AUTHOR BIO

Francis Russell is the coordinator of the Bachelor of Arts Honours course at Curtin University. He has a PhD in Literary and Cultural Studies from Curtin University, and researches the political and philosophical implications of mental illness, alongside conducting broader research into neoliberal culture. He has published in *Deleuze and Guattari Studies*, *Space and Culture*, *Ctrl-Z: New Media Philosophy*, *Cultural Studies Review*, and *Somatechnics*. Along with the artist David Attwood, he is the co-editor of the essay collection, The *Art of Laziness: Contemporary Art and Post-Work Politics*, published by Art + Australia in 2020.

Better Living Through Algorithms

Francis Russell

ABSTRACT

The use of machine learning and algorithmic processes to screen for mental illnesses, and to propose potential treatments, has caused some to be concerned about the possibility of increasingly impersonal and invasive forms of psychiatric surveillance. For some, the rise of big data and algorithmic psychiatry presents the possibility of a future where the mentally ill are increasingly dominated by machines. Concerns about the use of algorithms in the diagnosis of mental illnesses, and the devising of treatments, perhaps overlooks the extent to which an algorithmic revolution has been facilitated by existing human-centred psychological therapies. Through the dominance of cognitive-behavioural therapy (CBT), commonly experienced mental illnesses such as anxiety or depression have come to be understood as the result of faulty processes of recursive thinking. The model of consciousness that underpins cognitive-behavioural approaches, suggests that to be human is to already engage in forms of cognition that are open to algorithmic manipulation, insofar as to think is to produce recursive rules of self-conduct. Accordingly, this paper seeks to articulate the inhuman model of thought that is assumed by CBT, and to consider how it has opened the space for an algorithmic revolution in mental health.

KEYWORDS

CBT, algorithmic thinking, biopolitics, therapeutic culture, mental illness

In contemporary mental health therapy, therapeutic "chatbots" have come to be seen as part of an innovative new age of medical care. While the first computer-based mental health therapist is arguably ELIZA, developed in the 1960s at the MIT Artificial Intelligence Lab, it is in the last decade or so that enthusiasm for digitised therapists has become visible in mainstream media and in policy debates. Such enthusiasm seems to be driven by a recognition of both the growing capacities of machine learning and speech and facial recognition technologies, and the growing number of mental health sufferers – a number that is putting increased pressure on underfunded mental health services. In 2015 *The Atlantic* reported that researchers at Carnegie Mellon University were utilising machine learning in order to "trace connections between facial expressions and emotional states among depressed people" (Lafrance). Ellie, a virtual therapist, has been developed in order to screen for mental illness through analysing "facial and body movement and tone of voice" (Robinson). Such automated forms of diagnosis are hoped to "help transform cash-strapped mental health services" insofar as they can provide assessments of mental suffering quickly, cheaply, and in a purportedly objective manner (Robinson). For those researching into the development of "virtual humans" like Ellie, key social skills such as "active listening, mimicry, and gesture" are advancing rapidly, and, accordingly, machines will be purportedly able to "develop intimacy" with human beings for the first time (Lucas, Gratch, King et al. 94).

The possibility that algorithmic processes might come to replace human therapists and the human diagnosis of mental illnesses, can make such technology appear as a besieging force, an alien threat to the psychiatric establishment that will have to be fended off if human beings are going to maintain their professional status as expert practitioners. Indeed, from the perspective of talk therapies that emphasise the significance of the unconscious in the treatment of mental illnesses, the possibility of human symptoms being reduced to calculable and predictable datasets would be an absurdity. Despite this, recent developments in algorithmic decision-making have helped usher in new automated approaches to mental health care. Moreover, such approaches work, so it is claimed, not in spite of practices espoused by more conventional talk therapists, but because of them. One example is Woebot, an online application that "uses brief daily chat conversations, mood tracking, curated videos, and word games to help people manage mental health" (Molteni). In a blog post for the Woebot website, psychologist and CEO Alison Darcy writes that the application provides an "automated coach" that helps the user to practice "good thinking hygiene" (Darcy). Interestingly, Woebot is not positioned as being antithetical to conventional talk therapies, and its promotion does not draw on the conventional "two cultures" opposition between the humanities and STEM. Instead, Darcy states that a specific form of talk therapy, cognitive behavioural therapy or CBT, has allowed mental health treatment to open itself up to digital forms of proliferation. Darcy writes that,

> the popular idea about therapy is that it holds a special kind of magic that can only be delivered by individuals who are highly trained in this mysterious art form. The truth is that modern approaches to mental

health revolve around practical information gathering and problem solving. (Darcy)

For Darcy, the best example of such "modern approaches to mental health" is CBT, since this form of therapy is "highly structured and practical, and involves a lot of learning," which, for Darcy, means that it "lends itself well to being delivered over the internet" (Darcy).

This article argues that digital platforms like Woebot present a new challenge to cultural theorists interested in studying and critiquing the ways in which newly developing digital technologies are impacting our ontological understanding of mental illness, and how practices of mental health therapy are developing. While there is growing concern around the uses of algorithmic decision-making, facial recognition technology, and machine learning in the diagnosis and treatment of mental illness, such concerns are usually connected to some overarching concept of what has been varyingly referred to as "algorithmic governmentality" (Rouvroy & Berns), "surveillance capitalism" (Zuboff), or the "automatic society" (Stiegler). While these cannot be taken as synonymous terms, all three nevertheless attempt to describe the emergence of social formations in which the widest possible range of decisions are open to intertwined processes of automation and surveillance. Today, algorithms "distribute goods and services, classify persons (potential partner, customer, criminal), try to detect terrorists and much more" (Matzner 123). Accordingly, while these concepts are of great critical import, they are best equipped for approaching modes of rationality and relationships of power immanent to algorithmic systems, "big data", and the emergence of digital platforms. It is not the intention of this article to argue against such an approach, but, instead, to pose the adjacent question of how certain non-algorithmic and even seemingly humanistic practices of mental health care may have paved the way or opened up the space for furthering algorithmic interventions. Put more succinctly, this article poses the question of how CBT, as a dominant mental health therapy, might have helped to reframe mental illness and mental health therapy so as to make both more amenable to algorithmic processes and non-human decision making. In order to pursue such a question, this article will begin by providing a brief overview of the computational discourse that shapes our understanding of algorithms. In particular, this section will explore the fantasies of control that help to mobilise the desire of algorithmic solutions. This will then lead to a discussion of the work of one of CBT's progenitors, Aaron Beck, and his reframing of mental illness and mental health therapy. By way of conclusion, this article will look to offer a critique of the alliance between CBT and algorithmic processes, as embodied in digital platforms like Woebot.

Computational Discourse and Control

As humanities scholars have shown increased interest in the cultural, political, and economic power of algorithms, the problem of a lack of a stable definition for the term "algorithm" has become a pressing concern. The difficulty of the situation is not merely a product of the "two cultures" divide, but, as Ed Finn notes, is partly a product of the fact that "for computer scientists the term

remains more of an intuitive, unexamined notion than a delineated logical concept grounded in a mathematical theory of computation" (Finn 17-18). Algorithms have been varyingly defined as systems of knowledge and problem solving strategies – or "logic + control" (Kowalski 424); as a "recipe, an instruction set, a sequence of tasks to achieve a particular calculation or result" which could include "the steps needed to calculate a square root or tabulate the Fibonacci sequence" (Finn 17); and, in a more metaphysical register, as "exteriorised reason" (Hui 123). Rather than attempting to locate a singular definition for the term "algorithm," this article is instead interested in the discursive framing of algorithms, which has allowed the term to become increasingly prominent in our cultural imaginaries. That is to say, rather than pretending to offer a technical definition that uncovers the essence of algorithms, this article is instead interested in a set of idealised characteristics of algorithms, characteristics that are vital to the propagation of algorithmic interventions at a cultural and political level. It is this article's contention that the apparent utility, objectivity, and universal functionality of algorithms is made intelligible and desirable by way of an already influential computational discourse that radically reframes understandings of truth and meaning.

In the documentary *Transcendent Man*, best-selling author and influential futurist Rey Kurzweil provides the following enigmatic and evocative description of computation:

> Well I was thinking about how much computation is represented by the ocean. I mean it's all these water molecules interacting with each other. That's computation. It's quite beautiful. And I've always found it very soothing. And that's really what computation's all about. The capture of these transcendent moments of our consciousness. (Quoted in Finn 184)

While this is certainly not an attempt at providing a technical definition of computation, Kurzweil's poetic language is worth unpacking, as we can detect here one of the key rhetorical gestures of computational discourse. Kurzweil begins by discussing the ocean in terms of discrete units (water molecules) and their interactions. Whether or not an ocean can be reduced to such elements, and whether the ocean can be separated from geological, lunar, and cultural processes is left aside. Perhaps an ocean is inseparable from its various contexts, and, moreover, perhaps it is the case that oceans are a continuous or analog phenomenon – as opposed to one comprised of discrete elements that come to interact. Despite these possibilities, the phenomenon of "ocean" is here taken as separable from what surrounds it and contaminates it, and is reduced down to those elements that can be neatly studied. Such a manoeuvre is detectable in a whole range of different academic and popular conversations that participate in the discourse of computationalism. As David Golumbia has observed, fundamental to computational discourse is the striation of "otherwise-smooth details, analog details", so that they are better suited to calculation and hierarchisation (Golumbia 11). For Golumbia, the phenomena that computational theories attempt to capture – brains, languages, societies, ecologies, etc. – are analog, that is to say "they are gradable and fuzzy; they are rarely if ever exact" (Golumbia 21). From within computational discourse, however, it is not simply a matter of strategically or methodologically

substituting continuous phenomena for calculable abstractions, but, instead, taking the seemingly continuous to be a product of the discrete and calculable. Indeed, Golumbia notes that if we look at a broad range of research disciplines we find it proposed that "everything from DNA, to the interactions between subatomic particles to the shape of space-time, might be constructed from the algorithmic passing of information in some abstract sense" (Golumbia 19). [1]

The rise of algorithmic interventions in a host of institutions – whether educational, judicial, medical, etc. – cannot be simply reduced to matters of efficiency and usefulness. While algorithmic processes certainly can offer material benefits, both to those who manage them and those who are managed by them, the rise of algorithms as innovative solutions to a range of social issues has been accompanied by a particular discursive framing, one that informs our understanding of their relationship to truth, objectivity, and impartiality. As Tarleton Gillespie puts it, "more than mere tools, algorithms are also stabilisers of trust, practical and symbolic assurances that their evaluations are fair and accurate, and free from subjective error, or attempted influence" (Gillespie 179). From within computational discourse, the world is comprised of information that, once correctly gathered and sorted, can allow for utopian projects to be conceived and pursued. Accordingly, algorithmic processes are not understood as being hindered by the particular way in which they encounter the world – i.e., as fragmented data that must be creatively and usefully recombined. Instead, such processes are approached by way of an opportunity for realigning our understanding of the world's ontological status. Rather than these processes being understood as inherently limited by their need to reduce the world of phenomena to information, they are instead viewed as encountering a reality that is inaccessible to human beings; "blinded" as we are by our supposedly limited subjective view of the world. Such an understanding of algorithms as a technology that can uncover the true state of the present – and perhaps even the future – requires a great deal of discursive work, since, for developers of algorithmic processes, "the performance of algorithmic objectivity has become fundamental to the maintenance of these tools as legitimate brokers of relevant knowledge" (Gillespie 180).

Again, such a performance of objectivity is reliant on cultural narrative as much as it is dependent on displays of scientific evidence or brute efficacy. It is for this reason that theorists like Finn and Wendy Chun have associated computational discourse with early mythological understandings of language. As Finn notes, the cultural power of algorithms is inseparable from the mythological notion of language as magical, that is to say, the notion that "words are codes that change reality" (Finn 1). While sorcery, enchantments, and other superstitious notions of language might seem far removed from the rationalistic precision of algorithms, Finn maintains that today, "this figure of the algorithm as a quasi-mystical structure of implemented knowledge is both pervasive and poorly understood" (Finn 6). Such a mystical understanding of algorithms often results in their being made to appear immaterial – with information, data, and calculation appearing as somehow transcendent of the material world of labour and matter. As Chun writes, "we 'primitive folk' worship source code as a magical entity – as a source of causality – when in truth the power lies elsewhere, most importantly, in social and machinic relations" (Chun 51). Despite the economic and political forces that make

[1] N. Katherine Hayles has referred to the belief of a "computational universe" within the natural sciences (Hayles 3). Hayles states that the most full-throated endorsement of this view potentially comes from computer scientist, physicist, and businessman Stephen Wolfram. For Wolfram, as Hayles observes, "biological systems and, indeed, complex behaviours of every kind, including social and cultural systems" can be understood as forms of computation (Hayles 3).

algorithmic processes functional, the power of technical processes is often exaggerated and overemphasised. As such, "knowing (or using the right) software has been made analogous to man's release from his self-incurred tutelage" (Chun 21). [2]   For prominent figures such as Barack Obama (Shapiro) and Tim Cook (Griffin), coding will allow today's children to mitigate the uncertainties of the future labour market, and to be best positioned to help innovate our cities. The reality might instead be that algorithmic processes render human software engineers increasingly unnecessary, as the processing power of computers automates a larger and larger proportion of software development and management. Regardless, both the utopian and dystopian understanding of code as a cultural and technical practice functions to obscure the material realities and embedded power relations that form them. From within this discourse, the optimised algorithm, and the highly trained coder, are both granted enormous power to manipulate the world around them – not because the world is being politically and economically restructured for the sake of computational interventions, but, supposedly, because the world is ontologically structured in terms of information, and is best approached by way of computational logics and practices of calculation. As Evgeny Morozov puts it, such discourse recasts "all complex social situations either as neatly defined problems with definite, computable solutions or as transparent and self-evident processes that can be easily optimised – if only the right algorithms are in place!" (Morozov 5).

Within this discourse, algorithms do not appear as politically contestable technologies that reveal certain facets of the world, but instead as expressions of the world's rationality. As Zuboff has observed, this requires the constant framing of these interventions as historically necessary – simply part of the evolution of human societies, and signalled through the language of the "new 'age,' 'era,' 'wave,' 'phase,' or 'stage'" of whatever technological solution is being discussed (Zuboff 222-223). Whether in the assessment of work performance, health status, suitability for a loan, or likelihood of committing an illegal offence or reoffending on release from a correctional institution, algorithms offer the promise of transcending bias, ideology, or subjective preference by way of producing exact measurements and assessments of phenomena. Despite the growing body of literature that points to the various inequalities and injustices either produced or intensified by augmenting human decision-making with algorithms, failure and disaster is increasingly woven back into the "solutionism" proffered by computational discourse. As Roberge and Seyfert observe, there are limitations to simply drawing attention to the implausibility of the promises made by champions of algorithmic interventions, since, "if everything were functioning smoothly, these promises would be superfluous and would simply disappear" (Roberge & Seyfert 28). Such optimism is perhaps not the result of an ignorance of algorithmic failure, but is instead the discursive condition of possibility for their flourishing: "the dream of algorithmic objectivity, of smooth operations and efficiencies, of autonomy and the hope of a higher rationality makes sense especially in contrast to constant failures" (Roberge & Seyfert 28).

By engaging with computational discourse, we have attempted to discuss algorithms as cultural artefacts rather than as purely technical systems or tools. Viewed as part of this discourse, we can see algorithms as embodiments of

[2] The "learn to code" meme is an interesting example of the adoption of coding as a symbol of sovereignty or self-mastery in contemporary culture. See: https://knowyourmeme.com/memes/learn-to-code.

desires for optimisation and control, and we can better understand the willingness of some to overlook the limitations and injustices that are inseparable from their implementation. Framed by computational discourse, the failures and limitations of algorithms are always a result of a lack of sufficient optimisation, rather than the product of incommensurability. Once phenomena are reduced to a single stratum as information, they can then be more effectively ranked and hierarchised. If, however, we wish to oppose the inhuman depiction of the world offered by computational discourse, how do we explain its adoption in institutions concerned primarily with human meaning? That is to ask, if there is nothing neutral or inevitable about the adoption of algorithms to solve social problems, how do we explain their adoption within institutions whose practices are not obviously compatible with those preferred by algorithmic systems? Indeed, and to return to the question of contemporary mental health, computational discourse's emphasis on objectivity or impartiality, functionality, and universality would suggest a necessarily antithetical relationship with conventional talk therapies.

While talk therapies such as psychoanalysis have certainly produced their own theoretical systems, which often abstract from the individual case to speculate on the broader character of symptom formation and dissolution, there is nevertheless a sense in which talk therapies emphasise the singularity of the patient's speech. Indeed, for one of the intellectual fathers of CBT, Aaron Beck, mental health therapy should orient itself as much as possible to the meaningful statements patients make. Writing in the 1970s, Beck asked, "why do so many students of human nature and its aberrations turn away from conscious meaning?" (Beck 47). At this time, Beck, a former Freudian analyst, was attempting to popularise "cognitive therapy," a methodological forerunner to today's CBT. Beck was so motivated by a need to return to human meaning – that which "provides the richness of life [and] transforms a simple event into an experience" (Beck 47) – that he undertook to develop a radical new form of mental health therapy. Such interests in fixing human meaning at the centre of treatment might create the impression that CBT would be the most resistant to automation through algorithmic processes. Indeed, one of Beck's goals was to produce a form of treatment that avoided the "elaborate infrastructure of symbolic meanings" that he saw in those therapies that have their origins in Freudian theory, and that he felt took therapy away from the lived world of the patient. However, in today's milieu of mental health treatment, it appears conversely that CBT has opened up the space for algorithmic interventions, whereas psychoanalytic practitioners have resisted such developments and "innovations." Accordingly, in the next section we will look to unpack the theoretical underpinnings of CBT by providing an overview of some of Beck's key ideas. Moreover, by exploring these theoretical underpinnings, we will argue for the active role played by humanistic disciplines such as cognitive therapy in ushering in automated forms of therapy.

Testable, Teachable, Economical – Beck's Cognitive Therapy

Modern CBT is an amalgamation of at least two therapeutic practices developed in the twentieth-century: Aaron Beck's cognitive therapy, and Albert Ellis's rational emotive therapy. CBT training manuals, such as Stefan

Hofmann's *An Introduction to Modern CBT: Psychological Solutions to Mental Health Problems*, credit both Beck and Ellis as founding CBT (Hofmann xviii). While Beck's and Ellis's therapies cannot be made entirely synonymous, they nevertheless share the fundamental notion that the role of therapy is to "help patients realise that their own beliefs contribute greatly to, maintain, and even cause their psychological problems" (Hofmann vxiii). Hofmann claims that this is the "central notion" of modern CBT, and, for this reason, the works of Beck and Ellis can provide us with a great deal of insight into the theoretical underpinnings of contemporary CBT practice (Hofmann xix). Before moving on to a discussion of such underpinnings, it is worthwhile describing the extent of CBT's influence within contemporary mental health care. For the psychotherapist Farhad Dalal, CBT is one of the most dominant mental health practices in the liberal-democratic world. "If you go to your GP because of feeling depressed for some reason," Dalal writes, you will almost certainly be offered "anti-depressants or/and the 'one-size-fits-all' manualised treatment called CBT" (Dalal 1). Although Dalal is writing here about the British context, CBT is an enormously popular form of treatment in countries like the United States and Australia. [3] It is beyond the scope of this article to engage with the singular histories that are inseparable from CBT's rise to prominence across the globe. Nevertheless, CBT's success is broadly attributable to its capacity to be marketed as more scientific and empirical than other forms of therapy (Maxwell & Tappolet 3) and because it is vastly cheaper than existing talk therapies. CBT is usually administered for between 10 and 20 sessions (Leichsenring, Hiller, Weissberg, et al. 235), and so is much quicker to complete than psychoanalytic or psychodynamic treatments in which patient and therapist may work together for years. [4] As one of the most popular forms of mental health therapy, it is perhaps unsurprising that CBT has been adopted as the candidate for digitisation and automation. Despite the plausibility of this explanation, by turning to the foundational work of Beck we can begin to see how cognitive therapy powerfully transforms the very notion of mental illness, thereby opening up mental health care, albeit unwittingly, to further levels of automation.

Finding himself frustrated with the then-dominant behavioural, psychoanalytic, and neuropsychiatric explanations of mental illness, Beck began to abandon the principles of his training as a Freudian and to develop his own form of "cognitive therapy." As the name implies, cognitive therapy focuses on the cognitions experienced by patients, and attempts to understand how certain cognitive "schema" cause, shape, and prolong mental illnesses. Beck claimed that "schemas" are the cognitive structures that "mould" the "raw data" offered by sense experience "into thoughts or cognitions" (Beck, "Thinking" 562). While schemas are hypothetical descriptions of the mind's structure, and are not accessible to observation, schemas are believed to produce "automatic thoughts" which are directly observable by patients – albeit once they have received proper training and guidance from a cognitive therapist. Or as Beck puts it, "when the schemas become activated, the belief becomes operative in providing the content of an interpretation" (Beck, *Personality Disorders* 35). As Frank Wills elaborates, "a schema is hypothesised to be a general cognitive structure" comprised of the connections between "experiences, memories, thoughts, attitudes, and beliefs" (27-28). For Beck, a mental illness such as depression is caused by "the dominance of certain

[3] For an illuminating discussion of the political struggles between promoters of CBT and more established psychoanalysts in France, see: (Aflalo)

[4] As Beck himself put it, "since collection and use of data in cognitive therapy do not require much time, this approach allows for brief psychotherapy. Hence, it is *economical.* In many cases, short-term, structured cognitive therapy may take only ten to twenty sessions" (Beck 317). Later in the same text he proclaims that, "in contrast to psychoanalysis, cognitive therapy is readily *comprehensible* to the patient, *testable* by the researcher, *teachable* to the student, and *economical* in terms of time and money" (Beck 320).

cognitive schemas" through which the patient comes to "systemically misconstrue" themselves, their relationships with others, and their future (Beck 264). Or as he and Emily Haigh put it elsewhere, "biased perceptions are accumulated and stored in memory and lead to the formation of cognitive processing structures, or schemas, which incorporate the biased beliefs. Schemas are strengthened by exposure to a severe adverse event or repeated stressful experiences" (Beck & Haigh, "Generic Cognitive Model" 5). For example, once a person's mind becomes structured by way of depressive schema, they might find it difficult to interpret events positively. Even when an event occurs that others might perceive as positive, the sufferer's schema moulds the event into a negative one. As such, they might interpret the experience of success as inevitably leading to failure, they might feel that they are fundamentally unlikeable, and they might assume the worst of every situation.

While acknowledging that the schema related to psychopathology are connected to externally occurring events that have negatively impacted the individual, Beck nevertheless argued that it was insufficient to mobilise therapy at the level of external behaviour. Instead, Beck thought it "necessary to formulate the problems of the depressed patient in cognitive terms" (Beck 264). In other words, in order for the analyst and patient to truly understand how the techniques of cognitive therapy will be effective, there is a need to reimagine the illness or problem itself in cognitive terms, such that "the characteristic of depression can be viewed as expressions of an underlying shift in the depressed patient's cognitive organisation" (Beck 264). While cognitive approaches do not exclude the possibility of beginning therapy with a behavioural intervention – for instance, getting a depressed patient who is socially isolated to seek out contact with friends and family – the commitment to a fundamentally cognitive aetiology means that the patient is viewed as at risk of being incapable of appreciating the gains made through behavioural modification. Put differently, once it has been accepted that the source of the patient's problems lies with their cognitive schema, there will always be a need to optimise the patient at a cognitive level, as the patient's capacity to grow or transform via behavioural interventions will ultimately be interpreted through the patient's depressed cognitions. Or as Beck puts it,

> the therapeutic approach may be behavioural, for example, mobilising the patient into more activity and positively reinforcing certain types of activity. The underlying attitude, however, is the component that needs to be changed ultimately if the totality of the depression is to be influenced. Thus, the goal is cognitive modification. (Beck 268)

For Beck, and for contemporary practitioners of the CBT inspired by Beck's cognitive therapy, the key to helping the patient is to avoid as much as possible the quagmire of the biographical, social, or even political contexts through which symptoms may have arisen. As Hofmann writes, "knowing the initiating factors provides neither necessary nor sufficient information for treatment" (xxi). Hofmann explains this point by way of a medical illustration. If, contends Hofmann, we happen to break our arm, a doctor "may ask how it happened out of curiosity, but the information is rather unimportant for selecting the appropriate treatment – putting the arm in a cast" (xxi). Such contextual

information hinders the treatment process, since, following the ontology of mental illness offered by cognitive therapy, a mental illness is a maladaptive cognitive response. While Beck never denied that people experience painful or unjust circumstances, he nevertheless maintained that a mental illness is a fundamentally incorrect judgement that has become deeply engrained. Accordingly, the supposed power of cognitive therapy results from its capacity to forgo any complex and lengthy hermeneutic process, through which the patient attempts to understand themselves, and to instead focus on training the patient to notice the simple and straightforward thoughts that precipitate negative emotions. Following Beck, if the patient is to benefit from a cognitive approach: "first, he [sic] has to become aware of what he is thinking. Second, he needs to recognise what thoughts are awry. Then he has to substitute accurate for inaccurate judgements. Finally, he needs feedback to inform him whether his changes are correct" (Beck 217).

Writing on debates between psychoanalysts and cognitive therapists for *The Guardian*, Oliver Burkeman has neatly summarised why CBT is today more favourably viewed. Psychoanalysis, contends Burkeman, is convoluted and mysterious, whereas CBT is "a down-to-earth technique focused not on the past but on the present; not on mysterious inner drives, but on adjusting the unhelpful thought patterns that cause negative emotions" (Burkeman). Whereas Psychoanalysis commits itself to exploring the singular unconscious of the individual analysand, CBT views mental illness as being largely a technical problem that can be resolved with some simple "down-to-earth" modifications and optimisations in the patient's thought. Psychoanalysts like Freud or Lacan argued that the patient's utterances, the meanings of their dreams, slips of the tongue, or memories cannot be neatly or straightforwardly understood – and hence require a slow and careful collaboration between analyst and analysand. In contrast, CBT holds that mental illnesses result from transparently faulty patterns of reasoning – perhaps not unlike a poorly designed algorithm – in which external data is poorly ordered and hierarchised. Indeed, we can find Beck explicitly relying on this kind of conceptual language when he writes that the "rapid cognitive processing" that is required to navigate everyday experiences, "is facilitated by the use of opposing categories to sort information" and the application of "relevant formulas or algorithms" (Beck, *Personality Disorders* 35). Such an appeal to formulas and algorithms, which present thought as a process of rule following, is deeply antithetical to any therapy that takes the unconscious seriously. Whereas Lacan famously stated that "there is no metalanguage" within the field of speech (Lacan 688), the cognitive therapist acts as if "common sense" can allow one access to the "metalanguage" of the schemas and their algorithmic processes. Whereas a psychoanalyst might caution against simple or univocal interpretations of thoughts or statements, CBT holds that thoughts and statements lead us directly to the schematised structure of the sufferer's mind. As the Lacanian analyst David Ferraro has commented, within CBT "there are presumptions of the perfect self-reflexivity, of singularity of volition and will, and a lack of division" with regards to subjectivity (Ferraro). Within the work of Beck and contemporary CBT practitioners, one finds the presumption that "what is 'good', 'enjoyable', 'pleasant' and so forth are more or less self-evident, and that any rational person would pursue these things as ends" (Ferraro).

Beck's cognitive therapy, and the ensuing practices that utilise the label CBT, have been successful in repositioning mental illness as a technical problem that can be mitigated, if not solved, by way of practical and affordable interventions. Indeed, Beck was fond of the characterisation of the patient and therapist "as scientists who collaborate in investigating the patient's personal constructs" (Beck 296). It is arguably this commitment to therapy as problem-solving, and mental illness as the distortion or maladaptation of a rule following process that has allowed CBT to serve as a fertile space for automated interventions in therapy. Just as Golumbia views computational discourse as inviting us to understand language by way of the logics of computers – i.e., as "closed systems, subject to univocal, correct 'activating' interpretations" and lacking in "ambiguity, context, and polysemy" (Golumbia 84) – CBT helps train the patient to look for the "closed system" of their "cognitive organisation," that is to say, a series of thoughts and beliefs characterised by monosemy and transparency (Beck 270). Furthermore, just as computational discourse frames complex social and political conflicts as being resolvable through discrete technical solutions produced through attention to the correct data, Beck argued that his cognitive approach to therapy could help foster "a sense of mastery" insofar as "solving one problem frequently inspires the patient to approach and solve other problems" (Beck 232). Indeed, for Beck, "a bonus of successful therapy is not only freedom from the original problems, but a thorough psychological change that prepares him [sic] to meet new challenges" (Beck 232).

The developers of Woebot and similar applications such as Wysa frame CBT's reductive interpretation of mental illnesses, and its focus on efficiency and optimisation, as strengths that should be embraced rather than fundamental limitations. CBT's framing of context, polysemy, and ambiguity as superfluous, if not detrimental with regards to therapy has allowed it to function as a powerful ally in the automation of mental health care. While the chatbot component of apps like Woebot have shown severe deficiencies, such issues are entirely compatible with the discursive framing of algorithmic interventions discussed earlier in the previous section. [5] Indeed, the failure of an algorithmic process is often used to reinforce the apparent need for further funding, research, and user-interaction, so that future interventions will be more successful. As Hui puts it, "failures and errors are accepted not only as a necessity for technological progress, but also have become immanent to its operation and maintenance" (Hui 131-132). [6] However, it would be vastly more difficult to utilise such failures as a means of reinforcing the cultural value of algorithms if dominant understandings of mental illness and mental health treatment sat in opposition to computational discourse's emphasis on functionality, efficiency, and utility. CBT has offered the opposite, providing a framing of mental illness as a technical problem that can be solved with limited resources and training, and one that requires limited specialist knowledge. According to CBT, mental illnesses are problems of information processing, whereby faulty systems of ordering data – both in terms of importance and probability – have come to dominate the sufferer's life. Any concern for the unconscious dimension of language and thought, or the singularity of the individual sufferer, is sidestepped in the promotion of a "down-to-earth," affordable practice, which renders problems of human self-understanding and meaning as relatively simple cognitive distortions. For this reason, the

[5] Perhaps the most disturbing example of such deficiencies was uncovered by the BBC in 2018. Technology reporter Geoff White notes that in an experiment conducted by the BBC, Woebot was unable to detect clear statements relating to sexual abuse. See: (White).

[6] Hui provides the following illuminating quote from the Amazon engineer Jinesh Varia: "in particular, assume that your hardware will fail. Assume that outages will occur. Assume that some disaster will strike your application. Assume that you will be slammed with more than the expected number of requests per second someday. Assume that with time your application software will fail too. By being a pessimist, you end up thinking about recovery strategies during design time, which helps in designing an overall system better" (Hui 132). Indeed, this process of learning via failure is itself often automated.

proliferation of CBT as a response to the ever-growing rates of mental illness has helped to weaken the claim that the pain caused by depression, anxiety, and psychosis cannot be abstracted from the life of the individual sufferer. For this reason, it is appropriate that applications like Woebot are understood as relying on the model of mental illness and mental health therapy provided by CBT. Rather than having to challenge talk therapy as a rival practice, the automation of mental health therapy can be intensified once it has been redefined in the highly reductive terms promoted by Beck.

Conclusion

As Rosalind Cooper and others have argued, algorithmic power is "characterised by an imbalance regarding how transparency and (in)visibility are distributed amongst humans and machines" (43). While this is a crucial point, the character of the subject's visibility should not be taken for granted. Indeed, since Foucault it has been clear that rendering the individual visible has been inseparable from contemporary forms of governance. As Foucault so succinctly put it, with regards to disciplinary power, "visibility is a trap" (Foucault, *Discipline* 200). Especially in his lectures on abnormality (Foucault *Abnormal*) and psychiatric power (Foucault *Psychiatric*), Foucault argued that psychiatric and psychoanalytic therapies could not be separated from the production of visible subjects of governance, and, accordingly, posed the risks of exposure. Nevertheless, the attempt to produce visible individuals should be distinguished from the more contemporary power of algorithms to produce transparent, but fragmented dividuals (Cooper 30-31). Taken from Deleuze's "Postscript on the Societies of Control" (Deleuze), the term "dividual" refers to the granular subjects – if this term can still be applied – revealed through analysis of "sub- and trans-individual arrangements of matter and function (forces, genetic codes, affects, capacities, desires)" and that are "modelled as samples, data, propensities, populations and markets" (Thoburn 83). The difference between these two processes of subject-formation is perhaps found in the first's emphasis on the singularity and uniqueness of a life – uncovered through confession and the building of a case study – and the emphasis on the aggregation of data, detached from any notion of singularity. In the model described by Foucault, psychiatric power produces the sufferer's visibility by constituting their desires, dreams, and perversions into an individual identity. By contrast, algorithmic power renders the subject visible by way of tracked functionality – dividualising the subject as something that exists only in terms of what it connects to, is proximal to, and acts like.

If we can agree that talk therapies like CBT and algorithmic interventions like Woebot belong to different forms of governance – one relying on the power to individualise and the other relying on the power to dividualise – doesn't this surely weaken the argument that Beck's cognitive approach to therapy functioned to make room for further automated mental health services? Indeed, and as Celia Lury and Sophie Day have indicated, the rhetorical opponent of algorithmic healthcare – with its power to deliver "personalised" services and treatments – is the "one size fits all" system of the past (Lury & Day 18). Following this logic, whereas the mental health practitioners of the past attempted to produce a coherent identity for the patient by way of limited

psychopathological categories, algorithmic processes will supposedly allow more nuanced diagnoses by way of attending to the dividual as a node in a greater network of connections. So, if Dalal and other critics of CBT have characterised this practice as deficient due to its "one size fits all" approach, doesn't an application like Woebot fundamentally transform CBT into something dramatically new – shifting therapy away from strategies that are applicable regardless of the client, to strategies that are increasingly personalised? Such questions should hinder any attempt to make Beck's cognitive approach to mental health, or the ensuing practices labelled CBT, simply or straightforwardly synonymous with algorithmic governmentality. While they are certainly commensurable, the cognitions theorised by Beck and the computation critiqued by Golumbia, Chun, and others, should not be conflated. Moreover, while CBT and computational discourse privileges certain linguistic codes as being supremely self-reflexive, transparent, and optimisable, it is not true that CBT engages with dividuals in the manner aforementioned.

Nevertheless, this article has proposed that therapeutic practices like CBT have functioned as a space of transition between the governance of visible individuals and the governance of transparent dividuals – of patients rendered visible by the medical gaze, to patients whose singularity is seen through, so as to reveal their functionality within a surveilled network. By eliminating context – whether biographical, social, or political – and by focusing on thought as controllable, optimisable and predictable, cognitive approaches to therapy have helped to make the non-human treatment of mental illness both more conceivable and desirable. Woebot founder and CEO Alison Darcy is clear that her application "will never replace therapy or therapists, and it's not trying to" (Darcy). According to the Woebot blog, the human connection between patient and therapist – or what it sometimes called the "therapeutic alliance" – is too powerful to be replaced by a digital interface. Despite this, and according to the contemporary philosopher Catherine Malabou, algorithmic processes will not only outstrip the human capacity for calculation, but, transcending such quantitative thought, have the potential to simulate "noncalculation, that is, spontaneity, creative freedom, and the directness of emotion" (Malabou 150). Such developments in the emotional capacities of AI and human machine interactions shows that the notion of a human sharing their most intimate feelings with a machine therapist – one capable of simulating warmth, compassion, and interest – is not mere science fiction. Indeed, one can interpret such offers of reassurance as warnings of the real scope of ambition held by the developers of applications like Woebot. While Darcy believes that there will always be a place for human therapists, she nevertheless contends that applications like Woebot can help to elevate the messiness of human interaction. "It's almost borderline illegal to say this in my profession" she admits when interviewed for an article by Wired, "but there's a lot of noise in human relationships" (Molteni).

It is in such statements that we can see a deeper connection between Beck's ideas and applications like Woebot. While the difference between the individual patient of CBT and the dividual user of an application should be maintained, there is a shared understanding of the polysemy, ambiguity, and the unconscious character of human meaning as troublesome "noise" to be

eliminated. Due to such a shared commitment, even if inhuman processes don't entirely replace human therapists, there is a sense in which dominant mental health therapies seek to cast mental illness in inhuman terms – that is to say, in terms that bypass the unstable and open-ended character of human meaning. As a result, the cognitive approach to mental illness favours the view of suffering as a set of discrete technical problems. The reduction of mental illnesses to faulty propositions that can be simply decoded and neatly optimised, creates the space for computational understandings of meaning and truth to become increasingly dominant. Accordingly, while it is important to scrutinise the biases and injustices that are perpetrated and maintained when supposedly neutral algorithmic processes replace human decision making, it is just as important to consider the ways in which seemingly unrelated theories and practices of comportment open the space for such automated interventions.

Works Cited

Aflalo, Agnès. *The Failed Assassination of Psychoanalysis: The Rise and Fall of Cognitivism.* Trans A.R. Price. London: Karnac Books, 2015.

Beck, Aaron T. "Thinking and Depression II: Theory and Therapy." *Archives of General Psychiatry* 10 (1964): 561-571.

---. *Cognitive Therapy and the Emotional Disorders.* New York: International Universities Press Inc, 1976.

Beck, Aaron T. Davis, Denise D. Freeman, Arthur. *Cognitive Therapy of Personality Disorders, third edition.* New York: The Guilford Press, 2015.

Beck, Aaron T. & Haigh, Emily. "Advances in Cognitive Theory and Therapy: The Generic Cognitive Model." *Annual Review of Clinical Psychology* 10.1 (2014): 1-24.

Berns, Thomas & Rouvroy, Antoinette. "Algorithmic Governmentality and Prospects of Emancipation: Disparateness as Precondition for Individuation Through Relationships?" Trans. Elizabeth Libbrecht. *Réseaux* 177 (2013): 163-196.

Burkeman, Oliver. "Therapy Wars: The Revenge of Freud." *The Guardian*, 7 January 2016. https://www.theguardian.com/science/2016/jan/07/therapy-wars-revenge-of-freud-cognitive-behavioural-therapy.

Chun, Wendy Hui Kyong. *Programmed Visions: Software and Memory.* Massachusetts: MIT Press, 2011.

Cooper, Rosalind. "Pastoral Power and Algorithmic Governmentality." *Theory, Culture & Society* 37.1 (2020): 29-52.

Dalal, Farhad. *CBT: The Cognitive Behavioural Tsunami - Managerialism, Politics, and the Corruptions of Science*. London: Routledge, 2018.

Darcy, Alison. "Why we need mental health chatbots." *Woebot*, 30 March 2018. https://woebot.io/blog/why-we-need-mental-health-chatbots/.

Deleuze, Gilles. "Postscript on the Societies of Control." Negotiations: 1972-1990. Trans. M. Joughin. New York: Columbia University Press, 1995.

Ferraro, David. "Psychology, Epistemology, Theory and CBT: A Critique of CBT as Ideology (Part 3)." *Archives of a Divided Subject: Psychology & Psychoanalysis in the 21st Century* 5 February 2014. https://melbournelacanian.wordpress.com/2014/02/05/psychology-epistemology-theory-and-cbt-a-critique-of-cbt-as-ideology-part-3/.

Finn, Ed. *What Algorithms Want: Imagination in the Age of Computing*. Massachusetts: MIT Press, 2017.

Foucault, Michel. *Discipline and Punish: The Birth of the Prison*. Trans. Alan Sheridan. London: Penguin, 1991.

---. *Psychiatric Power: Lectures at the Collège de France, 1973-1974*. Trans. Graham Burchell. Gordonsville: Palgrave MacMillan, 2008.

---. *Abnormal: Lectures as the Collège de France, 1974-1975*. Trans.Graham Burchell. London: Verso, 2016.

Gillespie, Tarleton. "The Relevance of Algorithms." *Media Technologies: Essays on Communication, Materiality, and Society*. Eds. Tarleton Gillespie, Pablo J. Boczkowski, & Kristen A. Foot. Massachusetts: MIT Press, 2014. 167-193.

Golumbia, David. *The Cultural Logic of Computation*. Massachusetts: Harvard University Press, 2009.

Griffin, Andrew. "Kids Coding Languages Are The Future, Says Apple Boss Tim Cook From China." *Independent* 4 December 2017.

Hayles, N. Katherine. *My Mother was a Computer: Digital Subjects and Literary Texts*. Chicago: The University of Chicago Press, 2005.

Hofmann, Stefan. *An Introduction to Modern CBT: Psychological Solutions to Mental Health Problems*. Chichester: Wiley-Blackwell, 2012.

Hui, Yuk. "Algorithmic Catastrophe – The Revenge of Contingency." *Parrhesia* 23 (2015): 122-143.

Kowalski, Robert. "Algorithm = Logic + Control." *Communications of the ACM* 22.7 (1979): 424-436.

Lacan, Jacques. *Écrits: The First Complete English Edition*. Trans. Bruce Fink. New York: W.W. Norton & Company, 2006.

Lury, Celia. & Day, Sophie. "Algorithmic Personalisation as a Mode of Individuation." *Theory, Culture & Society* 36.2 (2019): 17-37.

Lafrance, Adrienne. "Machines That Can See Depression on a Person's Face." *The Atlantic*, 19 October, 2015.

Leichsenring, Falk. Hiller, Wolfgang. & Weissberg, Michael et al. "Cognitive Behavioural Therapy and Psychodynamic Psychotherapy: Techniques, Efficacy, and Indications." *American Journal of Psychotherapy* 60.3 (2006): 233-259.

Lucas, M. Gale. Gratch, Jonathan. King, Aisha. et al. "It's only a computer: Virtual humans increase willingness to disclose." *Computer in Human Behaviour* 37 (2014): 94-100.

Malabou, Catherine. *Morphing Intelligence: From IQ Measurement to Artificial Brains.* Trans. Carloyn Shread. New York: Columbia University Press, 2019.

Matzner, Tobias. "The Human is Dead – Long Live the Algorithm! Human-Algorithmic Ensembles and Liberal Subjectivity." *Theory Culture & Society* 36.2 (2019): 123-144.

Maxwell, Bruce. & Tappolet, Christine. "Rethinking Cognitive Mediation: Cognitive Behavioural Therapy and the Perceptual Theory of Emotions." *Philosophy, Psychiatry, & Psychology* 19.1 (2012): 1-12.

Molteni, Megan. "The Chatbot Therapist Will See You Now." *Wired*, 6 July 2017.

Morozov, Evengy. *To Save Everything, Click Here: The Folly of Technological Solutionism.* New York: Public Affairs, 2014.

Roberge, Jonathan. & Seyfert, Robert. "What are Algorithmic Cultures?" *Algorithmic Cultures: Essays on Meaning, Performance and New Technologies.* Eds. Jonathan Roberge & Robert Seyfert. London: Routledge, 2016. 13-35.

Robinson, Ann. "Meet Ellie, the machine that can detect depression." *The Guardian*, 17 September, 2015.

Shapiro, Jordan. "President Obama Wants Every Kid To Learn -- For All The Wrong Reasons." *Forbes* 31 January 2016.

Stiegler, Bernard. *Automatic Society Volume 1: The Future of Work.* Trans. Daniel Ross. London: Polity, 2016.

Thoburn, Nicholas. "Patterns of Production: Cultural Studies after Hegemony." *Theory Cultural & Society* 24.3 (2007): 79-94.

White, Geoff. "Child Advice Chatbots Fail to Sport Sexual Abuse." *BBC* 11 December 2018. https://www.bbc.com/news/technology-46507900.

Wills, Frank. *Beck's Cognitive Therapy: Distinctive Features*. London: Routledge, 2009.

Zuboff, Shoshana. *The Age of Surveillance Capitalism: The Fight for a Human Future at the Frontier of Power*. London: Profile Books, 2019.