Mental Disorder: A Conceptual Engineering Approach

Luca Malatesti*
lmalatesti@ffri.uniri.hr
https://orcid.org/0000-0001-9911-0637

https://doi.org/10.31192/np.23.3.12
UDK / UDC: 159.972:159.9.016
Izvorni znanstveni rad /
Original scientific paper
Primljeno / Received:
7. svibnja 2025. / May 7, 2025
Prihvaćeno / Accepted:
23. rujna 2025. / Sep 23, 2025

This paper explores the possibility of defining the concept of mental disorder through conceptual engineering. This method proposes shaping a concept according to the goals it should serve. I argue that mental disorders should be understood as unitary mental conditions involving harm. Such harm should be assessed against justifiable standards and arise from factors beyond an individual's ordinary control. The paper examines this proposal through five components of the proposed concept of mental disorder: unity, harm, normative standards, factors beyond control, and mental nature. The aim is to provide a framework that could support clearer diagnosis, ethical justification of treatment, and improved theoretical understanding.

Key words: *mental disorder, conceptual engineering, control, harm, normative standards.*

* Luca Malatesti, PhD, Full Prof., University of Rijeka, Faculty of Humanities and Social Sciences, Department of Philosophy and Division of Cognitive Sciences; Address: Sveučilišna avenija 4, HR-51000 Rijeka, Croatia.

Introduction

The definition of the concept of mental disorder is a central problem in the philosophy of psychiatry.[1] How we define the concept matters because it affects diagnosis, treatment, and classification in manuals such as the *Diagnostic and Statistical Manual of Mental Disorders* (hereafter DSM)[2] and the *International Classification of Diseases* (hereafter ICD).[3] A broad definition pathologizes normal variation; a narrow one may deny access to needed care.

This paper proposes a general definition of the concept of mental disorder. This proposal is based on conceptual engineering, a method increasingly discussed and used in contemporary philosophy.[4] Rather than merely describing the use of concepts, this approach recommends constructing or revising them to satisfy justified goals.[5] I argue that mental disorder should be engineered as a unitary mental condition involving harm. This harm should be identified in relation to deviation from properly justified normative standards and caused by factors beyond the individual's ordinary control.

Around this proposal, I organise a survey of open issues concerning the core components of the concept of mental disorder: unity, harm, justified standards, factors beyond control, and mentality. Each raises unresolved conceptual and empirical questions. I review the relevant current debates in philosophy and psychiatry to suggest paths for further conceptual inquiry. The aim is to clarify central debates, expose conceptual tensions, and promote a focused research agenda within the philosophy of psychiatry.

In the paper, I proceed as follows. Following the method of conceptual engineering, Section 1 sets out the goals the concept of mental disorder should serve. Section 2 introduces the requirement that mental disorders should be unitary conditions and examines the conceptual challenges this poses. In Section 3, I argue that mental disorders must involve harm to advance psychiatry's core practical aims. Moreover, I maintain that psychiatrically relevant harm

[1] Sam WILKINSON, *Philosophy of Psychiatry: A Contemporary Introduction*, New York and London, Routledge, 2023.

[2] AMERICAN PSYCHIATRIC ASSOCIATION, *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*, Washington, D.C., American Psychiatric Publishing, 2013.

[3] WORLD HEALTH ORGANIZATION, *International Classification of Diseases, 11th Revision (ICD-11)*, Geneva, World Health Organization, 2019.

[4] Georg BRUN, Explication as a Method of Conceptual Re-engineering, *Erkenntnis*, 81(2016) 6, 1211–1241; Herman CAPPELEN, *Fixing Language. An Essay on Conceptual Engineering*, Oxford, Oxford University Press, 2018.

[5] This method can be exemplified with a case outside philosophy. To serve the goals of classification, research, and theoretical explanation, in 2006 the International Astronomical Union formally redefined the ordinary concept of planet in our Solar System as a body that (1) orbits the Sun, (2) is nearly round due to its own gravity, and (3) has cleared its orbital neighbourhood. See: INTERNATIONAL ASTRONOMICAL UNION, Resolution B5: "Definition of a Planet in the Solar System", adopted 24 August 2006, IAU General Assembly, Prague, 2006, https://drive.google.com/file/d/1ckh-gW-oMYBbxKd8PgVyKwdMg1ywhA7h/view?pli=1 (23.09.2025).

should be identified as involving deviation from properly justified normative standards. Section 4 adds a further condition: psychiatric harm must stem from a condition that bypasses ordinary control. I present two currently dominant approaches that satisfy this requirement: one appeals to dysfunction, the other to inability or limited ability. I note unresolved issues in both. Section 5 considers the final condition: that mental disorders must be mental. There I outline challenges in clarifying how mentality should feature in the engineering of the concept of mental disorder.

1. Goals for engineering the concept of mental disorder

The main positions in the debate on the definition of the concept of mental disorder are naturalism, normativism, and hybrid accounts.[6] Each treats disorder as a deviation from certain standards but differs in how those standards are defined. Naturalists appeal to standards from the natural sciences, such as anatomical and physiological ones[7], biological survival and reproductive advantage[8], or evolutionary fixed functions.[9] Often, these standards are taken as objective and independent of social or personal values. Normativists, instead, appeal to value-laden standards tied, for instance, to human flourishing[10] or the preconditions of agency.[11] Hybrid views combine these elements. The most influential hybrid account is the harmful dysfunction analysis by Jerome Wakefield.[12] This account links dysfunction to failure of psychological mechanisms that are based on biological evolution and harm to cultural evaluation.

Recent debates in the analytic philosophy of psychiatry manifest methodological divisions.[13] Traditionally, philosophers used conceptual analysis to clarify the concept of mental disorder. This method aims to define concepts

[6] Tim THORNTON, *Essential Philosophy of Psychiatry*, Oxford, Oxford University Press, 2007; Jonathan Y. TSOU, *Philosophy of Psychiatry*, Cambridge, Cambridge University Press, 2021.

[7] Thomas S. SZASZ, The Myth of Mental Illness, *American Psychologist*, 15 (1960) 2, 113–118.

[8] Robert E. KENDELL, The Concept of Disease and Its Implications for Psychiatry, *The British Journal of Psychiatry*, 127 (1975) 305–315.

[9] Christopher BOORSE, A Second Rebuttal on Health, *The Journal of Medicine and Philosophy*, 39 (2014) 6, 683–724.

[10] Christopher MEGONE, Aristotle's Function Argument and the Concept of Mental Illness, *Philosophy, Psychiatry & Psychology*, 5 (1998) 3, 187–201.

[11] K. W. M. FULFORD, *Moral Theory and Medical Practice*, Cambridge, Cambridge University Press, 1989.

[12] Jerome C. WAKEFIELD, The Concept of Mental Disorder: On the Boundary Between Biological Facts and Social Values, *American Psychologist*, 47 (1992) 3, 373–388.

[13] Bengt BRÜLDE, On Defining 'Mental Disorder': Purposes and Conditions of Adequacy, *Theoretical Medicine and Bioethics*, 31 (2010) 1, 19–33; Rachel V. COOPER, The Concept of Disorder Revisited: Robustly Value-Laden Despite Change, *Aristotelian Society Supplementary Volume*, 94 (2020) 1, 141–161; Maël LEMOINE, Defining Disease beyond Conceptual Analysis: An Analysis of Conceptual Analysis in Philosophy of Medicine, *Theoretical Medicine and Bioethics*, 34 (2013) 4, 309–325.

by accurately describing their ordinary or expert usage, guided by intuitions concerning the appropriate use of linguistic expressions.[14] Yet, its perceived lack of progress, shifts in expert usage[15], and broader criticisms of the method of conceptual analysis[16] have led to growing scepticism about its value in this domain.[17]

Motivated by these challenges, some philosophers have turned to an alternative method: conceptual engineering.[18] This method, rather than aiming to describe a concept, *prescribes* one that would help achieve certain preferred aims.[19] Thus, satisfactorily engineering a concept involves two main steps. First, it is essential to demonstrate that certain aims are worth pursuing. Second, a concept should be engineered to effectively achieve these aims.

To engineer the concept of mental disorder, we must first identify the aims this concept is meant to serve. This can be achieved by examining recent proposals for engineering this concept, which explicitly list the intended aims. Additionally, the philosophical debate on defining this concept, whether based on ordinary or expert use, sometimes implicitly offers valuable insights into the goals the concept should fulfil. Finally, we should also consider the goals reflected in official expert documents, institutional aims, and the psychiatric or medical tradition.[20]

The concept of mental disorder should clearly distinguish between pathological conditions and non-pathological ones.[21] Thus, the concept should enable us to distinguish mental disorder from mental health. This is a fundamental function of the concept. Moreover, the concept of mental disorder should also distinguish mental disorders from deviations that are not relevant to psychiatry. This requirement is justified by the history of psychiatry, which has seen the medicalisation, often repressive, of conditions involving mere deviations

[14] Michael BEANEY, Thomas RAYSMITH, Analysis, in: Edward N. ZALTA, Uri NODELMAN (eds.), *The Stanford Encyclopedia of Philosophy*, Stanford, Metaphysics Research Lab, 2024. https://plato.stanford.edu/archives/fall2024/entries/analysis/ (21.09.2025).
[15] Cooper, *The Concept of Disorder Revisited…*
[16] Herman CAPPELEN, *Philosophy without Intuitions*, Oxford, Oxford University Press, 2012.
[17] Lemoine, *Defining Disease Beyond Conceptual Analysis…*
[18] Mia BITURAJAC, Marko JURJAKO, Reconsidering Harm in Psychiatric Manuals within an Explicationist Framework, *Medicine, Health Care and Philosophy*, 25 (2022) 2, 239–249; Sanja DEMBIĆ, *Philosophy of Mental Disorder: An Ability-Based Approach*, New York, Routledge, 2023; Marko JURJAKO, Luca MALATESTI, In What Sense Are Mental Disorders Brain Disorders? Explicating the Concept of Mental Disorder within RDoC, *Phenomenology and Mind*, 18 (2020) 182–198; Elisabetta LALUMERA, 'Are Mental Disorders Brain Disorders?' is a Question of Conceptual Choice, *Philosophical Psychology*, 37 (2023) 3, 1–13.
[19] Elisabetta LALUMERA, Conceptual Engineering of Medical Concepts, in: Manuel Gustavo ISAAC, Steffen KOCH, Kevin SCHARP (eds.), *New Perspectives on Conceptual Engineering*, Cham, Springer, forthcoming, https://link.springer.com/book/9783031985195 (21.09.2025).
[20] Lalumera, *Conceptual Engineering of Medical Concepts...*
[21] See: Dembić, *Philosophy of Mental Disorder…*, 20–21.

from social norms.[22] Additionally, the importance of this distinction is underscored by challenges from anti-psychiatric movements[23] and specific calls from advocacy groups to reassess, in certain cases, what constitutes a disorder versus natural neuropsychological variation.[24]

The concept of mental disorder should clarify how mental disorders differ from other medical conditions. This requires specifying the sense in which they involve mental phenomena.[25] This aim underscores psychiatry's disciplinary identity, grounded in its focus on mental conditions, and distinguishing it from medical specialties that address bodily disorders.

A further aim noted in the literature is that the concept of mental disorder should foster effective communication among experts from differing theoretical traditions. As Alfredo Gaete points out, the concept must remain independent of any specific theoretical framework.[26] For instance, effective communication cannot be achieved if mental disorders are defined solely through a psychoanalytic perspective, such as unresolved childhood conflicts. Simultaneously, the concept should align with current scientific research and clinical practice aimed at treating mental disorders and remain adaptable to future developments.

In addition to these theoretical aims, the concept of mental disorder must also serve practical goals. Sanja Dembić argues convincingly that, when correctly applied to an individual's condition, the concept provides a *pro tanto* reason for that person to seek medical treatment.[27] This reason is not absolute; it may be outweighed by other factors, but it establishes a strong basis for seeking help. When it comes to treatment, however, a nuanced approach is essential. While medical or therapeutic interventions are often beneficial, they are not the only workable solutions. Some patients may prefer or succeed with alternative forms of management, such as self-care, community support, or lifestyle adjustments. Thus, while having a mental disorder may ground a *pro tanto* reason for medical intervention, this should not be overly prescriptive. Instead, it should allow for personal autonomy, respecting individual preferences in care.

Furthermore, the concept should provide society with a *pro tanto* reason to offer (or, at times, require) medical treatment. However, this objective must also be balanced with the protection of individual rights. As Mia Biturajac and

[22] Andrew SCULL, *Madness in Civilization: A Cultural History of Insanity, from the Bible to Freud, from the Madhouse to Modern Medicine*, Princeton (N.J.), Princeton University Press, 2015.

[23] Rob WHITLEY, The Antipsychiatry Movement: Dead, Diminishing, or Developing? *Psychiatric Services*, 63 (2012) 10, 1039–1041.

[24] Robert CHAPMAN, *Empire of Normality: Neurodiversity and Capitalism*, London, Pluto Press, 2023.

[25] See: Dembić, *Philosophy of Mental Disorder…*, 28–29.

[26] Alfredo GAETE, The Concept of Mental Disorder: A Proposal, *Philosophy, Psychiatry & Psychology*, 15 (2008) 4, 327–339.

[27] See: Dembić, *Philosophy of Mental Disorder…*, 63.

Marko Jurjako persuasively argue, another important aim in the engineering of the concept of mental disorder is to avoid unnecessary medicalisation.[28] The goal here is to safeguard individual freedoms to the greatest extent possible.

The next section begins by demonstrating why the notion of unity should be a component of the concept of mental disorder.

2. Mental disorders as unitary conditions

The concept of mental disorder should concern unitary conditions. By this I mean that all individuals, or the same individual over time, with a condition classified as a mental disorder should share a significant similarity that supports the description, explanation, prediction, and treatment of that condition. This requirement meets the fundamental aims of the scientific study of disorders and clinical practice in their treatment. Moreover, it aligns with the principle of effective communication among experts. At this level of generality, the requirement of unity does not align with any specific theoretical framework. Instead, it relates the concept of mental disorder to existing psychiatric classification systems, such as DSM or ICD, while allowing for future revisions.

Articulating unity more precisely raises challenges, the most fundamental of which is determining what defines a valid scientific or medical classification.[29] This question is linked to the philosophical debate on natural kinds.[30] In philosophy, natural kinds are categories that group things by essential, inherent characteristics, independent of human interpretation. Elements like gold or water, with objective natural properties, are typical examples. The question for psychiatry, then, is whether mental disorders should be classified as natural kinds, with shared, objective characteristics, or viewed through a more flexible, pragmatic outlook.[31]

In psychiatry, classification debates often mirror these philosophical divisions but add further complexity. A central issue in psychiatry concerns the structure of classification. Should disorders be categorised into distinct kinds, or understood as variations along continuous dimensions? The categorical approach treats mental disorders as discrete, akin to diseases in general medicine. In contrast, the dimensional view sees mental conditions as differing in degree rather than kind, reflecting variations in traits shared with the general population. This debate shapes diagnostic criteria, treatment protocols, and research

[28] See: Biturajac, Jurjako, *Reconsidering Harm in Psychiatric Manuals*…, 245.
[29] Lara K. KUTSCHENKO, In Quest of »Good« Medical Classification Systems, *Medicine Studies*, 3 (2011) 1, 53–70.
[30] Zdenka BRZOVIĆ, Natural Kinds, in: *Internet Encyclopedia of Philosophy*, https://www.iep.utm.edu/nat-kind/ (21.09.2025).
[31] Harold KINCAID, Jacqueline A. SULLIVAN, *Classifying Psychopathology: Mental Kinds and Natural Kinds*, Cambridge (Mass.), MIT Press, 2014.

strategies. Choosing between these approaches affects how mental phenomena are conceptualised. Emil Kraepelin, a key figure in psychiatric classification, emphasized categorizing mental illnesses based on observable symptoms and clinical features.[32] His work laid the groundwork for systems like the DSM and ICD, which rely on categorical classification to diagnose mental disorders, offering clarity and uniformity.

However, categorical systems have been criticized for oversimplifying mental health by failing to capture the fluidity and overlap of various mental disorders.[33] High comorbidity rates among mental disorders challenge this approach, as overlapping symptoms may suggest shared underlying mechanisms rather than distinct conditions.[34] This has led to dimensional models, such as the Hierarchical Taxonomy of Psychopathology (HiTOP), which recognises symptoms on a continuum.[35] The DSM-5, reflecting this shift, incorporates a hybrid model that combines categorical and dimensional approaches, acknowledging that while categories aid treatment, they may not capture the full complexity of mental health.[36]

Another significant debate concerns the level of description needed to establish unity in mental disorders. Categorical systems like the DSM and ICD are based on clusters of symptoms, behaviours, mental states, personality traits, and physical signs. Critics argue that these classifications are limited by their lack of causal information.[37] Some suggest that mental disorder unity should be grounded in biological causes, asserting that only similarities in brain function or neurobiological mechanisms provide a solid basis for classification.[38] For example, the *Research Domain Criteria* (RDoC) framework emphasises a dimensional, integrative approach that combines biological, psychological, and

[32] Kenneth S. KENDLER, Assen JABLENSKY, Kraepelin's Concept of Psychiatric Illness, *Psychological Medicine*, 41 (2011) 6, 1119–1126.

[33] Nick HASLAM, Elise HOLLAND, Peter KUPPENS, Categories versus Dimensions in Personality and Psychopathology: A Quantitative Review of Taxometric Research, *Psychological Medicine*, 42 (2012) 5, 903–920.

[34] Tanja M. BRÜCKL et al., The Biological Classification of Mental Disorders (BeCOME) Study: A Protocol for an Observational Deep-Phenotyping Study for the Identification of Biological Subtypes, *BMC Psychiatry*, 20 (2020) 1, 213.

[35] Roman KOTOV et al., The Hierarchical Taxonomy of Psychopathology (HiTOP): A Dimensional Alternative to Traditional Nosologies, *Journal of Abnormal Psychology*, 126 (2017) 4, 454–477.

[36] American Psychiatric Association, *DSM-5…*

[37] Kenneth S. KENDLER, Peter ZACHAR, Carl CRAVER, What Kinds of Things Are Psychiatric Disorders?, *Psychological Medicine*, 41 (2011) 6, 1143–1150.; Dominic MURPHY, Psychiatry and the Concept of Disease as Pathology, in: Matthew R. BROOME, Lisa BORTOLOTTI (eds.), *Psychiatry as Cognitive Neuroscience: Philosophical Perspectives*, Oxford, Oxford University Press, 2009, 103–117.

[38] Jonathan Y. TSOU, Natural Kinds, Psychiatric Classification and the History of the DSM, *History of Psychiatry*, 27 (2016) 4, 406–424.

social factors.[39] Developed by the National Institute of Mental Health, RDoC seeks to understand mental disorders through underlying functional domains, such as cognition, emotion, and motivation, rather than relying solely on symptom-based diagnostic categories. This framework supports the development of targeted treatments grounded in empirical findings across multiple levels of analysis, from genetic to behavioural.

Others propose that unity may lie in consistent mental symptoms or behavioural patterns across cases. The network approach to mental disorders suggests that symptom interactions form meaningful connections essential for research, diagnosis, and treatment. Denny Borsboom argues that mental disorders should be modelled as interconnected nodes within a network of symptoms that causally influence and exacerbate each other.[40]

Another challenging aspect of psychiatric classification is understanding how the psychometric concepts of validity and reliability interrelate in establishing effective taxonomies of mental disorder.[41] Validity refers to how accurately a classification or, more precisely, the measurements within it, reflects the mental disorders it aims to describe. Reliability refers to the consistency with which the classification can be applied across various contexts and by different clinicians. A reliable measure produces consistent diagnoses but may lack validity if it does not accurately represent the disorder. Conversely, a highly valid measure that captures the complexity of a mental disorder may have low reliability if it is challenging to apply consistently. Balancing these concepts remains an ongoing challenge in psychiatric theory and practice.

The question of how to explain the unity of mental disorders, and its role, intersects with debates on whether and how psychiatric classification should shape diagnosis and treatment.[42] Some critics question the utility of unity if it is too closely tied to diagnosis, as psychiatric classifications can shape and influence individuals' behaviours, identities, and self-understanding.[43] Additionally, diagnoses may lead to stigma and do not necessarily result in better therapeutic outcomes.[44]

[39] Bruce N. CUTHBERT, The RDoC Framework: Facilitating Transition from ICD/DSM to Dimensional Approaches that Integrate Neuroscience and Psychopathology, *World Psychiatry*, 13 (2014) 1, 28–35; Thomas R. INSEL et al., Research Domain Criteria (RDoC): Toward a New Classification Framework for Research on Mental Disorders, *American Journal of Psychiatry*, 167 (2010) 7, 748–751.

[40] Denny BORSBOOM, A Network Theory of Mental Disorders, *World Psychiatry*, 16 (2017) 1, 5–13.

[41] Haslam, Holland, Kuppens, *Categories versus Dimensions...*

[42] John Z. SADLER, Diagnosis/Antidiagnosis, in: Jennifer RADDEN (ed.), *The Philosophy of Psychiatry: A Companion*, New York, (N.Y.), Oxford University Press, 2004, 163–179.

[43] Ian HACKING, The Looping Effects of Human Kinds, in: Dan SPERBER, David PREMACK, Ann James PREMACK (eds.), *Causal Cognition: A Multidisciplinary Debate*, Oxford, Oxford University Press, 1996, 351–383.

[44] Sami TIMIMI, No More Psychiatric Labels: Why Formal Psychiatric Diagnostic Systems Should Be Abolished, *International Journal of Clinical and Health Psychology*, 14 (2014) 3,

These significant worries about the use of diagnosis, however, do not undermine the theoretical value of aiming to individuate mental disorders at a level of unity that enables the identification of shared processes across individuals or within individuals over time. Such unity supports explanation, treatment, and prediction of disorder progression by reference to general shared features across cases, as is typical of scientific practice in other medical fields. For instance, in oncology cancer staging systems classify tumours and guide expectations about development and treatment.

To recapitulate, I have defended the claim that the notion of unity should be part of the engineered definition of the concept of mental disorder. Such a general conceptual requirement, however, highlights open conceptual and empirical research avenues concerning the substantiation of the relevant concept of unity.

In any case, unity alone is not enough. Even if a condition is unitary under a satisfactory construal, it may not be pathological. For example, a stable pattern of introversion may involve unified cognitive, affective, and behavioural traits, as well as identifiable neurological correlates, yet it is not a disorder. We can imagine such unitary conditions shared by non-pathological individuals. The next section shows that harm is also essential in engineering the concept of mental disorder.

3. Psychiatrically relevant harm

The debate surrounding the concept of mental disorder often centres on whether harm should be part of its definition. Some scholars argue that the notion of mental disorder inherently concerns conditions that harm the individual.[45] Others emphasize that harm is central in clinical contexts.[46] However, some, notably Christopher Boorse[47], argue that harm is not an inherent part of the concept of disease in scientific research. Others defend the DSM-5's

208–215.

[45] Derek BOLTON, *What Is Mental Disorder? An Essay in Philosophy, Science, and Values*, Oxford, Oxford University Press, 2008; Rachel V. COOPER, Disease, *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 33 (2002) 2, 263–282; Jonathan GLOVER, *Responsibility*, London, Humanities Press, 1970; George GRAHAM, *The Disordered Mind: An Introduction to Philosophy of Mind and Mental Illness*, 2nd ed., London, Routledge, 2013; Lawrie REZNEK, *The Nature of Disease*, London, Routledge & Kegan Paul, 1987; Wakefield, *The Concept of Mental Disorder…*; Jerome C. WAKEFIELD, The Biostatistical Theory versus the Harmful Dysfunction Analysis, Part 1: Is Part-Dysfunction a Sufficient Condition for Medical Disorder?, *The Journal of Medicine and Philosophy*, 39 (2014) 6, 648–682.

[46] Elselijn KINGMA, Naturalism About Health and Disease: Adding Nuance for Progress, *The Journal of Medicine and Philosophy*, 39 (2014) 6, 590–608.

[47] Boorse, *A Second Rebuttal on Health…*

purported exclusion of harm as a necessary criterion for mental disorder.[48] In DSM-IV, mental disorder was defined partly by reference to distress or disability, alongside dysfunction. Many maintain that DSM-5 has revised this by placing greater emphasis on dysfunction in psychological, biological, or developmental processes, while treating harm, such as distress or impairment, as a common but not essential feature.[49]

In the engineering approach adopted here, we should prescribe that the concept of harm is linked to that of mental disorder. In fact, this enables the concept to serve the practical aims outlined in Section 1. First, harm gives individuals a *pro tanto* reason to seek treatment when necessary.[50] Being in a harmful, pathological condition provides a practical reason to act, either to restore health or prevent deterioration. When someone suffers harm from a disorder, healthcare providers are expected to intervene to alleviate suffering and prevent further damage. Moreover, linking the concept of mental disorder to harm helps guard against unnecessary medicalisation, ensuring mental healthcare focuses on patient well-being.[51] An instructive case is the 1973 revision of the DSM-II, which removed homosexuality as it was no longer seen as harmful, and the subsequent inclusion of harm as a necessary condition for mental disorder in DSM-III.[52]

The concept of harm, however, must be clarified to support a workable account of mental disorder. Psychiatry might recognise various harms, including distress, poor social ties, unstable relationships, reduced work capacity, and shorter lifespan. But there is debate over which harms matter, how to interpret them, and whether they must be directly caused or only risked by the condition. Harm thresholds and durations add further complexity.

A very demanding normative challenge concerns how to determine the harmfulness of a condition and to respect maximally the patient's perspective.

[48] M. Cristina AMORETTI, Elisabetta LALUMERA, Harm Should Not Be a Necessary Criterion for Mental Disorder: Some Reflections on the DSM-5 Definition of Mental Disorder, *Theoretical Medicine and Bioethics*, 40 (2019) 4, 321–337.

[49] For a historical and theoretical account that supports this reading, see Rachel COOPER, Must Disorders Cause Harm? The Changing Stance of the DSM, in: Steeves DEMAZEUX, Patrick SINGY (eds.), *The DSM-5 in Perspective: Philosophical Reflections on the Psychiatric Babel*, Dordrecht, Springer, 2015, 83–96. For doubts about this reconstruction, see Jerome WAKEFIELD, Must Social Values Play a Role in the Harm Component of the Harmful Dysfunction Analysis? Reply to Rachel Cooper, in: Luc FAUCHER, Delphine FOREST (eds.), *Defining mental disorder*, The MIT Press, 2021, 553–576.

[50] For forceful formulations of this argument, see: Biturajac and Jurjako, *Reconsidering Harm in Psychiatric Manuals within an Explicationist Framework…*; Dembić, *Philosophy of Mental Disorder…*

[51] Biturajac and Jurjako, *Reconsidering Harm in Psychiatric Manuals within an Explicationist Framework…*; Mia BITURAJAC, Marko JURJAKO, Važnost pojma štete u raspravi o mentalnim poremećajima [The Importance of the Concept of Harm in the Debate on Mental Disorders], *Arhe*, 19 (2022) 37, 341–361.

[52] Jack DRESCHER, Joseph P. MERLINO, *American Psychiatry and Homosexuality: An Oral History*, New York, Routledge, 2007.

Clear cases of mental disorder, such as major depression, involve distress, and thus harm, that the patient explicitly recognises and wishes to remove. However, it is recognised that some individuals with mental illness may lack insight into their condition and its harmfulness.[53] People with obsessive-compulsive disorder, for instance, may feel distress but not judge it as a ground for medical treatment. These cases raise a key tension: psychiatry must respect autonomy while identifying harm in patients who do not acknowledge it. Overriding autonomy may be justified to prevent harm, but this requires ethical or normative justification.

Some philosophers have stressed the need for *normatively justified* standards in identifying psychiatric harm.[54] Reliance on prevailing social norms is problematic. Such norms may legitimise repression, especially given psychiatry's history of pathologizing dissent and enforcing dominant ideologies.[55] Furthermore, there is persistent disagreement, within and across cultures, about the criteria for psychiatric harm. These concerns highlight the need for standards that can be justified through reflective engagement with, but not reduced to, prevailing social standards.

While I endorse the need for normative philosophical justification, this commitment forms only part of the broader framework I propose. There are, in fact, open issues about the shape such justifications should take, and significant philosophical disagreement remains. Some authors appeal to accounts of human flourishing or wellbeing. Christopher Megone, influenced by Aristotelian ethics and metaphysics, argues that a condition is psychologically harmful when it disrupts rational capacities essential to human flourishing.[56]

However, the assumption of a unitary account of human flourishing, as proposed by Megone, conflicts with the pluralism of admissible conceptions of the good life that are recognised in contemporary democratic societies.[57] Others focus, thus, less on flourishing and more on the criteria of the standards that are relevant for psychiatric harm. On this view, philosophy evaluates which standards ought to count, rather than identifying substantive ideals of human flourishing. Russell Powell and Eric Scarffe call for justifications rooted in prin-

[53] Ivana S. MARKOVÁ, *Insight in Psychiatry*, Cambridge, Cambridge University Press, 2005.

[54] Megone, *Aristotle's Function Argument…*; Russell POWELL, Eric SCARFFE, Rethinking »Disease«: A Fresh Diagnosis and a New Philosophical Treatment, *Journal of Medical Ethics*, 45 (2019) 579–588.

[55] See Tyler FEIGHT, Wakefield's Harmful Dysfunction Analysis of Disorder: Too Much Dysfunction, Too Little Harm, *The Journal of Medicine and Philosophy*, 47 (2022) 4, 409–433; Jerome C. WAKEFIELD, Jordan A. CONRAD, Social Values and the Harm Component of Disorder: The Case of the DSM's Definition of Mental Disorder, *The Journal of Medicine and Philosophy*, 45 (2020) 5-6, 580–603.

[56] Megone, *Aristotle's Function Argument…*

[57] See Shane GLACKIN, Three Aristotelian Accounts of Disease and Disability, *Journal of Applied Philosophy* 33 (2016) 3, 251–270; Luca MALATESTI, Elvio BACCARINI, The Disorder Status of Psychopathy, in: Luca MALATESTI, John McMILLAN, Predrag ŠUSTAR (eds.), *Psychopathy: Its Uses, Validity and Status*, Cham, Springer, 2021, 291–309.

ciples that are rational and open to scrutiny.[58] However, this approach might also disqualify certain standards that, although not strictly rationally justified, may still be admissible within liberal democracies. For example, consider the case of people who are harmed by their psychological incapacity to align their behaviour and mental life to the requirement of a legitimate religious practice that they devoutly follow.[59] Finally, there is an emerging trend in offering justifications of standards relevant for mental disorder in terms of a liberal framework, often inspired by the work of the political philosopher John Rawls.[60] These justifications aim to respect the plurality of the conceptions of the good life admissible in contemporary liberal democracies.[61]

All these contrasting approaches point to unresolved foundational questions that demand further philosophical investigation. However, a condition may be harmful by deviating from justified standards without being a mental disorder, as in the case of harm caused by justified punishment.

4. Factors beyond ordinary control

Establishing what constitutes psychiatrically relevant harm, beyond ascribing it based on justified normative standards, requires that remedying it lies outside individual control. In fact, this type of harm should provide a reason for medical intervention. However, not all harm warrants medical attention. Some accounts plausibly suggest that psychiatrically relevant harm cannot be alleviated by simply using ordinary means available to the individual.[62]

Specifying the notion of control required in this context is a demanding task. First, it must be recognised that this notion relates to issues of justification for psychiatrically relevant harm discussed above. Understanding, in general terms, factors that bypass ordinary capacities is intuitive in certain paradigmatic cases. For example, persons with depression may wish to avoid the distress associated with their condition but, despite their efforts, may be unable to

[58] See: Powell, Scarffe, *Rethinking »Disease«…*, 583.

[59] For this line of reasoning, see Jerome C. WAKEFIELD, Jordan A. CONRAD, Does the Harm Component of the Harmful Dysfunction Analysis Need Rethinking? Reply to Powell and Scarffe, *Journal of Medical Ethics*, 45 (2019) 9, 594–596.

[60] John RAWLS, *A Theory of Justice*, Cambridge (Mass.), Harvard University Press, 1971; John RAWLS, *Political Liberalism*, New York, Columbia University Press, 1993.

[61] See, for the concept of mental health: Anna ALEXANDROVA, Sam WREN-LEWIS, Mental Health Without Well-being, *The Journal of Medicine and Philosophy*, 46 (2021) 6, 690–708; for the concept of disability: Elvio BACCARINI, Kristina LEKIĆ BARUNČIĆ, Public Justification, Evaluative Standards, and Different Perspectives in the Attribution of Disability, *Philosophies*, 8 (2023) 5, 87; for the concept of mental disorder in general: Kristina LEKIĆ BARUNČIĆ, Rawls' Theory of Justice in the Context of Mental Disorders, *Filozofska istraživanja*, 43 (2023) 3, 451–467.

[62] Fulford, *Moral Theory and Medical Practice…*

do so. Here, it seems intuitively clear that relieving the harmful condition bypasses the individual's control. However, someone might embrace the distress associated with depression, for instance, as a form of deserved punishment.[63] In such cases, if there are justified grounds for recognizing the harmfulness of the condition independently of the individuals' perspectives, and their rights and autonomy are respected, the harmful condition can be said to bypass their control. This holds true even if they do not personally experience it that way. Similar considerations apply to a person who lacks insight into their justifiably harmful condition.

While psychiatric harm arises from factors that are partly or wholly beyond an individual's ordinary control, it is crucial to recognise that this does not require mental disorders to eliminate personal agency entirely, nor even in the specific case of harmful symptoms or other associated factors.[64] It should be recognised that individuals, especially in protective or therapeutic settings, often maintain some degree of responsibility or resilience in managing or mitigating their mental disorders.

Debates persist over which concept best captures the factors that should determine psychiatric harm in ways that bypass or undermine individuals' control in ameliorating their situation. Some argue that mental disorder should involve harm deriving from psychological dysfunctions. The DSM-5 echoes this view, emphasizing psychological or behavioural harmful dysfunctions as criteria for categorizing disorders. Wakefield advocates this account, emphasizing that dysfunction should be based on the failure of a biological mechanism to perform its function within an etiologically grounded evolutionary psychology framework.[65] Dysfunctions highlight issues that often arise within individuals, beyond their control, and treating or compensating for dysfunctions requires specialized scientific research and psychiatric intervention, since ordinary resources are insufficient. In cases of biological dysfunction, this perspective aligns psychiatry with the scientific framework of evolutionary biology.

However, it remains contentious whether biological dysfunction, or another dysfunction concept, is the correct foundation for defining mental disorder. Some argue that Wakefield's approach is overly revisionary and impractical for current psychiatry.[66] Understanding the biological functions of mental mechanisms is highly speculative. Defining the natural functions of many cognitive processes is challenging, and traits that may have been adaptive in the past may not be relevant today. Additionally, there is no consensus on what constitutes a »normal« evolutionary function. The variability in cognition and behaviour

[63] Thomas STOMPE et al., Guilt and Depression: A Cross-Cultural Comparative Study, *Psychopathology*, 34 (2001) 6, 289–298.

[64] Steve PEARCE, The Place of Free Will and Agency in Psychiatric Practice: Commentary on William James and British Thought: Then and Now, *BJPsych Bulletin*, 44 (2020) 2, 57–60.

[65] Wakefield, *Conrad, Does the Harm Component…*

[66] See: Bolton, *What Is Mental Disorder?…*, 160–161; Graham, *The Disordered Mind…*, 121–126.

across cultures and individuals complicates establishing universal norms based on evolutionary biology.

Inability is another concept proposed to account for mental disorder.[67] Evaluating the status of a condition based on whether it involves significant inabilities or diminished abilities may offer insight. This concept, like dysfunction, points to impairments that interfere with an individual's control. If an inability causes harm because the individual cannot meet justified social or mental standards, we may have a *prima facie* reason to consider it a disorder.

Some proponents of the inability account argue that functional accounts of mental disorder are untenable.[68] However, without taking sides, it can be noted that certain dysfunctions are harmful in a psychiatrically relevant sense, as they involve inabilities or reduced abilities. For example, dyslexia is considered a disorder under this account because it entails a dysfunction that harms the individual by impairing the ability to read. Furthermore, the appropriateness of these proposals depends on balancing the current need for practical intervention with the goal of establishing long-term research projects aimed at scientifically describing and explaining mental disorder. Dysfunction is a central concept in various areas of biological science and can also be expanded within computational accounts of mental functioning, which are at the core of contemporary cognitive science.

However, even harm that meets justified psychiatric standards and stems from factors beyond ordinary control may not suffice for classifying a condition as a mental disorder. This raises a further question: how can such a concept help distinguish mental from bodily disorders?

5. What makes mental a mental disorder

Philosophers have overlooked how mentality defines mental disorders, focusing mostly on the disorder aspect.[69] Psychiatry has traditionally focused on disturbances in mental states, categorized under areas such as mood, thought, perception, and cognition. Philosophers of psychiatry have examined whether consciousness or intentionality characterises mental disorders.[70] Even if these

[67] Dembić, *Philosophy of Mental Disorder…*; Gaete, *The Concept of Mental Disorder…*; Graham, *The Disordered Mind…*; Luca MALATESTI, Psychopathy and Failures of Ordinary Doing, *Etica & Politica/Ethics & Politics*, 2 (2014) 1138–1152; Malatesti, Baccarini, *The Disorder Status of Psychopathy…*, Lennart NORDENFELT, *Health, Science, and Ordinary Language*, Amsterdam, Rodopi, 2001.

[68] Dembić, *Philosophy of Mental Disorder…*; Graham, *The Disordered Mind…*

[69] For exceptions, see Bengt BRÜLDE, Filip RADOVIC, What Is Mental About Mental Disorder?, *Philosophy, Psychiatry & Psychology*, 13 (2006) 2, 99–116; Dembić, *Philosophy of Mental Disorder…*; Gaete, *The Concept of Mental Disorder…*; Jerome C. WAKEFIELD, What Makes a Mental Disorder Mental?, *Philosophy, Psychiatry & Psychology* 13 (2006) 2, 123–131.

[70] Gaete, *The Concept of Mental Disorder…*, 200.

characterisations suffice, how mentality relates to mental disorders remains an open issue.[71]

Mental disorders may be termed *mental* due to their mental symptoms identified through descriptive psychopathology.[72] However, mental symptoms also appear in physical disorders, making this distinction insufficient.[73] Another view might hold that mental disorders are mental because mental states cause them. But this is also problematic, as the explanatory role of mentality remains contested. The issue links to wider debates on the proper level of description (see Section 2) and on psychiatric explanation more broadly.[74] Moreover, some argue that the ultimate causes of mental disorders lie in the brain[75], while others contend that they involve a complex interplay of physical, mental, and social factors.[76]

Another view holds that mental disorders deviate from standards definable only via mental features. This raises complex issues. First, it touches on the mind-body problem, central in the philosophy of mind.[77] Some views on this problem do not give a univocal answer on whether standards concerning the mind are relevant. Interactive dualists, for instance, might take different positions. They could ground mental disorder in a deviation from irreducible standards governing mental states. Alternatively, they might argue that deviations from standards concerning the body, namely the brain, cause disturbances in mind and behaviour. Finally, they might point to deviations from standards governing the causal connections between mind and body. Yet, some positions bear directly on whether disorder can be grounded in mentality. Eliminative materialists, for example, reject mental states as outdated constructs, favouring brain-state descriptions. This removes the notion of »disordered mentality« and replaces it with physical disorder. Likewise, Thomas Szasz rejects the category of mental disorder, as his view of mind denies the validity of mentality-specific standards for defining it.[78]

---

[71] Brülde, Radovic, *What Is Mental…*; Gaete, *The Concept of Mental Disorder…*

[72] Femi OYEBODE, *Sims' Symptoms in the Mind: Textbook of Descriptive Psychopathology*, 7th ed., Philadelphia, Elsevier, 2022.

[73] Brülde, Radovic, *What Is Mental…*

[74] Kenneth S. KENDLER, Josef PARNAS (eds.), *Philosophical Issues in Psychiatry: Explanation, Phenomenology, and Nosology*, reprint ed., Baltimore, Johns Hopkins University Press, 2015.

[75] Thomas R. INSEL, Bruce N. CUTHBERT, 'Brain Disorders? Precisely', *Science*, 348 (2015) 6234, 499-500.

[76] Derek BOLTON, Grant GILLETT, *The Biopsychosocial Model of Health and Disease: New Philosophical and Scientific Developments*, Cham, Palgrave Macmillan, 2019.

[77] Tim CRANE, *Elements of Mind*, Oxford, Oxford University Press, 2001; Marko JURJAKO, Luca MALATESTI, *Filozofija uma: suvremene rasprave o odnosu uma i tijela* [Philosophy of Mind: Contemporary Debates on the Mind–Body Relationship], Rijeka, Faculty of Humanities and Social Sciences, University of Rijeka, 2022.

[78] Miguel NÚÑEZ DE PRADO-GORDILLO, Broken Wills and Ill Beliefs: Szaszianism, Expressivism, and the Doubly Value-Laden Nature of Mental Disorder, *Synthese*, 203 (2024) 1, 24.

Theoretical issues also arise if one holds that mental disorders deviate from standards unique to mentality. Graham, for example, stresses that mental disorders involve irrational or reason-unresponsive states.[79] Moreover, he claims this level alone explains how such states are both mental and disordered. He then argues that mental disorders can exist without brain disorders.[80] This view holds that mentality-specific standards, not biological ones, define what makes a disorder mental.

However, not all who link mentality to mental disorder through standards for mental states see these as independent of brain-based or other biological or computational standards. Wakefield's concept of mental disorder as harmful dysfunction offers a biologically grounded view.[81] He holds that mental disorders result from failures in evolved psychological mechanisms. Here, disorder stems from biologically grounded dysfunction in mental systems. Likewise, newer neurocomputational models, such as predictive coding, treat the brain as a predictive system adapting to input.[82] On this view, disorders arise from disruptions in computational processes and their functional roles.[83]

These debates show that the role of mentality in defining mental disorder remains unsettled. While much attention has gone to what makes something a disorder, less has been said about what makes it mental. This gap must be addressed to develop a concept of mental disorder that captures its distinctiveness.

Conclusion

I have proposed a framework with some core elements for shaping the concept of mental disorder: unity, justified harm beyond control, and mentality. These, in turn, are constraints to guide further conceptual work. They prompt questions about what unifies disorder, how harm is justified, what the relevant explanatory factors of this harm are, and whether and how mentality is required. The framework allows for refinement across theoretical, clinical, and social contexts. It leaves room for philosophical reflection on the assumptions and values underlying psychiatry.

[79] Graham, *The Disordered Mind*…

[80] Graham, *The Disordered Mind*…, 24.

[81] Jerome C. WAKEFIELD, Addiction from the Harmful Dysfunction Perspective: How There Can Be a Mental Disorder in a Normal Brain, *Behavioural Brain Research*, 389 (2020) 112665.

[82] Jakob HOHWY, *The Predictive Mind*, Oxford, Oxford University Press, 2013.

[83] Ryan SMITH, Paul BADCOCK, Karl J. FRISTON, Recent Advances in the Application of Predictive Coding and Active Inference Models within Clinical Neuroscience, *Psychiatry and Clinical Neurosciences*, 75 (2021) 1, 3–13.

Acknowledgements

I thank the reviewers for their positive comments and helpful suggestions. I also thank the participants at the following events for their valuable feedback on earlier drafts: the *Philosophy & Psychiatry Seminar 2023*, Institute of Social Sciences, Belgrade (13 December 2023), the online lecture in the *Institute of Applied Psychology Interdisciplinary Series*, Jagiellonian University, Kraków (24 April 2024); and, at the University of Rijeka, the workshops: *Pernicious and Beneficial* (12 June 2024), *Philosophy & Science* (24–25 June 2024), and *TIPPS-PACS Journal Club* (22 October 2024).

This work is an outcome of the project *Philosophical Analysis within Cognitive Science (PACS)*, funded by the University of Rijeka (UNIRI-ISKUSNI-HUMAN-23-14). I also gratefully acknowledge funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) within the Centre for Advanced Studies in the Humanities "Human Abilities", Berlin, grant number 409272951.

Luca Malatesti*

Mentalni poremećaj: Pristup inženjerstva pojmova

Sažetak

Ovaj rad istražuje mogućnost definiranja pojma mentalnog poremećaja primjenom pojmovnog inženjerstva. Ta metoda predlaže oblikovanje pojma u skladu s ciljevima koje bi trebao ispunjavati. Zastupam tezu da se mentalni poremećaji trebaju shvaćati kao jedinstvena mentalna stanja koja uključuju štetu. Štetu treba procjenjivati prema opravdanim standardima i mora proizlaziti iz čimbenika izvan kontrole pojedinca. Rad razmatra ovu definiciju prema pet kriterija: jedinstvu, šteti, normativnim standardima, čimbenicima izvan kontrole i mentalnoj naravi. Cilj je predložiti okvir koji može pridonijeti jasnijoj dijagnozi, etičkom opravdanju liječenja i boljem teorijskom razumijevanju.

Ključne riječi: mentalni poremećaj, pojmovno inženjerstvo, kontrola, šteta, normativni standardi.

* Prof. dr. sc. Luca Malatesti, Sveučilište u Rijeci, Filozofski fakultet, Odsjek za filozofiju i Katedra za kognitivne znanosti, Sveučilišna avenija 4, HR-51000 Rijeka; e-mail: lmalatesti@ffri.uniri.hr.