A NICE game of Minecraft: philosophical flaws underpinning UK depression guideline nosology

Susan McPherson

ABSTRACT

Categorising mental disorders for purposes of diagnosis, research and practice has historically been justified on philosophical terms as a pragmatic activity; categories which have been subject to wide-ranging philosophical critique have been defended on the grounds that they serve as heuristic devices providing loose representations of shared experiences, not labels for real structures. In acknowledgement of this, there has been increasing recognition that subclassifying multiple discrete forms of persistent depression moves too far away from the notion of a heuristic and that attempts to create more precise categories become less clinically useful. Hence the most recent Diagnostic and Statistical Manual of Mental Disorders (V.5) and International Classification of Diseases (V.11) both group persistent forms of depression together. However, the UK National Institute for Health and Care Excellence has delineated certain subclassifications of persistent depression in its new guideline, which grossly distorts the phenomenology of depression. This approach commits a fundamental philosophical error in conflating absence of knowledge with knowledge of absence. In this sense, the new guideline appears to be engaging in an activity akin to the digital game Minecraft, in which the craft of building structures from units of construction is largely divorced from the laws of physics. The risk of ignoring these philosophical errors and making false claims about scientific plausibility is that the guideline recommendations inevitably represent a highly distorted phenomenology of depression and will be of very little value to patients or practitioners looking for guidance on best possible treatment options.

INTRODUCTION

Minecraft is a computer game with no specific goals to accomplish. The game-world consists of three-dimensional (3D) cubes and objects which the player (Steve) can mine and build into infinitely complex (and logically impossible) structures. Steve sometimes encounters other characters ('mobs'), such as animals and hostile

Correspondence to Dr Susan McPherson, Health and Social Care, University of Essex, Colchester CO4 3SQ, UK; smcpher@essex.ac.uk

creatures; he can 'spawn' and destroy them. While it looks like a harmless game of logical construction, it conveys some worryingly delusive ideas about the real world. The difference between real and imagined structures is at the heart of the age-old debate around categorising mental disorders.

Classification in mental health has had various forms throughout history. Mack and colleagues set out a history of psychiatric classification beginning in 2600 BC with Egyptian references to melancholia and hysteria; through the Ancient Greeks with Hippocrates' phrenitis, mania, melancholia, epilepsy, hysteria and Scythian disease; through the Renaissance period; through to 19th-century psychiatry featuring Pinel (known as the first psychiatrist), Kraepelin (known for observational classification) and Freud (known for classifying neurosis and psychosis).[1]

Although the history of psychiatric classification identifies some common trends such as the labels 'melancholia' and 'hysteria' which have survived millennia, the label 'depression' is relatively new. The earliest usage noted by Snaith is from 1899: 'in simple pathological depression…the patient exhibits a growing indifference to his former pursuits…'.[2] Snaith noted that early 20th-century psychiatrists like Adolf Meyer hoped that 'depression' would come to encompass a broad category under which descriptions of subtypes would emerge. This did not happen until the middle of the 20th century. With the publication of the sixth International Classification of Diseases (ICD) in 1948 and the Diagnostic and Statistical Manual of Mental Disorders (DSM) in 1952 and their subsequent revisions, the latter half of the 20th century has seen depression subtype labels proliferate. In their study of the social determinants of diagnostic labels in depression, McPherson and Armstrong illustrate how the codification of depression subtypes in the latter half of the 20th century has been shaped by the evolving context of psychiatry, including power struggles within the profession, a move to community care and the development of psychopharmacology.[3]

During this period, McPherson and Armstrong describe how subsequent versions of the DSM served as battle-grounds for professional disputes and philosophical quarrels around categorisation of mental disorders. DSM I and DSM II have been described as products of an American Psychiatric Association dominated by psychoanalytic psychiatrists.[4] DSM III and DSM III-R have been described as a radical rejection of psychoanalytic thinking, a 'neo-Kraepelinian revolution', a reference to the observational descriptive techniques of 19th-century psychiatrist Emil Kraepelin who classified mental disorders into two broad categories: 'dementia praecox' and 'manic-depression'.[5] DSM III was seen by some as a turning point in the use of the medical model of mental illness, through provision of specific inclusion and exclusion criteria, and use of field trials and a multiaxial system.[6] These latter technocratic additions to psychiatric labelling served to engender a much closer alignment between psychiatry, science and medicine.

The codification of mental disorders in manuals has been described by Thomas Schacht as intrinsic to the relationship between science and politics and the way in which psychiatrists gain significant social power by aligning themselves to science.[7] His argument drew on Szasz, who saw the mental health establishment as a therapeutic state; Zimbardo, who described psychiatric care as a controlling force; and Foucault, who described the categorisation of the mentally ill as a force for isolating 'the other'. Diagnostic critique has been further developed through a cultural relativist lens in that what Western psychiatrists classify as a depression is constructed differently in other cultures.[8] Considering these limitations, some critics have gone so far as to argue that psychiatric diagnostic systems should be abolished.[9]

Yet architects of DSM manuals have worked hard to ensure the technology of classification is regarded as genuine scientific activity with sound roots in philosophy of science. In their philosophical defence of DSM IV, Allen Frances and colleagues address their critics under the headings 'nominalism vs realism', 'empiricism vs rationalism' and 'categorical vs dimensional'.[10] The implication is that there are opposing stances in which a choice must be made or a middle ground forged by those reasonable enough to recognise the need for pragmatism in the service of clinical utility. The nominalism–realism debate is illustrated using as

BMJ

IME

1

metaphor three different stances a cricket umpire might take on calling strikes and balls. The discussion sets out two of these as extreme views: 'at one extreme…those who take a reductionistically realistic view of the world' versus 'the solipsistic nominalists…might content that nothing exists'. Szasz, who is characterised as holding particularly extreme views, is named as an archetypal solipsist. There is implied to be a degree of arrogance associated with this view in the illustrative example in which the umpire states 'there are no balls and there are no strikes until I call them'. Frances therefore sets up a means of grouping two kinds of people as philosophical extremists who can be dismissed, while avoiding addressing the philosophical problems they pose.

Frances provides little if any justification for the middle ground stance, 'There are balls and there are strikes and I call them as I see them', other than to focus on its clinical utility and the lack of clinical utility in the alternatives 'naïve realism' and 'heuristically barren solipsism'. The natural conclusion the reader is invited to reach is that a middle ground of a heuristic concept is naturally right because it is not extreme and is naturally useful clinically, without specifying in what way this stance is coherent, resolves the two alternatives, and in what way a heuristic construct that is not 'real' can be subject to scientific testing.

Similarly, in discussing the 'categorical vs dimensional', Frances promotes the 'prototype approach'. Those holding opposing views are labelled as 'dualists' or 'dichotomisers'. The prototypical approach is again put forward as a clinically useful middle ground. Illustrations are drawn from natural science: 'a triangle and a square are never the same', inciting the reader to consider science as value-free. The prototypical approach emerges as a natural solution, yet the authors do not address how a diagnostic prototype resolves the issues posed by the two alternatives, nor how a prototype can be subjected to natural science methods.

The argument presented here is not a defence of solipsism or dualism; rather it aims to illustrate that if for pragmatic purposes clinicians and policymakers choose to gloss over the philosophical flaws in classification practices, it is then risky to move beyond the heuristic and apply natural science methods to these constructs adding multiple layers of technocratic subclassification. Doing so is more like playing Minecraft than cricket. The National Institute for Health and Care Excellence (NICE) guideline for depression is taken as an example of the philosophical errors that can follow from playing Minecraft with unsound heuristic devices, specifically subcategories of persistent forms of depression. As well as serving a clinical purpose, diagnosis in medicine is a way of allocating resources for insurance companies and constructing clinical guidelines, which in turn determine rationing within the National Health Service. The consequences for recipients of healthcare are therefore significant; clinical utility is arguably not being served at all and patients are left at risk of poor-quality care.

Heterogeneity of persistent depression

Andrea Jobst and colleagues note that 'because of their chronic clinical course, approximately 40% of CD [chronic depression] patients also fulfil criteria for TRD [treatment resistant depression]… usually defined by the number of non-successful biological treatments'.[11] This position is reflected in the DSM VAmerican Psychiatric Association (2013), the European Psychiatric Association (EPA) guidance and the ICD-11(World Health Organisation, 2018), which all use a 'persistent' depression category, acknowledging a loosely defined mixed group of long-term, difficult-to-treat depressive conditions, often associated with dysthymia and comorbid common mental disorders, various personality traits and psychosocial disability.

In contrast, the NICE 2018 draft guideline separates treatments into those for 'new episodes' of depression: 'further-line' treatment of depression (equivalent to TRD), CD and 'depression with co-morbidities'. The latter is subdivided into treatments for 'complex depression' and 'psychotic depression'. These categories and subcategories introduce an unfortunate sense of certainty as though these labels represent real things. An analysis follows of how these definitions play out in terms of grouping of randomised controlled trials in the NICE evidence review. Specifically, the analysis reveals the overlap between populations in trials which have been separated into discrete categories, revealing significant limitations to the utility of the category labels.

The NICE definition of CD requires trial samples to meet the criteria for major depressive disorder (MDD) for 2 years. Dysthymia and double depression (MDD superimposed on dysthymia) were included. If 75% of the trial population met these criteria, the trial was reviewed in the CD category.[12] The definition of TRD (or 'further-line treatments') required that the trial sample had demonstrated a 'limited response to previous treatment' and randomised to the further-line treatment at this point. If 80% of the trial participants met these criteria, it was reviewed in the TRD category.[13] Complex depression was defined as 'depression co-existing with personality disorder'. To be classed as complex, 51% of trial participants had to have personality disorder (PD).[14]

It is immediately clear from these definitions that there is a potential problem with attempting to categorise trial populations into just one of these categories. These populations are likely to overlap, whether or not a trial protocol sets out to explicitly record all of this information. The analysis below will illustrate this using examples from within the NICE review.

Cataloguing complexity in trial populations

Within the category of further-line treatments (TRD), 64 trials were reviewed. Comparisons within these trials were further subcategorised into 'dose escalation strategies', 'augmentation strategies' and 'switching strategies'. In drilling down by way of illustration, this analysis considers the 51 trials in the augmentation strategy evidence review. Of these, two were classified by the reviewers as also fulfilling the criteria for CD but were not analysed in the CD category (Study IDs: Fonagy 2015 and Kocsis 2009[15]). About half of the trials (23/51) did not report the mean duration of episode, meaning that it is not possible to know what percentage of participants also met the criteria for CD. Of trials that did report episode duration, 17 reported a mean duration longer than 24 months. While the standard deviations varied in size or were unreported, the mean indicates a good likelihood that a significant proportion of the participants across these 51 trials met the criteria for CD.

Details of baseline employment, trauma history, suicidality, physical comorbidity, axis I comorbidity and PD (all clinical indicators of complexity, severity and chronicity) were not collated by NICE. For the present analysis, all 51 publications were examined and data compiled concerning clinical complexity in the trial populations. Only 14 of 51 trials report employment data. Of those that do, unemployment ranges from 12% to 56% across trial samples. None of the trials report trauma history. About half of the trials (26/51) excluded people who were

considered a suicide risk; the others did not.

A large proportion of trials (30/51) did not provide any data on axis 1 comorbidity. Of these, 18 did not exclude any diagnoses, while 12 excluded some (but not all) disorders. The most common diagnoses excluded were psychotic disorders, substance or alcohol abuse, and bipolar disorder (excluded in 26, 25 and 23 trials, respectively). Only 7 of 51 trials clearly stated that all axis 1 diagnoses were excluded. This leaves only 13 studies providing any data about comorbidity: of these, 9 gave partial data on one or two conditions, while 4 reported either the mean number of disorders (range 1.96–2.9) or the percentage of participants (range 68.1–96.7) with any comorbid diagnosis (Nierenberg 2003a, Nierenberg 2006, Watkins 2011a, Town 2017[15]).

The majority of trials (46/51) did not report the prevalence of PD. Many stated PD as an exclusion criterion but without defining a threshold for exclusion. For example, PD could be excluded if it 'impacted' the depression, if it was 'significant', 'severe' or 'persistent'. Some excluded certain PDs (such as antisocial or borderline) and not others but without reporting the prevalence of those not excluded. In the five trials where prevalence was clear, prevalence ranged from 0% (Ravindran 2008a[15]), where all PDs were excluded, to 87.5% of the sample (Town 2017[15]). Two studies reported the mean number of PDs: 2.0 (Nierenberg 2003a) and 0.85 (Watkins 2011a[15]).

The majority of trials (43/51) did not report the prevalence of physical illness. Many stated illness as an exclusion criterion, but the definitions and thresholds were vague and could be interpreted in different ways. For example, illness could be excluded if it was 'unstable', 'serious', 'significant', 'relevant', or would 'contraindicate' or 'impact' the medication. Of the eight trials reporting information about physical health, there was a wide variation. Four reported prevalence varying from 7.6% having a disability (Eisendrath 2016[15]) to 90.9% having an illness or disability (Town 2017[15]). Four used scales of physical health: two indicating mild problems (Nierenberg 2006, Lavretsky 2011[15]) and two indicating moderately high levels of illness (Thase 2007, Fang 2010[15]).

The NICE review also divided trial populations into a dichotomy of 'more severe' and 'less severe' on the grounds that this would be a clinically useful classification for general practitioners. NICE applied a bespoke methodology for creating this dichotomy, abandoning validated measure thresholds in order first to generate two 'homogeneous' groups to 'facilitate analysis', and second to create an algorithm to 'read across' different measures (such as the Beck Depression Inventory, the Hamilton Rating Scale for Depression (HRSD) and the Montgomery-Asberg Depression Rating Scale).[16] Examining trials which use more than one of these measures reveals problems in the algorithm. Of the 51 trials, there are 6 instances in which the study population falls into NICE's more severe category according to one measure and into the less severe category according to another. In four of these trials, NICE chose the less severe category (Souza 2016, Watkins 2011a, Fonagy 2015, Town 2017[15]). The other two trials were designated more severe (Barbee 2011, Dunner 2007[15]). Only 17 of 51 trials reported two or more depression scale measures, leaving much unknown about whether other study populations could count as both more severe and less severe.

Absence of knowledge or knowledge of absence?

A key philosophical error in science is to confuse an absence of knowledge with knowledge of absence. It is likely that some of the study populations deemed lacking in complexity or severity could actually have high degrees of complexity and/or severity. Data to demonstrate this may either fall foul of a guideline committee decision to prioritise certain information over other conflicting information (as in the severity algorithm); the information may be non-existent as it was not collected; it may be somewhere in the publication pipeline; or it may be sitting in a database with a research team that has run out of funds for supplementary analyses. Wherever those data are or are not, their absence from published articles does not define the phenomenology of depression for the patients who took part. As a case in point, data from the Fonagy 2015 trial presented at conferences but not published reveal that PD prevalence data would place the trial well within the NICE complex depression category, and that the sample had high levels of past trauma and physical condition comorbidity. The trial also meets the guideline criteria for CD according to the guideline's own appendices.[17] Reported axis 1 comorbidity was high (75.2% had anxiety disorder, 18.6% had substance abuse disorder, 13.2% had eating disorder).[18] The mean depression scores at baseline were 36.5 on the Beck Depression Inventory and 20.1 on the HRSD (severe and very severe, respectively, according to published cut-off scores). NICE categorised this population as less severe TRD, not CD and not complex.

CONCLUSION: CLINICAL FUTILITY OF CUBOIDISM

There are philosophical flaws in the NICE guideline's overly technocratic approach to subclassification of depression. Information available in a journal publication about study populations is not a clear and fulsome representation of the varied experiences of the individuals concerned, particularly as journal word limits often prevent fuller reporting of trial data. Moreover, carving up human experience into more and more cuboids does not make a heuristic labelling device more clinically useful; it does the opposite. NICE's cuboidism concerning persistent depression deviates markedly from the ICD and DSM manuals, as well as the EPA guideline and the American Psychological Association draft depression guideline.[19] The result is that the UK guideline for depression represents an impossible structure that would horrify a structural engineer. While science should be employed to help us understand reality, this approach to the philosophy of science obscures realities encountered by patients and practitioners because the underpinning philosophy has warped into a gameworld logic.

This is one among a number of epistemological concerns expressed by professional bodies and patient groups about the NICE draft depression guideline and which led to a joint statement warning that the guideline will be unfit for purpose.[20] Unlike Minecraft, NICE has a specific goal to accomplish: guidance which enables best possible patient care which will ultimately dictate rationing of services for UK patients. With a new draft scheduled for publication in February 2020, there may yet be scope for NICE to move away from a warped 3D gameworld and reconsider professional and patient concerns about the clinical futility of cuboidism.

OPEN ACCESS

(🔴) Check for updates

NOTES

1. Avram H. Mack et al. (1994), "A Brief History of Psychiatric Classification: From the Ancients to DSM-IV," *Psychiatric Clinics* 17, no. 3: 515–9.
2. R. P. Snaith (1987), "The Concepts of Mild Depression," *British Journal of Psychiatry* 150, no. 3: 387.
3. Susan McPherson and David Armstrong (2006), "Social Determinants of Diagnostic Labels in Depression," *Social Science & Medicine* 62, no. 1: 52–7.
4. Gerald N. Grob (1991), "Origins of DSM-I: A Study in Appearance and Reality," *The American Journal of Psychiatry*: 421–31.
5. Wilson M. Compton and Samuel B. Guze (1995), "The Neo-Kraepelinian Revolution in Psychiatric Diagnosis," *European Archives of Psychiatry and Clinical Neuroscience* 245, no. 4: 198–9.
6. Gerald L. Klerman (1984), "A Debate on DSM-III: The Advantages of DSM-III," *The American Journal of Psychiatry*: 539–42.
7. Thomas E. Schacht (1985), "DSM-III and the Politics of Truth," *American Psychologist*: 513–5.
8. Daniel F. Hartner and Kari L. Theurer (2018), "Psychiatry Should Not Seek Mechanisms of Disorder," *Journal of Theoretical and Philosophical Psychology* 38, no. 4: 189–204.
9. Sami Timimi (2014), "No More Psychiatric Labels: Why Formal Psychiatric Diagnostic Systems Should Be Abolished," *Journal of Clinical and Health Psychology* 14, no. 3: 208–15.
10. Allen Frances et al. (1994), "DSM-IV Meets Philosophy," *The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine* 19, no. 3: 207–18.
11. Andrea Jobst et al. (2016), "European Psychiatric Association Guidance on Psychotherapy in Chronic Depression Across Europe," *European Psychiatry* 33: 20.
12. National Institute for Health and Care Excellence (2018), Depression in Adults: Treatment and Management. Draft for Consultation, https://www.nice.org.uk/guidance/gid-cgwave0725/documents/full-guideline-updated, 507.
13. Ibid., 351–62.
14. Ibid., 597.
15. Note that in order to refer to specific trials reviewed in the guideline, rather than the full citation, the Study IDs from column A in appendix J5 have been used. See www.nice.org.uk/guidance/gid-cgwave0725/documents/addendum-appendix-9 for details and full references.
16. National Institute for Health and Care Excellence (2018), Depression in Adults: Treatment and Management. Second Consultation on Draft Guideline – Stakeholder Comments Table, https://www.nice.org.uk/guidance/gid-cgwave0725/documents/consultation-comments-and-responses-2, 420–1.
17. National Institute for Health and Care Excellence (2018), Depression in Adults, appendix J5.
18. Peter Fonagy et al. (2015), "Pragmatic Randomized Controlled Trial of Long-Term Psychoanalytic Psychotherapy for Treatment-Resistant Depression: The Tavistock Adult Depression Study (TADS)," *World Psychiatry* 14, no. 3: 312–21.
19. American Psychological Association (2018), *Clinical Practice Guideline for the Treatment of Depression in Children, Adolescents, and Young, Middle-aged, and Older Adults. Draft*.
20. Jacqui Thornton (2018), "Depression in Adults: Campaigners and Doctors Demand Full Revision of NICE Guidance," *BMJ* 361: k2681.

BIBLIOGRAPHY

American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Health Disorders: DSM-5*, 5th ed. Washington DC: American Psychiatric Publishing, 2013.

——————. *Clinical Practice Guideline for the Treatment of Depression in Children, Adolescents, and Young, Middle-aged, and Older Adults. Draft*, 2018.

Compton, Wilson M, and Samuel B Guze. "The Neo-Kraepelinian Revolution in Psychiatric Diagnosis." *Eur Arch Psychiatry Clin Neurosci* 245, no. 4 (1995): 196–201.

Fonagy, Peter, Felicitas Rost, Jo-Anne Carlyle, Susan McPherson, Rachel Thomas, Pasco, Fearon, David Goldberg, and David Taylor. "Pragmatic Randomized Controlled Trial of Long-Term Psychoanalytic Psychotherapy for Treatment-Resistant Depression: The Tavistock Adult Depression Study (TADS)." *World Psychiatry* 14, no. 3 (2015): 312–21.

Frances, Allen, Leslie Forman, Avram H Mack, and Micheal B First. "Thomas A Widiger, and Wendy Wakefield Davis. "DSM-IV Meets Philosophy." *J Med Philos* 19, no. 3 (1994): 207–18.

Grob, Gerald N. *"Origins of DSM-I: A Study in Appearance and Reality." The American Journal of Psychiatry*. USA: American Psychiatric Association, 1991.

Hartner, Daniel F, and Kari L Theurer. "Psychiatry Should Not Seek Mechanisms of Disorder." *J Theor Philos Psychol* 38, no. 4 (2018): 189–204.

Jobst, Andrea, Eva-Lotta Brakemeier, Anna Buchheim, Franz Caspar, Pim Cuijpers, Klaus P Ebmeier, Peter G Falkai., *et al*. "European Psychiatric Association Guidance on Psychotherapy in Chronic Depression Across Europe." *Eur Psychiatry* 33 (2016): 18–36.

Klerman, Gerald L. "A debate on DSM-III: the advantages of DSM-III." *Am J Psychiatry* (1984).

Mack, Avram H, Leslie Forman, Rebekah Brown, and Allen Frances. "A Brief History of Psychiatric Classification: From the Ancients to DSM-IV." *Psychiatric Clinics* 17, no. 3 (1994): 515–23.

McPherson, Susan, and David Armstrong. "Social Determinants of Diagnostic Labels in Depression." *Soc Sci Med* 62, no. 1 (2006): 50–8.

National Institute for Health and Care Excellence. "Depression in Adults: Treatment and Management. Draft for Consultation.", 2018a. https://www.nice.org.uk/guidance/gid-cgwave0725/documents/full-guideline-updated

——————. "Depression in Adults: Treatment and Management. Second Consultation on Draft Guideline – Stakeholder Comments Table.", 2018b. https://www.nice.org.uk/guidance/gid-cgwave0725/documents/consultation-comments-and-responses-2

Schacht, Thomas E. *"DSM-III and the Politics of Truth." American Psychologist*. USA: American Psychological Association, 1985.

Snaith, R P. "The concepts of mild depression." *Br J Psychiatry* 150, no. 3 (1987): 387–93.

Thornton, Jacqui. "Depression in Adults: Campaigners and Doctors Demand Full Revision of NICE Guidance." *BMJ* 361 (2018): k2681.

Timimi, Sami. "No More Psychiatric Labels: Why Formal Psychiatric Diagnostic Systems Should Be Abolished." *Int J Clin Health Psychol* 14, no. 3 (2014): 208–15.

World Health Organisation. *ICD-11 International Classification of Diseases*. Geneva: World Health Organisation, 2018.