MARKO JURJAKO
University of Rijeka
mjurjako@gmail.com

LUCA MALATESTI
University of Rijeka
lucamalatesti@gmail.com

IN WHAT SENSE ARE MENTAL DISORDERS BRAIN DISORDERS? EXPLICATING THE CONCEPT OF MENTAL DISORDER WITHIN RDOC[1]

abstract

Recently there has been a trend of moving towards biological and neurocognitive based classifications of mental disorders that is motivated by a dissatisfaction with the syndrome-based classifications of mental disorders. The Research Domain Criteria *(indicated with the acronym RDoC) represents a bold and systematic attempt to foster this advancement. However, RDoC faces theoretical and conceptual issues that need to be addressed. Some of these difficulties emerge when we reflect on the plausible reading of the slogan "mental disorders are brain disorders", that according to proponents of RDoC constitutes one of its main presuppositions. Some authors think that endorsing this idea commits RDoC to a form of biological reductionism. We offer empirical and theoretical considerations for concluding that the slogan above should not be read as a reductionist thesis. We argue, instead, that the slogan has a pragmatic function whose aim is to direct research in psychopathology. We show how this function might be captured in the framework of a Carnapian explication as a methodological tool for conceptual engineering. Thus, we argue that a charitable interpretation of the aims of the proponents of RDoC should be understood as an attempt at providing an explication of the concept of mental disorder in terms of brain disorder whose main goal is to provide a more precise and fruitful notion that is expected to have a beneficial impact on classification, research, and treatment of psychiatric conditions.*

keywords

conceptual engineering/explication, biological reductionism, DSM-5, ICD-10, mental disorder, natural kinds, Research domain criteria (RDoC), philosophy of psychiatry

[1] **Acknowledgements**: we want to thank Mia Biturajac for her valuable proofreading of this article. Marko would like to thank the hosts of the BIAS institute (Nerezine) where part of the paper was written and its newest member David Boycie Bizovich for their creative, emotional, and financial support. This paper is an outcome of the Project RAD, financed by the Croatian Science Foundation (HRZZ) (grant IP-2018-01-3518). Marko Jurjako also acknowledges the support from the University of Rijeka (project Bcon, small grant 17.05.2.2.04).

Phenomenology and Mind, n. 18 - 2020, pp. 182-198
DOI: 10.17454/pam-1814
Web: http://www.rosenbergesellier.it/elenco-libri?aaidriv=14

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1. Introduction

Recently there has been a trend towards biological and neurocognitive based classifications of mental disorders (Brazil, van Dongen, Maes, Mars, & Baskin-Sommers, 2018; Insel & Cuthbert, 2015; Wiecki, Poland, & Frank, 2015). The *Research Domain Criteria* (RDoC), for instance, aims at determining categories of mental disorders that would reflect, to a greater extent, the biological underpinnings of psychological disturbances and health problems (Cuthbert & Insel, 2013; Insel *et al.*, 2010). The motivation of these approaches is replacing syndrome-based classifications as encapsulated in many diagnoses in classificatory systems such as the *Diagnostic statistical manual of mental disorders* (American Psychiatric Association & American Psychiatric Association, 2013, APA) or the *International classification of diseases* (World Health Organization, 1992, WHO). In these systems, mental disorders are categorised in terms of symptomatic behaviours, mental disturbances, and maladaptive personality traits, and not on the basis of their aetiology or underlying biological factors (Murphy, 2006; Tabb, 2015).

In this paper we aim at explicating and clarifying some methodological assumptions of RDoC and similar biologically based approaches to the classification of mental disorders. We focus on the issue of the correct interpretation of the slogan "mental disorders are brain disorders" that is often associated with such classifications (Insel & Cuthbert, 2015). We argue that this claim should not be read as a commitment to a form of biological reductionism, as, for instance, Borsboom, Cramer, and Kalis do in a recent target paper in *Behavioral and Brain Sciences* (2019). Instead, we maintain that the slogan has a pragmatic function that should direct research in psychopathology. We show how this function might be captured in terms of a broadly construed notion of explication as elaborated by Rudolf Carnap (1971) and contemporary scholars such as Catarina Dutilh Novaes (2018). We argue, thus, that the function of the explication of the concept of mental disorder in terms of that of brain disorder is to provide a more precise and fruitful notion that is expected to have a beneficial impact on the classification, research, diagnostics, and the treatment of psychiatric conditions.

The paper is structured as follows. In Section 2, we introduce RDoC and describe the main reasons that have motivated its creation and adoption. In Section 3, by means of an example concerning antisocial personality disorders, we argue that this type of approach is not committed to a classical form of reductionism. In section 4, we argue, instead, that the idea that mental disorders are brain disorders has a pragmatic function that should direct research in psychopathology. Finally, in the same section we show how this idea might be outlined in terms of a Carnapian explication as conceptual engineering. Here the explication of the concept of mental disorder as a disorder of brain circuits is expected to integrate psychiatry with the rest of medicine with the final aim of improving research and clinical practice in treating psychopathology.

2. The RDoC approach to the classification of mental disorders

The Research *Domain Criteria* (RDoC) project is a recent, biologically oriented approach to categorization of mental disorders (see, e.g. Insel *et al.*, 2010). Its final goal is to develop new classification systems of psychiatric conditions based on data spanning from genetics and neurobiology to self-reports and behavioural tasks. Integrating new with already available biological data RDoC create more valid measures of disorders which would aid clinical practice and improve health outcomes (Cuthbert, 2014; Cuthbert & Insel, 2013; Insel *et al.*, 2010). While the long-term goal of RDoC is to develop personalised form of psychiatric treatment, the short term goal is to provide a platform for "research that can produce pioneering new findings and approaches to inform future versions of psychiatric nosologies" (Cuthbert & Insel, 2013, p. 7).

The main reason for introducing RDoC is to overcome serious problems in mental health research and clinical practice. In the last 50 years there has been a considerable advancement and impact of research in treating bodily illnesses that improved health outcomes and reduced mortality rates (Bethesda, 2011). However, in the case of mental disorders there are neither similar improvements in research and diagnosis, nor in treatment that led to a reduction of mortality rates and improvement of health outcomes (Cuthbert & Insel, 2013). The proponents of the RDoC project argue that some of the impediments to progress in clinical practice can be traced back to the currently dominant syndrome-based categorizations of mental disorders, as embodied in the DSMs (APA 2013) and ICDs (WHO 1992) (Buckholtz & Meyer-Lindenberg, 2012; Cuthbert & Insel, 2013; Lilienfeld, 2014). Let us consider some of these problems (for a thorough comparative analysis, see Clark, Cuthbert, Lewis-Fernández, Narrow, & Reed, 2017).

Syndrome based classifications that are based on DSM-5 and ICD-10 delineate categories of mental illnesses in terms of clusters of symptoms. Starting with DSM III (APA, 1980), the main goal of such approaches has been to devise reliable diagnoses which would enable communication between different researchers, epidemiological studies, psychiatric treatment, and practical applications, such as for insurance purposes (Cooper, 2005). It was thought that the best way to accomplish this was by building "a-theoretical" classifications, in the sense that categories would be based on observable symptoms and not on specific theories about the causal aetiology of mental disorders developed by different schools of thought in psychiatry (Tsou, 2011).

For instance, according to DSM-5, the core of the diagnosis of a major depression is stated as a list of symptoms:

> Five (or more) of the following symptoms have been present during the same 2-week period and represent a change from previous functioning; at least one of the symptoms is either (1) depressed mood or (2) loss of interest or pleasure (APA 2013, p. 160).
> Some of the symptoms are:
>
> 1. Depressed mood most of the day, nearly every day, as indicated by either subjective report (e.g., feels sad, empty, hopeless) or observation made by others (e.g., appears tearful).
> 2. Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (as indicated by either subjective account or observation.)
> 3. Insomnia or hypersomnia nearly every day.
> 4. Feelings of worthlessness or excessive or inappropriate guilt (which may be delusional) nearly every day (not merely self-reproach or guilt about being sick).
> 5. Recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide. (APA 2013, pp. 160–161)

Despite the research and practical virtues of an "a-theoretical" system of classification, in recent years the syndrome-based approach has been criticised for its many misgivings, including the fact that it disregards the causal underpinnings of diseases in their classification (Murphy, 2006; Tabb, 2015). The problems with syndrome-based approaches to mental disorders as captured by DSM-5 and ICD 10 are numerous and already well known (see, e.g. Buckholtz & Meyer-Lindenberg, 2012; Lilienfeld, Smith & Watts, 2013). To summarize the gist of these difficulties, we can cast them by using the notion of natural kind. This concept is generally taken to refer to a good scientific category (see Brzović, 2018).

Many philosophers of psychiatry agree that mental disorder categories purport to capture natural kinds (see, e.g. Beebee & Sabbarton-Leary, 2010; Brzović, Hodak, Malatesti, Šendula-Jengić, & Šustar, 2016; Samuels, 2009; Kendler, Zachar, & Craver, 2011; Tsou, 2016; cf. Tabb, 2019a; Haslam, 2014). On these views, natural kinds should not be seen as possessing biological essences, in the sense of possessing necessary and sufficient conditions that determine when a particular instance falls under the kind term. Rather, natural kinds are construed as denoting clusters of properties or causal structures that enable reliable predictions and explanatory generalizations about its instances (Boyd, 1991; Khalidi, 2013; Slater, 2015; for a review, see Brzović, 2018). In the context of research and clinical practice in biomedicine, having such kinds should enable us to develop classifications which would capture the causal underpinnings of different illnesses and their characteristic trajectories, with the aim of successfully treating them (Brzović *et al.*, 2016; Brzović, Jurjako, & Malatesti, 2018; Brzović, Jurjako, & Šustar, 2017; Kendler *et al.*, 2011). In order to serve these explanatory, inductive, and clinical purposes, the kinds or classifications in psychiatry should enable reliable diagnosis, predict temporal trajectories of illnesses, and enable preventive interventions and the design of effective therapies.

The basic complaint against the categories in currently dominant syndrome-based classification systems is that in many cases they do not capture natural kinds and thus that psychiatric practice might benefit from revising them (Brzović *et al.*, 2017; Murphy, 2006). This complaint is expressed by claiming that the current syndrome-based classifications, for the most part, have low validity. "Validity" can mean different things in biomedical research. In this context it refers to the idea that a good category of a mental disorder will denote a specific set of symptoms that differentiates it from other disorders, correlates with different behavioural, cognitive and biological measures, has a specific development trajectory and a specific response to treatments (Aboraya, France, Young, Curci & LePage, 2005). In other words, a valid psychiatric category should provide grounds for inductive generalizations and explanatory information characteristic of natural kinds that can be used in clinical practice for prediction, interventions, and treatment.

Current categorizations are not valid in this sense because they cover heterogeneous groups of people which undermines reliable prediction and treatment outcomes (Lilienfeld, 2014; Lilienfeld *et al.*, 2013). Moreover, different categories in DSM and ICD show extensive comorbidity. For instance, studies indicate that around 20% of individuals who received a diagnosis on DSM-IV also satisfy criteria for three or more other disorders in the same manual (see Lilienfeld *et al.*, 2013). Such comorbidity indicates that the diagnostic category is not well formed. This also precludes successful treatment, given that a clinician does not know which therapy to administer. In general, the fact that different DSM and ICD categories share symptomatology creates difficulties in investigating biological correlates of mental disorders and using such information for devising new research and clinical studies.

Bruce Cuthbert and Thomas Insel indicate how the problems with DSM reflect on scientific and clinical research:

> Decades of research have increasingly revealed that neural circuits and systems are a critical factor in how the brain is organized and functions, and how genetics and epigenetics exert their influence. However, this knowledge cannot be implemented in clinical studies as readily as might be hoped. Any one mechanism, such as fear circuits or working memory, is implicated in multiple disorders as currently defined; it is difficult to know which diagnostic category to select first to explore any promising leads, and a positive result immediately raises the question of whether the demonstration of efficacy must be extended to all similar disorders (a time-consuming and expensive proposition). (Cuthbert & Insel, 2013, p. 3)

In addition, studies show that most categories in DSM-IV do not have a categorical structure (Haslam, Holland, & Kuppens, 2012), rather they denote disorders whose symptom severity is dimensional. Thus, there is not a clear boundary between people suffering from a mental disorder and those who are not. DSM 5 is, thus, often criticized for presupposing that mental disorders have a categorical structure (Buckholtz & Meyer-Lindenberg, 2012; Lilienfeld, 2014). This is another sense in which DSM categories do not designate unified clusters of symptoms that might be viewed as natural kinds (Haslam, 2014).

The RDoC presents a serious and in some ways radical response to the problems mentioned above (Cuthbert & Insel, 2013; Insel *et al.*, 2010). Its aim is to replace the old schemes of psychiatric classifications and build, almost from scratch, new systems of classification that should have beneficial and far-reaching consequences for research and treatment of mental illnesses. The basic idea is to reconfigure psychiatric categories by creating research platforms that would enable researchers to gather and integrate already available genetic, neurobiological, cognitive, self-report, and behavioural data in addition to producing new data. This would help psychiatry devise valid categories of groupings that are more homogenous and providing grounds for developing therapies that are more effective.

In this sense, one could say that the goal of the RDoC project is to rebuild psychiatric categories in order to reflect natural kinds regarding mental disorders which might at first glance seem incorrect since, in contrast to the syndrome-based approaches, RDoC takes a dimensional approach to classifying psychiatric disorders (see Buckholtz & Meyer-Lindenberg, 2012; cf. Haslam, 2014). However, this issue is resolved by emphasizing that the cluster view of natural kinds (as expounded earlier in the text), does not presuppose that kinds have clear-cut boundaries. Rather, the main claim is that natural kinds, or in other words, good scientific classifications, are those that support explanations and inductive inferences that play a role in specific domains of research or practice (see Slater, 2015).

The RDoC's emphasis on biological factors in devising classifications is a built-in feature of this approach. Thomas Insel and colleagues, who initiated the RDoC project, indicate that RDoC has three assumptions:

> First, the RDoC framework conceptualizes mental illnesses as brain disorders. In contrast to neurological disorders with identifiable lesions, mental disorders can be addressed as disorders of brain circuits. Second, RDoC classification assumes that the dysfunction in neural circuits can be identified with the tools of clinical neuroscience, including electrophysiology, functional neuroimaging, and new methods for quantifying connections in vivo. Third, the RDoC framework assumes that data from genetics and clinical neuroscience will yield biosignatures that will augment clinical symptoms and signs for clinical management. (Insel *et al.*, 2010, p. 749)

Emphasis on the biological factors and the claim that mental disorders are brain disorders and the philosophical issues that it raises will become important in the next section. In the rest of the section we will outline how RDoC is expected to be and already has been implemented in practice. Insel and Cuthbert explain that:

> (t)he approach proceeds in two steps. The first step is to inventory the fundamental, primary behavioral functions that the brain has evolved to carry out, and to specify the neural systems that are primarily responsible for implementing these functions. [...] The second step then involves a consideration of psychopathology in terms of dysfunction of various kinds and degrees in particular systems, as studied from an integrative, multi-systems point of view. (Cuthbert & Insel, 2013, p. 4)

The first step led to defining, on empirical grounds, 6 major domains that are further divided into measurable dimensions (see table 1 below). These dimensions are neuropsychological constructs that provide the backdrop for conducting research in the next step. As such they are subject to continuous scientific validation and revision.[1]

Negative Valence Systems	**Positive Valence Systems**	**Cognitive Systems**	**Social Processes**	**Arousal and Regulatory Systems**
Acute Threat (“Fear”)	Reward Responsiveness	Attention	Affiliation and Attachment	Arousal
Potential Threat (“Anxiety”)	Reward Learning	Perception	Social Communication	Circadian Rhythms
Sustained Threat	Reward Valuation	Declarative Memory	Perception and Understanding of Self	Sleep-Wakefulness
Loss		Language	Perception and Understanding of Others	
Frustrative Nonreward		Cognitive Control		
		Working Memory		

Table 1: *The research domain criteria, November 2019 based on RDoC*
https://www.nimh.nih.gov/research/research-funded-by-nimh/rdoc/constructs/rdoc-matrix.shtml

The envisioned new research design based on RDoC domains and constructs can be seen as also involving two steps (Cuthbert & Insel, 2013, p. 5). The first concerns the selection of the target group of people. Standardly, this group would be delineated by symptoms comprising a mental disorder category in the DSM-5 or ICD 10. Given that RDoC is not constrained by such diagnostic categories, the target sample could be delineated by using other criteria. For instance, these could include all patients at a clinic exhibiting anxiety symptoms or even inmates in a forensic

[1] See https://www.nimh.nih.gov/research-priorities/rdoc/index.shtml.

institution exhibiting externalizing behaviour (Brazil *et al.*, 2018; Cuthbert & Insel, 2013).
In the next step an independent and a dependent variable would be chosen from dimensions relating to the constructs that need to be measured from the six domains. These dimensions can be defined on "different levels of analysis, from genetic, molecular, and cellular levels, proceeding to the circuit-level [...], and on to the level of the individual, family environment, and social context" (Insel *et al.*, 2010, p. 749). See, for instance table 2 below, for these levels in relation to the construct Acute Threat "Fear".

Construct/Sub construct	Genes	Molecules	Cells	Circuits	Physiology	Behavior	Self-Report
Acute Threat ("Fear")		BDNF CCK Cortisol/ Corticosterone CRF family Dopamine Endogenous cannabinoids FGF2 GABA Glutamat...	GABAergic cells Glia Neurons Pyramidal cells	autonomic nervous system BasAmyg Central Nucleus d-hippocampus ...	Elem BP Context Startle EMG Eye Tracking Facial EMG Fear Potentiated Startle ...	Analgesia approach (early development) Avoidance Facial expression ...	Fear survey schedule SUDS

Table 2: *Levels of analysis in RDoC, November 2019 based on RDoC*
https://www.nimh.nih.gov/research/research-funded-by-nimh/rdoc/constructs/rdoc-matrix.shtml

Thus, in devising constructs that should augment clinical practice and research there are no privileged levels of analysis, although there is an invitation to tend more to the neural circuits that can be seen as playing the mediating role between the lower genetic and molecular levels and higher cognitive/affective functions and behavioural symptoms.

3. In what sense mental disorders are not brain disorders

Given that one of the presuppositions of RDoC is that mental disorders are brain disorders, it might be assumed that this approach endorses a form of reductionism (see, e.g. Borsboom *et al.*, 2019). Traditional forms of reductionism in philosophy of psychiatry purport to reduce or identify mental disorders with neural disorders (see, e.g. Szasz, 1974). If such a reduction could be made to work, the imperative would be to ground classifications of mental disorders on classifications of brain disorders. In this case, the role of biological or neurological variables would be constitutive of the classification.
However, we think that RDoC should not be associated with this type of reductionism (see, also, our manuscript Jurjako, Malatesti, & Brazil, 2019a). Traditionally, reduction is conceived as a relation between theories T_1 and T_2 where theory T_1 is reduced to T_2, if it is possible, by means of bridging principles that characterise the concepts in T_1 in terms of the concepts of T_2, to logically derive the statements of T_1 from those of T_2 (Nagel, 1987). A classic example is the reduction of thermodynamic theory of gases to the molecular kinetic theory of gases. There, the crucial element is the bridging principle that identifies the thermodynamic concept of temperature with mean kinetic energy, which allows a reduction of the equations describing the behaviour of macroscopic gases to the equations of statistical mechanics describing the behaviour of microscopic molecules that compose gases.
The classical view of reduction seems to presuppose that theoretical roles of the terms that are being reduced can be specified independently from the terms that play similar roles in the reductive theories. When translated to the present context, the idea would be that a reduction requires providing independent identification criteria for symptoms that define a syndrome

of a mental disorder and features that individuate its neural reductive base (Borsboom *et al.*, 2019).

As we have seen, however, the RDoC project is principally motivated by the dissatisfaction with syndrome-based categorizations of mental disorders as captured in different versions of DSM and ICD (Buckholtz & Meyer-Lindenberg, 2012; Cuthbert & Insel, 2013; Lilienfeld, 2014). Accordingly, the aim of RDoC is to revise our current classifications of mental disorders by including a dataset from genetic and neurobiological to cognitive and behavioural factors, rather than reducing them to unique neurobiological mechanisms (Cuthbert, 2014; Cuthbert & Insel, 2013; see, also, Tabb, 2019b). Moreover, in these revised data-driven classifications the focus is on the newly available neurobiological data that are *correlated* with different behaviourally defined disorders, but not to the exclusion of other behavioural, psychological, phenomenological, social and even normative factors (Insel & Cuthbert, 2015). The aim of this reclassification is to find cognitive, genetic, neurobiological or even behavioural differences that might be conducive to better diagnosis, treatment, and prediction of health outcomes.

A recently proposed RDoC type of approach to antisocial personality disorder and psychopathy illustrates these points (we develop these points in more detail in Jurjako *et al.*, 2019a). Antisocial personality disorder (ASPD) is in the DSM-5 characterized as a pervasive disposition towards violating social and moral norms, that starts before the age of 15 and is underpinned by hostile attitudes and lack of remorse for persistent antisocial behaviour. Psychopathy, which should not be confused with the general diagnosis of ASPD, can be seen in forensic manifestations as an especially severe form of ASPD (Hare, 2003). Psychopaths are characterised by callous attitudes, lack of empathy and remorse, complete disregard for other people's interests and safety, manipulative and conning interpersonal styles that tend to be accompanied by different forms of aggressive and generally antisocial behaviour (for a review, see Brazil & Cima, 2016).

Given that standard measures of ASPD and psychopathy are grounded in syndrome-based approaches to psychiatric classifications (see, e.g. Međedović, Bulut, Savić, & Đuričić, 2018) they inherit all of the problems related to such classifications (see section 2 above). They include heterogeneity and low construct validity which lead to forming groups of people that often do not share meaningful cognitive, biological, or aetiological underpinnings (Cooke, 2018; Jurjako, Malatesti, & Brazil, 2019b; Međedović, Petrović, Kujačić, Đorić, & Savić, 2015; Rosenberg Larsen, 2018). In that sense they fail to fulfil the standards for being a biomedical natural kind (Brzović *et al.*, 2016, 2017; Maibom, 2018). It is assumed that these problems explain the lack of successful therapies for reducing antisocial behaviour and devising coherent policies for regulating the social response to these individuals when they offend (Jurjako & Malatesti, 2018b, 2018a; Jurjako *et al.*, 2019b).

The concrete proposals, in the spirit of RDoC, to overcome the problems that afflict syndrome-based classifications of antisocial personalities, involve revisions of these categories in accordance with the already available and expected new cognitive and biological data underpinning antisocial behaviour (see Brazil *et al.*, 2018; Jurjako *et al.*, 2019a, 2019b). In accordance with RDoC, attempts to stratify and rebuild categories of persistent antisocial behaviour start with behaviourally defined criteria for individuating groups of antisocial individuals (Jurjako *et al.*, 2019b). In practice this will involve going to clinical and forensic facilities in order to investigate individuals who show extremely aggressive, maladaptive, and generally antisocial behaviour. Once we have individuated such groups we can investigate the genetic, neurobiological, cognitive, psychological, environmental, self-reports, neuropsychological, and other behavioural correlates to form inductively more robust groupings, in the sense that such groupings would enable better predictions regarding group inclusivity for research purposes, and finally treatment outcomes. From this procedure it should be clear that antisocial behaviour and its psychological concomitants will not be reduced to genetic and neurobiological mechanisms because environmental, behavioural, and psychological criteria provide data points whose epistemic value is on a par with other more biological factors that help to determine the revised groupings. This procedure is just an example of how RDoC is supposed to be applied in practice. Therefore, there is nothing in the RDoC approach to psychiatric research that presupposes classical reductionism and the ability of identifying different levels of description independently from each other.

It remains, thus, to be explained how we should understand the slogan that "mental disorders are brain disorders" if not in the classically reductionist sense. Addressing this problem requires, preliminarily, distinguishing different important theoretical issues to which the slogan might apply. We begin next section by drawing these distinctions.

4. Engineering the concept of mental disorder

Jerome Wakefield (2014) usefully distinguishes between conceptual validity and construct validity of disorder categories. In this context, construct validity refers to explanatory factors that delineate one disorder category from another. For instance, it is expected that two different disorders will have different risk factors, causal antecedents, development trajectories, and so on. Conceptual validity refers to criteria that determine when a category designates a disorder as opposed to a normal variation. In other words, conceptual validity pertains to offering criteria about what confers a disorder status to a condition. With respect to this distinction, a category might have construct validity, but not conceptual validity, and *vice versa*. For instance, if psychopathy is not a mental disorder then it is not conceptually valid although it still might be valid as a construct because it delineates a scientifically relevant cluster of personality traits (Jurjako, 2019). Alternatively, psychopathy might be a mental disorder and thus conceptually valid, but lack construct validity, because, for instance, it is comprised of a group of people that are too heterogenous (Brzović *et al.*, 2017).

There are different philosophical views on what confers a "disorder" status to a mental or a bodily condition. According to Thomas Szasz (1974; see, also, Boorse, 2014) for a mental disorder to be real, it must consist of a neural deviance from "objective" standards of brain anatomy and physiology. From this, Szasz concludes that mental illnesses, although legitimate problems of living, are either mythological medical entities or neurological disorders. Other views adopt a more normative approach, indicating that the concept of a mental disorder cannot be defined without some reference to social norms or value judgments (see Kingma, 2013, 2014). These approaches would admit the importance of neural causes and markers in the categorisation and explanation of mental illnesses but would not require that what confers an illness status is a disorder of the brain specified independently from any normative considerations about the role that the brain plays in cognitive, behavioural, and social matters. Several contemporary authors endorse this position by arguing that the disorder or illness status must partly (Wakefield, 1992) or completely (Bolton, 2008; Fulford, 1989) be a matter of societal or other type of evaluations (for a recent discussion of this issue in the context of DSM, see Amoretti & Lalumera, 2019b, 2019a; Cooper, 2013).

The supporters of RDoC seem to be silent on the issue of conceptual validity. According to Cuthbert and Insel:

> RDoC is committed to studying the 'full range of variation, from normal to abnormal.' In some cases, only one end of a dimension may involve problem behavior (for instance, one is seldom likely to complain of an outstanding memory or keen vision), but often both extremes of a dimension may be considered as 'abnormal' [...]. (Cuthbert & Insel, 2013, p. 5)

Investigating variations from normal to abnormal brain function indicates that RDoC is not primarily about determining what makes some set of properties symptomatic of a mental disorder. Thus, they allow that there might be some other criterion, beside the independently specified brain function that determines whether some condition is a psychiatric disorder worth treating. Moreover, given that Cuthbert and Insel (2013) hold that brain functions could be specified by behavioural functions the brain evolved to implement, it is not reasonable to expect that the assumption that mental disorders are brain disorders will be specified by independently identifiable biological causes. In this spirit, Anneli Jefferson (2018) has argued that brain dysfunctions are sufficient to confer a disorder status to a brain state, however, what provides the normative criterion for deciding when a brain is dysfunctional might rely on mental or, more generally, social considerations. In this sense, we might identify mental disorders with brain disorders by mentalizing the brain, so to speak. This is compatible with Cuthbert and Insel's idea that we might determine the function of different brain areas and neural circuits by investigating what cognitive, behavioural or social functions the brain evolved to implement.

Thus, RDoC seems to be compatible with normative and strictly naturalistic ideas about what confers a mental disorder status to a condition, be it characterized as mental or bodily. Accordingly, we agree with Wakefield (2014) that RDoC does not offer a criterion for conceptual validity for categories of mental disorder. However, we do not see this as its weakness because RDoC is first and foremost a research project, while offering an account of what makes a mental condition a disorder as opposed to a normal variation belongs to perplexing conceptual issues of the philosophy of psychiatry.

So, if RDoC does not offer a criterion for conceptual validity, and is even compatible with views that would determine when a brain is disordered by reference to social criteria, we might again ask why is one of the assumptions of RDoC that mental disorders should be thought of as brain disorders? We think that this claim has a descriptive and a prescriptive dimension. Regarding the descriptive claim, some proponents of RDoC indicate that with new discoveries about the biological and neuroscientific underpinnings of many psychiatric conditions and their implementations in practice we will, as a matter of fact, start to see mental disorders as brain disorders (see, also, Murphy, 2017 for relating this claim to eliminativism in philosophy of psychology). Here is an indicative statement of this descriptive reading:

> recently psychiatry has undergone a tectonic shift as the intellectual foundation of the discipline begins to incorporate the concepts of modern biology, especially contemporary cognitive, affective, and social neuroscience. As these rapidly evolving sciences yield new insights into the neural basis of normal and abnormal behavior, syndromes once considered exclusively as "mental" are being reconsidered as "brain" disorders—or, to be more precise, as syndromes of disrupted neural, cognitive, and behavioral systems. (Insel & Cuthbert, 2015, p. 499)

In this sense, the descriptive claim concerns an empirical question for which time will show whether mental disorders will be reinterpreted as disturbances of brain networks underpinning cognitive and behavioural functioning.

The more interesting reading of the slogan that mental disorders are brain disorders is the one that highlights its prescriptive dimension. Among the proponents of the RDoC type of approach to classification there is a sense that progress in psychiatry will be achieved only if we reconceptualise mental disorders as brain disorders. Here the emphasis is on the *reconceptualization* of mental disorders as brain disorders and not on their *reduction*.

The RDoC approach to psychiatric classification and research can be conceptualised as recommending a *co-evolutionary model, instead of a reductive one, of the relation between top-down descriptions and explanations of mental phenomena and bottom-up descriptions and explanations pertaining to neurobiological mechanisms* (Churchland, 2006; see, also, Bermúdez, 2005 ch. 5). In this case, top-down considerations would specify behavioural and psychological functions the brain is supposed to carry out, and the level of genetics and neural circuits would be accordingly used to specify normal and abnormal ways of their functioning (see section 2 above). In the present context, to discuss the details of how such a co-evolutionary model could apply to the overall RDoC approach is beyond the scope of this work. In fact, that would require addressing, amongst other issues, difficult problems concerning explanation and causation between the different levels touched upon by the different type of considerations described above (Bermúdez, 2005). We limit ourselves, instead, to discuss and clarify a conceptual dimension of the co-evolutionary way of thinking that we think can be plausibly related to RDoC.

We think that the plea for reconceptualization embedded in the slogan that "mental disorders are brain disorders" associated with RDoC might be interpreted as a plea for *explicating* the concept of mental disorders in terms of disturbances or dysfunctions of the neural networks (see Insel & Cuthbert, 2015; White, Rickards, & Zeman, 2012).

In recent discussions of philosophical methodology, explication is often understood as a process of conceptual re-engineering (Brun, 2016). The main aim is to make an imprecise or vague everyday concept (technically called *explicandum*) into a more precise concept (technically called *explicatum*) that is more suitable for theoretical, scientific or even pragmatic purposes (Brun, 2016; Dutilh Novaes, 2018).

Rudolf Carnap (1971), who provided one of the first systematic expositions of explication, gave four requirements that an adequate explication should satisfy:

1. Similarity: the *explicatum* should be similar to some degree in meaning to the *explicandum;*
2. Exactness: the *explicatum* should be more exact in meaning than the *explicandum*;
3. Fruitfulness: the *explicatum* should be fruitful with respect to accomplishing the aims of the research project (for example, for formulating theorems in logic or empirical laws in natural sciences);
4. Simplicity: it is expected that a vague concept could be explicated in more than one way. If two or more possible explications satisfy the above criteria than simplicity could be used to choose among the alternative explications (see Brun, 2016, p. 1215).

Successful explication does not have to satisfy all of the requirements of adequacy to the same degree, rather "an explicatum counts as adequate just in case it meets these criteria to a sufficient degree" (Brun, 2016, p. 1215). Here we will concentrate on the requirement of fruitfulness, because it has been argued that this is the "the crucial requirement for a successful explication" (Dutilh Novaes, 2018, p. 202; see, also, Carus, 2009; Dutilh Novaes & Reck, 2017).

Carnap originally explained a successful explication by using the example of Fish as an explicandum that can be adequately explicated in terms of the *explicatum* Piscis:

> When we compare the explicandum Fish with the explicatum Piscis, we see that they do not even approximately coincide [...]. What was [the zoologists'] motive for [...] artificially constructing the new concept Piscis far remote from any concept in the prescientific language? The reason was that [they] realized the fact that the concept Piscis promised to be much more fruitful than any concept more similar to Fish. A scientific concept is the more fruitful the more it can be brought into connection with

> other concepts on the basis of observed facts; in other words, the more it can be used for the formulation of laws. (Carnap, 1971, p. 6)

According to Carnap, the pre-theoretical or ordinary concept of Fish is replaced by a more precise concept of *Piscis*, which is more fruitful in scientific contexts because it allows us to formulate general (or empirical law-like) statements that in turn underpin successful explanatory practices, increase predictive power and empirical testability. For instance, although whales and dolphins might have fallen under our pretheoretical concept of Fish, they do not fall under the concept of Piscis, because the latter excludes mammals.

In addition to the more theoretical reading of fruitfulness, some authors emphasize that fruitfulness of an *explicatum* can be generally related to our research purposes, whether they be strictly scientific/theoretical or more broadly practical/political. Thus, in general we can say that the fruitfulness of an *explicatum* can be related to its ability to systematize a domain of inquiry according to our purposes or aims (Carus, 2009; Dutilh Novaes, 2018). In this respect, explication is a process of conceptual re-engineering where the criteria of successful explication will depend on what we need these concepts to do for us relative to some project or inquiry we find valuable.[2]

In the context of RDoC, we argue that its assumption that mental disorders should be reconceptualised as brain disorders can be plausibly justified as a plea for making the concept of mental disorder more fruitful in conducting psychiatric research and devising more effective therapies. The aim is to improve psychiatric practice and devise more effective treatments by theoretically and practically unifying knowledge about the biological, cognitive and behavioural systems underpinning what we currently call mental disorders (Insel & Cuthbert, 2015).

The kind of problems that proponents of RDoC see regarding current psychiatric classifications and how to address them testify the explicatory aim of these approaches. Insel and Cuthbert are rather clear about what propels their view that mental disorders should be redefined as brain disorders. First, they state that:

> before research on the convergence of biology and behavior can deliver on the promise of precision medicine for mental disorders, the field must address the imprecise concepts that constrain both research and practice. (Insel & Cuthbert, 2015, p. 499, emphasis added)

Further they emphasize that imprecise concepts:

> like “behavioral health disorders” or “mental disorders” or the awkwardly euphemistic “mental health conditions, ”when juxtaposed against brain science, invite continual recapitulation of the fruitless “mind-body” and “nature-nurture” debates that have impeded a deep understanding of psychopathology. (Insel & Cuthbert, 2015, p. 499)

They see the imprecision of the concept of mental disorder as impeding psychiatric research, and thus replacing this imprecise concept with a concept of brain disorder as referring to

[2] That is why Dutilh Novaes (2018) argues that Carnapian explication can be viewed as a methodology similar in spirit to ameliorative analysis as expounded by Sally Haslanger (2012), where the method of ameliorative analysis refers to an exercise of engineering concepts that is shaped by our political ideals and aims, such as correcting social injustices. These approaches to philosophical methodology are currently discussed under the heading of conceptual engineering. See Cappelen (2018) for a book length discussion of these issues.

"syndromes of disrupted neural, cognitive, and behavioral systems" (Insel & Cuthbert, 2015, p. 499) is expected to transform for the better diagnostic procedures and eventually improve health outcomes.

Similarly, Peter White, Hugh Rikards, and Adam Zeman's (2012) plea for redefining mental disorders as dysfunctions of the central nervous system could be charitably read as suggesting an explication of the concept of mental disorder in terms of brain dysfunctions. They argue that mental disorders should not be grouped separately from the disorders of the brain. Keeping separate categories for mental and brain disorders creates an illusion that psychological functions have a different ontology than brain functions. To ground a division between mental function and brain function is like treating heart function as fundamentally different and disconnected from heart anatomy. If this is implausible in the case of the heart, then it should be implausible in the case of the brain and psychological function.

The negative consequences of this division are, according to them, particularly noticeable in the "bizarre double accounting". For instance, in ICD-10 "dementia in Alzheimer's disease" is classified as a mental disorder (F00), while Alzheimer's disease is classified under neurology (G30)" (White *et al.*, 2012, p. 2). Furthermore, this conceptual division has negative practical consequences on institutional division between psychiatry and the rest of medicine. Thus, White and colleagues argue that "changing the classification" by reconceptualising mental disorders as disorders of the nervous systems "will epitomise an intellectual shift with far reaching beneficial consequences" which are expected to include research, medical, and social benefits (White *et al.*, 2012).

Thus, construing mental disorders as brain disorders should invite a more integrative perspective by thinking about the brain as the seat of psychological and behavioural functions (Jefferson, 2020). Given these facts, currently unsuccessful attempts at treating mental disorders and prospects for advancement in devising treatments and improving health outcomes is likely premised on our (in)ability to take into consideration the wealth of current and future knowledge of the biological factors that underpin our psychological and behavioural functions and their characteristic patterns of malfunctioning.

The idea that we should reconceive mental disorders as brain disorders should not be read as endorsing crude versions of explanatory reductionism. Instead, it should be read as an invitation to avoid the pitfalls of drawing arbitrary wedges within the field of medicine and engage the promising projects that aim at improving research, treatment, and health outcomes by readily integrating psychiatry with the rest of medicine.

5. Conclusion

There are several theoretical and empirical considerations for moving beyond syndrome-based classifications of mental disorders. RDoC represents a bold but a systematic attempt to foster this advancement. Besides the considerable empirical difficulties that this approach faces (Lilienfeld, 2014), there are important theoretical and conceptual issues that need to be addressed. Some of these difficulties emerge upon reflection on the plausible readings of the slogan "mental disorders are brain disorders", that is often associated with biologically grounded approaches to classification of mental disorders.

We have offered conceptual and theoretical considerations for concluding that the slogan above should not be read as an explanatory reductionist thesis. Moreover, current formulations of the biologically based classifications do not appear to involve commitments on the issue of conceptual validity. In these approaches there is no explicit or strictly logically required statement about what confers a disorder status to the investigated conditions.

Finally, while a descriptive reading of the thesis that "mental disorders are brain disorders" is an interesting prediction about our future conceptual and medical practices, we think that the most important reading is a prescriptive one. This is the idea that the categorisation of mental disorders should be motivated by the assumption that they are brain disorders.

We have shown how a prescriptive reading of "mental disorders are brain disorders" slogan should be viewed as recommending a revisionary project for conceptually reconfiguring the categories of mental disorders. Such a plausible and promising project of reconfiguration confers to biological variables a deserved, although not exclusive, role in the classification of mental disorders. We have further clarified this conceptual reconfiguration in the terms of a notion of explication, that, firstly formulated by Carnap, has an increasingly influential currency in contemporary philosophy.

REFERENCES

Aboraya A., France C., Young J., Curci K. & LePage J. (2005). The Validity of Psychiatric Diagnosis Revisited. *Psychiatry (Edgmont)*, *2*(9), pp. 48–55;

American Psychiatric Association & American Psychiatric Association (Eds.). (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (5th ed). Washington, D.C: American Psychiatric Association;

Amoretti M. C. & Lalumera E. (2019a). A Potential Tension in DSM-5: The General Definition of Mental Disorder versus Some Specific Diagnostic Criteria. *The Journal of Medicine and Philosophy*, *44*(1), pp. 85–108. https://doi.org/10.1093/jmp/jhy001;

Amoretti M. C. & Lalumera E. (2019b). 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*(4), pp. 321–337. https://doi.org/10.1007/s11017-019-09499-4;

Beebee H., & Sabbarton-Leary, N. (2010). Are Psychiatric Kinds Real? *European Journal of Analytic Philosophy*, *6*(1), pp. 11–27;

Bermúdez J. L. (2005). *Philosophy of Psychology: A Contemporary Introduction*. London: Routledge;

Bethesda (Ed.). (2011). *National Heart, Lung, and Blood Institute: In NHLBI Fact Book, Fiscal Year*. Retrieved from http://www.nhlbi.nih.gov/about/factpdf.htm;

Bolton D. (2008). *What is mental disorder? An essay in philosophy, science, and values*. Oxford: Oxford University Press;

Boorse C. (2014). A Second Rebuttal On Health. *The Journal of Medicine and Philosophy*, *39*(6), pp. 683–724. https://doi.org/10.1093/jmp/jhu035;

Borsboom D., Cramer A. O. J. & Kalis A. (2019). Brain disorders? Not really: Why network structures block reductionism in psychopathology research. *Behavioral and Brain Sciences*, *42*, pp. 1–54 https://doi.org/10.1017/S0140525X17002266;

Boyd R. (1991). Realism, Anti-foundationalism and the Enthusiasm for Natural Kinds. *Philosophical Studies*, *61*(1–2), pp. 127–48;

Brazil I. A. & Cima M. (2016). Contemporary Approaches to Psychopathy. In M. Cima (Ed.), *The Handbook of Forensic Psychopathology and Treatment* (pp. 206–226). London and New York: Routledge;

Brazil I. A., van Dongen J. D. M., Maes J. H. R. Mars R. B. & Baskin-Sommers A. R. (2018). Classification and treatment of antisocial individuals: From behavior to biocognition. *Neuroscience & Biobehavioral Reviews*, *91*, pp. 259–277. https://doi.org/10.1016/j.neubiorev.2016.10.010;

Brun G. (2016). Explication as a Method of Conceptual Re-engineering. *Erkenntnis*, *81*(6), pp. 1211–1241 https://doi.org/10.1007/s10670-015-9791-5;

Brzović Z. (2018). Natural Kinds. In *Internet Encyclopedia of Philosophy*. Retrieved from https://www.iep.utm.edu/nat-kind/;

Brzović Z., Hodak J., Malatesti L., Šendula-Jengić V., & Šustar P. (2016). Problem klasifikacije u filozofiji psihijatrije: Slučaj psihopatije (Eng. The Problem of Classification in the Philosophy of Psychiatry: The Case of Psychopathy). *Prolegomena*, *15*(1), pp. 21–41;

Brzović Z., Jurjako M. & Malatesti L. (2018). Il Modello Medico Forte E I Disturbi Antisociali Della Personalità (Eng. The Strong Medical Model and Antisocial Personality Disorders). *Sistemi Intelligenti*, *30*(1), pp. 175–188;

Brzović Z., Jurjako M. & Šustar P. (2017). The kindness of psychopaths. *International Studies in the Philosophy of Science*, *31*(2), V189–211. https://doi.org/10.1080/02698595.2018.1424761;

Buckholtz J. W., & Meyer-Lindenberg, A. (2012). Psychopathology and the human connectome: Toward a transdiagnostic model of risk for mental illness. *Neuron*, *74*(6), pp. 990–1004 https://doi.org/10.1016/j.neuron.2012.06.002;

Cappelen H. (2018). *Introduction to Conceptual Engineering*. Oxford: Oxford University Press;

Carnap R. (1971). *Logical foundations of probability* (4. impr). Chicago: University of Chicago Press;

Carus A. W. (2009). *Carnap and twentieth-century thought: Explication as enlightenment*. Cambridge: Cambridge Univ. Press;

Churchland P. S. (2006). *Neurophilosophy: Toward a unified science of the mind-brain* (15th ed.). Cambridge, Mass.: MIT Press;

Clark L. A., Cuthbert B., Lewis-Fernández R., Narrow W. E. & Reed G. M. (2017). Three Approaches to Understanding and Classifying Mental Disorder: ICD-11, DSM-5, and the National Institute of Mental Health's Research Domain Criteria (RDoC). *Psychological Science in the Public Interest*, *18*(2), pp. 72–145. https://doi.org/10.1177/1529100617727266;

Cooke D. J. (2018). Psychopathic Personality Disorder: Capturing an Elusive Concept. *European Journal of Analytic Philosophy*, *14*(1), 15–32. https://doi.org/10.31820/ejap.14.1.1;

Cooper R. V. (2005). *Classifying madness: A philosophical examination of the diagnostic and statistical manual of mental disorders*. Dordrecht and New York: Springer;

Cooper R. V. (2013). Avoiding False Positives: Zones of Rarity, the Threshold Problem, and the DSM Clinical Significance Criterion. *The Canadian Journal of Psychiatry*, *58*(11), pp. 606–611. https://doi.org/10.1177/070674371305801105;

Cuthbert B. N. (2014). The RDoC framework: Facilitating transition from ICD/DSM to dimensional approaches that integrate neuroscience and psychopathology. *World Psychiatry*, *13*(1), pp. 28–35. https://doi.org/10.1002/wps.20087;

Cuthbert B. N. & Insel T. R. (2013). Toward the future of psychiatric diagnosis: The seven pillars of RDoC. *BMC Medicine*, *11*(1). https://doi.org/10.1186/1741-7015-11-126;

Dutilh Novaes C. (2018). Carnapian explication and ameliorative analysis: A systematic comparison. *Synthese*. https://doi.org/10.1007/s11229-018-1732-9;

Dutilh Novaes, C., & Reck, E. (2017). Carnapian explication, formalisms as cognitive tools, and the paradox of adequate formalization. *Synthese*, *194*(1), pp. 195–215. https://doi.org/10.1007/s11229-015-0816-z;

Fulford K. W. M. (1989). *Moral Theory and Medical Practice*. Cambridge University Press;

Hare R. D. (2003). *The Hare Psychopathy Checklist Revised* (2nd ed.). Toronto, ON: Multi-Health Systems;

Haslam N. (2014). Natural kinds in psychiatry: Conceptually implausible, empirically questionable, and stigmatizing. In H. Kincaid & J. A. Sullivan (Eds.), *Classifying psychopathology* (pp. 11–28). Cambridge, Mass.: The MIT Press;

Haslam N., Holland E. & Kuppens P. (2012). Categories versus Dimensions in Personality and Psychopathology: A Quantitative Review of Taxometric Research. *Psychological Medicine*, *42*(5), pp. 903–920. https://doi.org/10.1017/S0033291711001966;

Haslanger S. (2012). *Resisting reality: Social construction and social critique*. Oxford: Oxford University Press;

Insel T. R., Cuthbert B., Garvey M., Heinssen R., Pine D. S., Quinn K., ... Wang P. (2010). Research Domain Criteria (RDoC): Toward a New Classification Framework for Research on Mental

Disorders. *American Journal of Psychiatry*, *167*(7), pp. 748–751. https://doi.org/10.1176/appi.ajp.2010.09091379;
Insel T. R. & Cuthbert B. N. (2015). Brain Disorders? Precisely. *Science*, *348*(6234), pp. 499–500. https://doi.org/10.1126/science.aab2358;
Jefferson A. (2018). What does it take to be a brain disorder? *Synthese*. https://doi.org/10.1007/s11229-018-1784-x;
Jurjako M. (2019). Is psychopathy a harmful dysfunction? *Biology & Philosophy*, *34*(1), 5. https://doi.org/10.1007/s10539-018-9668-5;
Jurjako M. & Malatesti L. (2018a). Neuropsychology and the Criminal Responsibility of Psychopaths: Reconsidering the Evidence. *Erkenntnis*, *83*(5), pp. 1003–1025. https://doi.org/10.1007/s10670-017-9924-0;
Jurjako M. & Malatesti L. (2018b). Psychopathy, executive functions, and neuropsychological data: A response to Sifferd and Hirstein. *Neuroethics*, *11*(1), pp. 55–65. https://doi.org/10.1007/s12152-016-9291-6;
Jurjako M., Malatesti L. & Brazil I. A. (2019a). Biocognitive classification of antisocial individuals without explanatory reductionism. Accepted for publication in *Perspectives on Psychological Science*, DOI: 10.1177/1745691620904160
Jurjako M., Malatesti L. & Brazil I. A. (2019b). Some Ethical Considerations about the Use of Biomarkers for the Classification of Adult Antisocial Individuals. *International Journal of Forensic Mental Health*, *18*(3), pp. 228–242. https://doi.org/10.1080/14999013.2018.1485188;
Kendler K. S., Zachar P. & Craver C. (2011). What Kinds of Things are Psychiatric Disorders? *Psychological Medicine*, *41*(6), pp. 1143–1150. https://doi.org/10.1017/S0033291710001844;
Khalidi M. A. (2013). *Natural Categories and Human Kinds: Classification in the Natural and Social Sciences*. Cambridge: Cambridge University Press;
Kingma E. (2013). Naturalist Accounts of Mental Disorder. In K. W. M. Fulford, M. Davies, R. G. T. Gipps, G. Graham, J. Z. Sadler, G. Stanghellini, & T. Thornton (Eds.), *The Oxford handbook of philosophy and psychiatry* (pp. 363–384). Oxford: Oxford University Press;
Kingma E. (2014). Naturalism about Health and Disease: Adding Nuance for Progress. *The Journal of Medicine and Philosophy*, *39*(6), pp. 590–608. https://doi.org/10.1093/jmp/jhu037;
Lilienfeld S. O. (2014). The Research Domain Criteria (RDoC): An Analysis of Methodological and Conceptual Challenges. *Behaviour Research and Therapy*, *62*, pp. 129–139. https://doi.org/10.1016/j.brat.2014.07.019;
Lilienfeld S. O., Smith S. F. & Watts A. L. (2013). Issues in diagnosis: Conceptual issues and controversies. In W. E. Craighead, Mikllowitz, D. J., & Craighead, L. W. (Eds.), *Psychopathology: History, diagnosis, and empirical foundations* (pp. 1–35). Hoboken, NJ: John Wiley & Sons;
Maibom H. L. (2018). What Can Philosophers Learn from Psychopathy? *European Journal of Analytic Philosophy*, *14*(1), pp. 63–78;
Međedović J., Bulut T., Savić D. & Đuričić N. (2018). Delineating Psychopathy from Cognitive Empathy: The Case of Psychopathic Personality Traits Scale. *European Journal of Analytic Philosophy*, *14*(1), pp. 53–62;
Međedović J., Petrović B. P., Kujačić D. K., Đorić J. Ž. Đ. Ž. & Savić M. S. (2015). What is the optimal number of traits to describe psychopathy? *Primenjena Psihologija*, *8*(2), pp. 109–130;
Murphy D. (2006). *Psychiatry in the Scientific Image*. Cambridge, Mass.: The MIT Press;
Murphy D. (2017). Can Psychiatry Refurnish the Mind? *Philosophical Explorations*, *20*(2), pp. 160–174. https://doi.org/10.1080/13869795.2017.1312499;
Nagel E. (1987). *The Structure of Science: Problems in the Logic of Scientific Explanation* (2nd ed.). Indianapolis, Ind.: Hackett;

Rosenberg Larsen R. (2018). False-Positives in Psychopathy Assessment: Proposing Theory-Driven Exclusion Criteria in Research Sampling. *European Journal of Analytic Philosophy*, *14*(1), pp. 33–52;
Samuels R. (2009). Delusion as a Natural Kind. In Broome, Matthew M. & Bortolotti, Lisa (Eds.), *Psychiatry as Cognitive Neuroscience: Philosophical Perspectives* (pp. 49–79). Oxford: Oxford University Press;
Slater M. H. (2015). Natural Kindness. *The British Journal for the Philosophy of Science*, *66*(2), pp. 375–411;
Szasz T. S. (1974). *The myth of mental illness: Foundations of a theory of personal conduct* (Rev. ed). New York: Harper & Row;
Tabb K. (2015). Psychiatric Progress and the Assumption of Diagnostic Discrimination. *Philosophy of Science*, *82*(5), pp. 1047–1058. https://doi.org/10.1086/683439;
Tabb K. (2019a). Philosophy of psychiatry after diagnostic kinds. *Synthese*, *196*(6), pp. 2177–2195. https://doi.org/10.1007/s11229-017-1659-6;
Tabb K. (2019b). Why not be pluralists about explanatory reduction? *Behavioral and Brain Sciences*, *42*, e27. https://doi.org/10.1017/S0140525X18002054;
Tsou J. Y. (2011). The Importance of History for Philosophy of Psychiatry: The Case of the DSM and Psychiatric Classification. *Journal of the Philosophy of History*, *5*(3), 446–470. https://doi.org/10.1163/187226311X599907;
Tsou J. Y. (2016). Natural Kinds, Psychiatric Classification and the History of the DSM. *History of Psychiatry*, *27*(4), 406–424;
Wakefield J. C. (1992). The concept of mental disorder. On the boundary between biological facts and social values. *The American Psychologist*, *47*(3), 373–388;
Wakefield J. C. (2014). Wittgenstein's nightmare: Why the RDoC grid needs a conceptual dimension. *World Psychiatry*, *13*(1), 38–40. https://doi.org/10.1002/wps.20097;
White P. D., Rickards H. H. & Zeman A. Z. J. (2012). Time to end the distinction between mental and neurological illnesses. *BMJ*, *344*, e3454. https://doi.org/10.1136/bmj.e3454;
Wiecki T. V., Poland J. & Frank M. J. (2015). Model-Based Cognitive Neuroscience Approaches to Computational Psychiatry: Clustering and Classification. *Clinical Psychological Science*, *3*(3), 378–399. https://doi.org/10.1177/2167702614565359;
World Health Organization (Ed.). (1992). *The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines*. Geneva: World Health Organization.