

## BOOK REVIEW

**Sanja Dembić**  
***PHILOSOPHY OF MENTAL DISORDER: AN ABILITY-  
BASED APPROACH***  
**Routledge, 2023**  
**ISBN: 9781032435466**  
**Hardcover: £108.00**  
**eBook: £31.99**

Luca Malatesti<sup>1</sup>

<sup>1</sup>University of Rijeka, Faculty of Humanities and Social Sciences, Croatia

Sanja Dembić in *Philosophy of Mental Disorder: An Ability-Based Approach* integrates two general ideas. Methodologically, she adopts an explicationist method, which defines concepts to serve theoretical and practical objectives. This method underpins her defence of the Reability View, an inability-based account of mental disorder. According to this view, mental disorders are conditions marked by harmful inability to respond to reasons.

The explicationist method and inability-in-reason-responsiveness accounts of mental disorder have already been proposed in the philosophy of psychiatry, yet Dembić's account is innovative. Traditional a priori descriptive methods favoured by analytic philosophers have faced criticism for their inconclusiveness in defining mental disorder. This has led some to adopt conceptual explication's prescriptive method. Moreover, there are other proponents of the inability-in-reason-responsiveness account, notably George Graham. However, Dembić is the first to defend this type of account explicitly through the explicationist method. Her further contributions to the debate are detailed in the following chapter summaries.

In the first chapter, Dembić advances and defends her explicationist project. She identifies four goals for the concept of mental disorder, that she calls adequacy conditions. Theoretically, the proper applications of the concept should (1) distinguish pathological from healthy conditions and (2) societal deviations, and (3) mental from bodily disorders. Practically, the concept of mental disorder should (4) provide the affected individuals

a *pro tanto* reason to seek treatment and mitigate some of their social, moral, or legal obligations. A *pro tanto* reason provides initial justification that remains valid unless outweighed by stronger, overriding considerations. She adopts a conservative approach to current use of the concept of mental disorder, insofar the stated goals remain attainable. She critiques a priori conceptual analysis and rejects extrapolating empirically the concept from paradigmatic cases, as proposed by Dominic Murphy.

The second chapter considers Harm Views and Action Views of mental disorder. Regarding the Harm Views, Dembić considers those advanced by Danner Clouser, Charles Culver, and Bernard Gert; Lawrie Reznick; and Rachel Cooper, who agree that harm is necessary but insufficient for defining disorders. Clouser, Culver, and Gert add that disorders lack sustaining external causes; for instance, strangulation is not a disorder since removing the external cause resolves the harm. Reznick defines disorder as an abnormal mental condition harming species-typical individuals in standard circumstances and requiring medical intervention. Cooper describes disorder as a bad, unfortunate, and potentially treatable condition. Regarding the Action Views, Dembić considers the positions of William Fulford and Lennart Nordenfelt. These accounts ground the concept of mental disorder in agential failure. According to Fulford, a disorder (illness in his terminology) involves the experience of a failure of ordinary doing, acting as we intend to act, in the absence of obstruction or opposition. For Nordenfelt, an agent is disordered when they do not have the agential abilities to achieve their vital goals, these goals being those whose fulfilment is necessary and jointly sufficient for a minimal amount of welfare and happiness.

Dembić criticizes all the above positions but also highlights certain positive aspects. She attacks each of them for specific shortcomings but emphasizes that, collectively, they fail to meet the adequacy conditions (58-64). She argues that they cannot, in relevant cases, distinguish mental disorders from non-pathological conditions, bodily disorders, or deviations from social, political, or ethical rules. Additionally, they fail to explain satisfactorily why a mental disorder provides a *pro tanto* reason for treatment. However, she acknowledges that harm is undoubtedly a central element of the concept of disorder, though it needs further specification beyond what the theories considered so far have provided. Similarly, she recognizes that agential inabilities are promising concepts for explicating that of mental disorder.

In the third chapter, the discussion shifts to Biological Function Views, which link mental disorders to biological dysfunctions. Dembić considers the influential proposals of Jerome Wakefield and Christopher Boorse.

According to Wakefield, a mental disorder constitutes harmful dysfunction, with dysfunction understood in evolutionary terms. Boorse, while acknowledging the clinical and practical relevance of harm, identifies dysfunction itself as central to mental disorder. Diverging from Wakefield, Boorse adopts a goal-directed account of dysfunction. Specifically, he defines the function of a mechanism in an organism as determined by the overall goals of that type of organism, which are, in turn, established by a reference class of exemplars of that organism.

In her critique of Wakefield's and Boorse's positions, Dembić considers key aspects of the broader debate they have generated. Regarding Wakefield's account, amongst other shortcomings that she identifies (73-80), she maintains that this account makes psychiatry overly and impractically dependent on speculative or future evolutionary accounts of mental mechanisms. Moreover, this account imposes overly stringent explanatory requirements: mental disorder causes must be tied to hereditary traits established through evolution.

Regarding Boorse's account, she focuses her criticism, amongst other issues (84-87), on his defining function and dysfunction based on a reference class of individuals. If all members of a species suffered from a particular disease, they would paradoxically be classified as healthy. Furthermore, she shows the limitations of these proposals when they are tested in terms of the criteria of adequacy (87-90). Both accounts fail in recognizing certain clear cases of mental disorder. While both can account for the difference between mental disorders and departures from social, political, or ethical norms, Boorse is not able to give a satisfactory account of the practical normative goals, while the concept of harm in Wakefield's account has this capacity.

The heart of her framework, the Reability View is presented in the fourth chapter. This account defines mental disorders as harmful inability to act on normative reasons. To give an idea of the complexity and precision of the definition, I quote here the full version of the account, that is meant to account for the clearest types of cases that should fall under the concept of mental disorder:

**RHapsy** An individual S has a mental disorder if and only if i S does not have the ability to respond adequately to some of their available (apparent) reasons for (or against) some of their reason-sensitive attitudes or actions; in view of their mental constitution and their life circumstances (where the threshold of inability is determined by the degree at which individuals in the relevant comparison class are, on average, harmed by their

condition C in some respect X) and ii S is harmed by their condition C in some respect X, where X is some component of what makes S's life a non-instrumentally good one for S.

For S to be harmed by C in some respect X, it is sufficient that i C is non-instrumentally bad for S in some respect X ii C causes S to be sufficiently worse than before or that they would otherwise have been in respect X; or iii C prevents S from receiving a good in respect X, thereby leaving S in a non-instrumentally bad state in respect X, where X is some component of what makes S's life a non-instrumentally good one for S. (Dembić 2024, 126-127)

Dembić discusses all the main concepts that enter this complex definition. Here I briefly examine her views on the central concepts of ability, (apparent) reason, and harm.

Other authors have identified abilities as central to analysing mental disorders. Dembić deepens this type of account with insights from ontological investigations on abilities developed outside the philosophy of psychiatry. She proposes four characteristics of abilities. First, an ability to  $\phi$  is a modal property of an agent, reflecting what they can or could do. Second, abilities can be agentive, involving actions, or non-agentive, where actions are not involved. For mental disorders, both types of inability are relevant. For example, according to Dembić, schizophrenia involves the non-agentive inability to avoid holding certain beliefs that they have reason not to endorse. Compulsive disorder, instead, involves an inability to resist acting on obsessive thoughts, although they have reason not to do so. Third, abilities exist in degrees. An individual's inability to respond to apparent reasons is significant when it is, compared to appropriately individuated peers, low enough to harm members of that reference group. Finally, abilities depend on specific facts. For example, a swimmer can swim, but this is compromised with a broken arm. Regarding mental disorders, relevant abilities or inabilities must be identified based on internal mental constitutions and external conditions, what Dembić terms "life circumstances", which the agent cannot easily change.

Central to Dembić's account are the inabilities to respond with appropriate attitudes and actions to normative reasons and *apparent* normative reasons (110-120). For Dembić, a normative reason is given by a fact or a true proposition that counts in favour of (or against) responding with an action or attitude. An apparent normative reason would count in favour (or against) a certain action if it were true. For example, anxiety disorder arises

from a harmful inability to avoid frequent fear despite beliefs providing (apparent) reasons against it (120-126).

Dembić addresses complex issues regarding the proper characterization of harm (113-119). Although she acknowledges the provisional nature of her proposal, she argues that psychiatric harm must consider both the individual's subjective perspective and objective criteria independent of it. Thus, she advocates a hybrid view. A central feature of her account is that the harmfulness of an inability to respond to reasons depends on the reference class. This requires identifying relevant classes, by means of age or other contextual factors, and determining the threshold where inability becomes psychiatrically harmful.

The fifth chapter defends the Reliability View. First, Dembić argues that this view meets the adequacy conditions outlined in the first chapter. For example, she tests its descriptive adequacy in distinguishing health from pathology by examining principal mental disorders classified so far. For each disorder type, she identifies the relevant agentive or non-agentive abilities to respond to reasons central to these disorders. Additionally, she demonstrates how her account avoids the difficulties faced by positions she criticized in the second and third chapter.

Moreover, she addresses two objections to her account (155-161). First, it could be argued that her account assumes the mental is constituted just by sensitivity to reasons, thus excluding mental states as our sensations. She states that she is not assuming this. Instead, she considers the concept of the mental that is relevant to psychiatry. For practical purposes, psychiatry focuses on conditions involving insensitivity to reasons, particularly in interventions like discursive therapy. Second, it could be argued that harm is not a necessary requirement for a mental disorder. She acknowledges the possibility of a purely theoretical concept of mental disorder based on inability to respond to (apparent) reasons. However, she argues that since her goal is to provide a relevant concept that offers *pro tanto* reasons for intervention, the concept of harm must be retained.

A case study on addictive disorders, in the last chapter, underscores the framework's practical value. Addiction, often framed as a failure of willpower or a biological defect, is reconceived as an inability to act on reasons. This reinterpretation demonstrates the Reliability View's capacity to inform both philosophical debates and clinical practice.

This book examines mental disorder, focusing on ability, reasons, harms, all concepts central to unresolved controversies in philosophy. For instance, Dembić endorses what appears a metaphysically realist account

of normative reasons. Not all readers may find this endorsement convincing.

There may also be internal tensions between Dembić's conceptual work and psychiatric theory and practice, rooted in her methodological assumptions. Her explication of the concept of mental disorder does not appear to be maximally conservative. In fact, it is not in line with descriptive psychopathology. For example, she tests the adequacy of her proposed explication of mental disorder against currently classified conditions, by showing that they might involve inability to respond to reasons. However, existing diagnostic categories prioritize behavioural symptoms and inferred mental states and traits rather than inability. Bridging this gap requires empirical research, which could shift her project from conservation to revision.

Her treatment of harm introduces further challenges. Referencing harm to a class of individuals raises two significant issues. First, her concept may inadequately safeguard individual perspective and rights. For example, individuals with comparable inability to respond to reasons may evaluate the associated harm differently and still be ascribed the same mental disorder. In this case, they should not have the same *pro tanto* reason for being treated and excused. Second, this framework risks criticisms like those Dembić has directed at Boorse's use of class of reference to fix biological dysfunction. Variations in the class of reference would determine too revisionary variations of the disorder status of certain conditions.

Finally, she incorporates reason-responsiveness into her explication, raising the question of whether this embeds a non-revisable assumption about the understanding of the notion of the mental relevant for mental disorder. Dembić asserts compatibility with most views of the mind-body problem, excluding eliminativism, but this claim requires further exploration. In the philosophy of psychiatry, the mind-body relation also involves determining the appropriate level of description or explanation for identifying standards relevant to disorder status. Even if her arguments against using biological dysfunctions are compelling, other standards for mental disorder might go beyond normatively individuated reason-responsiveness. For example, advances in computational accounts, such as the Bayesian brain framework in psychiatry, might suggest alternative empirically grounded views. Integrating the conceptual elaboration of disorder standards with empirical research should remain open to revision. However, Dembić seems to treat reason-unresponsiveness as a conceptual and fixed criterion for determining when a condition qualifies as a mental disorder.

Despite these tensions, *Philosophy of Mental Disorder* is a remarkable book for its clarity, depth, and innovation in understanding mental disorders. It seamlessly bridges abstract theory, often drawing on areas of philosophy not typically applied to the philosophy of psychiatry, with practical concerns. By employing the method of explication, it makes an invaluable contribution to conceptual debates on the nature and treatment of mental health. This ground-breaking book is essential reading for philosophers, clinicians, and anyone invested in the conceptual foundations of mental health.

### **Acknowledgements**

Many thanks to Sanja Dembić, Miguel Núñez de Prado Gordillo, and Marko Jurjako for reading and commenting on previous versions. This work was supported by the project PACS, funded by the University of Rijeka under contract UNIRI-ISKUSNI-HUMAN-23-14.

