dieter M, Gerlitz C, Helmond A, et al. Multi-situated app studies: methods and propositions. *Soc Med Soc* 2019; 5: 2056305119846486.

37. Roniotis A and Tsiknakis M. Detecting depression using voice signal extracted by chatbots: a feasibility study. In: *Interactivity, game creation, design, learning, and innovation*. Switzerland: Springer, Cham, 2017, pp. 386–392.

38. Patton MQ. *Qualitative evaluation and research methods*. 2nd ed. London: SAGE Publications, 1990.

39. Clarke J. Can artificial intelligence help with depression? *Verywell.mind*, https://www.verywellmind.com/can-artificial-intelligence-help-depression-4158330 (2019, accessed 1 October 2020).

40. Fitzpatrick KK, Darcy A and Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR Mental Health* 2017; 4: e19.

41. Dwyer R and Fraser S. Making addiction in screening and diagnostic tools used in AOD and other health settings. *Health Sociol Rev* 2016; 25: 223–239.

42. Tubb R. Woebot – an AI-based therapy chatbot to help with depression and anxiety, *Richard Tubb: the IT growth expert*, https://www.tubblog.co.uk/blog/tech-reviews/woebot-an-ai-based-therapy-chatbot-to-help-with-depression-and-anxiety/ (accessed 1 October 2010).

43. Ng A. Woebot: AI for mental health. *Medium*, https://medium.com/@andrewng/woebot-ai-for-mental-health-d0e8632b82ba (2017, accessed 1 October 2010).

44. Collins HM. *Artificial experts: social knowledge and intelligent machines*. Cambridge: MIT Press, 1990.

45. Sachs SE. The algorithm at work? Explanation and repair in the enactment of similarity in art data. *Inform Commun Soc* 2020; 23(11): 1689–1705.

46. Brodwin E. A Stanford researcher is pioneering a dramatic shift in how we treat depression—and you can try her new tool right now. *Business Insider*, https://www.businessinsider.com/stanford-therapy-chatbot-app-depression-anxiety-woebot-2018-1?r = US&IR = T (2018, accessed 1 October 2010).

47. Molteni M. The chatbot therapist will see you now. *Wired Magazine*, https://www.wired.com/2017/06/facebook-messenger-woebot-chatbot-therapist/ (2017, accessed 1 October 2010).

48. Pilgrim D. The hegemony of cognitive-behaviour therapy in modern mental health care. *Health Sociol Rev* 2011; 20:120–132.

49. Guilfoyle M. CBT's integration into societal networks of power. *Eur J Psychother Counsel* 2008; 10: 197–205.

50. Ruckenstein M. Toying with the world: children, virtual pets and the value of mobility. *Childhood* 2010; 17: 500–13.

51. Szasz TS. *The myth of mental illness*. *American Psychologist* 1960; 15(2): 113.

52. Pies R. The bereavement exclusion and DSM-5: an update and commentary. *Innov Clin Neurosci* 2014; 11: 19.

53. Nutt D, Wilson S and Paterson L. Sleep disorders as core symptoms of depression. *Dialog Clin Neurosci* 2008; 10: 329.

54. Anthony WA. Recovery from mental illness: the guiding vision of the mental health service system in the 1990s. *Psychosoc Rehabil J* 1993; 16: 11.

55. Slade M. Mental illness and well-being: the central importance of positive psychology and recovery approaches. *BMC Health Serv Res* 2010; 10: 26.

56. Slade M and Longden E. Empirical evidence about recovery and mental health. *BMC Psychiatry* 2015; 15: 285.

57. Light B, Burgess J and Duguay S. The walkthrough method: an approach to the study of apps. *New Media Soc* 2018; 20: 881–900.