Synthese (2024) 203:24
https://doi.org/10.1007/s11229-023-04427-5

ORIGINAL RESEARCH

Broken wills and ill beliefs: Szaszianism, expressivism, and the doubly value-laden nature of mental disorder

Miguel Núñez de Prado-Gordillo[1]

Received: 2 September 2023 / Accepted: 8 November 2023

Abstract
Critical psychiatry has recently echoed Szasz's longstanding concerns about medical understandings of mental distress. According to Szaszianism, the analogy between mental and somatic disorders is illegitimate because the former presuppose psychosocial and ethical norms, whereas the latter merely involve deviations from natural ones. So-called "having-it-both-ways" views have contested that social norms and values play a role in *both* mental and somatic healthcare, thus rejecting that the influence of socio-normative considerations in mental healthcare compromises the analogy between mental and somatic disorders. This paper has two goals. Firstly, I argue that having-it-both-ways views fail to provide a compelling answer to Szasz's challenge. The reason is that what is essential to Szasz's argument is not that mental disorder attributions involve value judgements, but that mental attributions *in general* do. Mental disorders are thus doubly value-laden and, *qua* mental, only metaphorically possible. To illustrate this, I construe Szasz's view and Fulford's having-it-both-ways approach as endorsing two different kinds of expressivism about mental disorders, pointing out their different implications for the analysis of delusions. Secondly, I argue, against Szaszianism, that Szasz's rejection of the analogy is relatively irrelevant for discussions about the appropriateness of medicalizing mental distress. Specifically, I draw from socio-normative approaches to the psychopathology/social deviance distinction and mad and neurodiversity literature to argue that a) it is still possible to distinguish social deviance from psychopathology once we reject the analogy; and b) that *both* medicalizing and normalizing attitudes to mental distress can harmfully wrong people from relevant collectives.

Keywords Expressivism · Philosophy of psychiatry · Critical psychiatry · Mental disorder · Social normativity · Delusion

---

✉ Miguel Núñez de Prado-Gordillo
m.nunezdepradogordillo@uu.nl

[1] Department of Philosophy and Religious Studies, Utrecht University, Janskerkhof 13, 3512 BL Utrecht, The Netherlands

Published online: 05 January 2024

In recent years, public controversies surrounding the publication of the DSM-5 and the DSM-5-TR (American Psychiatric Association, 2013, 2022) have led to a resurgence of Szasz's (1960, 1961a, 2011) longstanding criticisms against the medicalization of mental distress. A primary example are so-called "critical psychiatry" views which have gained increasing popularity in recent years (Johnstone & Boyle, 2018; Kinderman et al., 2013; Middleton & Moncrieff, 2019; Moncrieff, 2020; Pickard, 2009; Read & Moncrieff, 2022). Although not all critical views fully endorse Szasz's project, they undoubtedly rely on some of his core theoretical commitments (Chapman, 2023; Morgan, 2023). These views share with other critical perspectives the analysis of the shortcomings of the medical model and the emphasis on the need for alternative ways of conceptualizing and addressing the experiences of mental health-related collectives (see Chapman, 2020a; Curtis et al., 2000; Frazer-Carroll, 2023); however, what distinguishes them is their reliance on similar varieties of Szasz's core claim: that describing mental distress in medical terms (e.g., as disorders, diseases, illnesses, pathologies, etc.[1]) amounts to an *illegitimate analogy* between mental and somatic disorders -one which purportedly plays a mere self-serving role in the maintenance of the perceived scientific legitimacy of psychiatry and the superiority of psychiatric discourse over other accounts of distress. Szasz's attack on the 'myth of mental illness' was based on the observation that, while somatic disorders involve deviations from *natural* norms -i.e., those governing "the structural and functional integrity of the human body" (1960, p. 114)- mental disorders involve deviations from *psychosocial* and *ethical* ones. The upshot of this illegitimate analogy, according to Szasz, is the medicalizing of mere 'problems in living' and the pathologizing of social deviance, i.e., non-conformity to social values and expectations.

Szasz's stark observations about the legitimacy of psychiatry have been the object of continued countercriticism, mostly focused on the discussion of the broader notion of disorder. There are two common counter arguments. Firstly, so-called *naturalist* accounts attempt to ground the notion of disorder -or some central component of it, e.g., "dysfunction"- in some sort of natural norm that makes room for mental disorders as well (Boorse, 1976a, 1976b, 2014; Kendell, 1975; Wakefield, 1992, 2007). This natural norm is typically defined in terms of the physiologically normal functioning of an organism's internal mechanisms (Boorse, 1976a, 1976b, 2014) or its evolutionary history (Wakefield, 1992, 2007). However, these approaches have been largely criticized on account of their inability to offer truly 'value-free' analyses of disorder or *any* of its supposed central components (Chapman, 2023; Fulford, 1999; Kingma, 2013). Alternatively, *having-it-both-ways* views (Varga, 2015) argue that ascriptions of bodily or mental disorder, dysfunction, and so on, necessarily involve some sort of evaluative judgement, whereby a person's biological or dispositional makeup is assessed against a bedrock of social norms and values (Fulford, 1989; Graham, 2010; Thornton, 2007). On this view, mental disorders are in fact analogous to somatic ones; not because they

[1] Some authors clearly distinguish among these medical notions, precisely in response to Szasz's and others' challenges. For instance, Boorse (1975, 2014) distinguishes 'disease' or 'pathology', which he views as purely descriptive, from 'illness', which may involve an evaluative component. However, as shown below, Szasz's argument stands independently of the exact kind of medical language used to describe psychiatric conditions. Unless otherwise specified, here I'll use these terms interchangeably.

don't presuppose social norms and values, but precisely because they do. Having-it-both-ways theorists take this argument to reveal the ethical dimension intrinsic to *all* medical disciplines; and, in so doing, to undermine Szasz's attack on the "foundational myth" of psychiatry and preserve the legitimacy of medical understandings of mental malaise.

In this paper, I will mainly focus on the latter kind of counterargument to Szasz's criticisms. My primary goal is to argue that having-it-both-ways views, although right in pointing out the irreducible socionormative character of ascriptions of disorder and similar others, nonetheless fail to come to grips with Szasz's main point of contention. Therefore, they fail to provide a convincing account of the analogy between mental and somatic disorders. The reason is that Szasz's critique concerns not (or not primarily) the 'disordered' aspect of mental disorders, but the 'mental' one: what is essential to his argument is not that mental disorder attributions (unlike somatic disorder ones) involve value judgements, but rather that mental attributions *in general* (unlike bodily descriptions) do. I develop this point further by recasting having-it-both-ways approaches as endorsing some sort of expressivist analysis of pathology ascriptions; by contrast, I argue, Szasz's critique can be better understood as an expressivist analysis of mental ascriptions (Fernández Castro, 2023; Pérez-Navarro et al., 2019), akin to the non-descriptivist views of mind of early analytic philosophers like Ryle or Wittgenstein. The resulting "doubly value-laden" character of mental disorders is what primarily justifies Szasz's claim that, if genuinely mental, they are only metaphorically possible. I illustrate this point by comparing Wilkinson's (2020) recent expressivist analysis of delusion attributions with Pérez-Navarro et al. (2019) doxastic expressivism; not only delusion ascriptions express our adherence to certain (epistemological) values and standards, but belief ascriptions in general do so.

My secondary goal is to show that the argumentative appeal of Szaszianism ends right here. The main reason is that his success in attacking the analogy between mental and somatic disorders is orthogonal to his other claims about the legitimacy of medical discourse. Whether the argument for the analogy stands or fails is thus relatively irrelevant for discussions about the appropriateness of medicalizing mental distress. I illustrate this point by drawing from various sources, including a) various socio-normative accounts of the boundary or demarcation problem, i.e., approaches which aim to distinguish social deviance from psychopathology in terms of irreducibly social, rather than natural norms and values (de Haan, 2020; Rashed, 2021); and b) mad and neurodiversity activism and scholarship, which illustrate how *both* medicalizing and normalizing attitudes to mental distress can harmfully wrong people from relevant collectives (Carel, 2023; Chapman, 2023b; Chapman & Carel, 2022; Frazer-Carroll, 2023; Sedgwick, 1982).

In Sect. 1, I introduce the main motivations behind Szasz's attack on the analogy between mental and somatic disorders. In Sect. 2, I introduce the main counterarguments against Szaszianism, focusing primarily on Fulford's having-it-both-ways view. In Sects. 3 and 4, I compare Szasz's and Fulford's views in terms of the different kinds of expressivism that they seem to endorse, applying them to the case of delusion. This illustrates why having-it-both-ways views fail to provide a convincing reason for the analogy between mental and somatic disorders. Finally, in Sect. 5, I discuss why this doesn't necessarily lead to some of the main criticisms that Szasz raised against the medical understanding of mental distress.

1 The Szaszian crusade

The role of social norms and values in mental healthcare has been the object of harsh debates for at least the last six decades, since Szasz (1961a) and other critical voices questioned the scientific legitimacy of psychiatry as a branch of medicine. These debates have revolved around two inter-related issues: (a) *the analogy problem*, concerning the possibility of understanding mental conditions in terms of physical illnesses or disorders; and (b) *the boundary* or *demarcation problem*, concerning the possibility of distinguishing the "mad" from the "bad", i.e., telling apart psychopathology from mere social deviance. *Naturalist* approaches to both issues have been typically associated with medical models of psychiatric conditions -at least in their strong interpretation (Murphy, 2013). On an extreme interpretation of this perspective, social norms and values play no role whatsoever in psychiatric theory or practice: psychiatric conditions are medical conditions (e.g., mental disorders) that stand exactly on a par with physical illnesses, and this is precisely what distinguishes them from mere cases of social deviance. *Normativist* perspectives, by contrast, reject this "value-free" or "purely descriptive" view of psychiatric phenomena. Szaszianism and other critical perspectives that emerged during the 1960s and the 1970s are often understood as normativist positions. On an extreme understanding of it, psychiatric conditions are mere instances of social deviance, their alleged medical status amounting to "coercion masquerading as medical treatment" (Szasz, 2001, p. 140).

Ever since these early debates, medical and psychological anthropologists, cultural psychiatrists, sociologists, and philosophers of psychiatry have remarked the influence of social norms and values in different dimensions of mental health (e.g., in the expression of symptoms, lived experience, the origin or maintenance of psychopathology, etc.). This influence of socionormative considerations is nowadays widely accepted by most researchers, to a lesser or greater extent. Szaszianism, however, goes a step further. In his lifelong, uncompromising crusade against psychiatric and psychotherapeutic institutions, Szasz drew from these observations to contend that talk about 'mental illness' amounted to "a logically highly dubious proposition" (1961b, p. 59), "fallacious reasoning" (1960, p. 114), "a worthless and misleading definition" (1961a, p. 262), "a metaphor" (1961a, p. 267; 2001, p. 91; 2008, Chapter 1, section III; 2011, p. 180), a "socially useful fiction" (1977, p. xix), or, in his most famous expression, a plain *myth* (1960, 2011). His main argument draws from a Virchowian, "materialist-scientific" understanding of the concept of illness as "functional or structural abnormality of cells, tissues, or organs" (2001, p. 13), which he subscribed to during his whole career. The crucial point is that while physical illness involves deviations from *biological norms* governing "the structural and functional integrity of the human body", the norms at stake in mental health "must be stated in terms of *psychosocial, ethical,* and *legal* concepts" (1960, p. 114).

For Szasz, describing mental distress in medical terms amounted to a mistaken and misleading metaphor, a conflation of two "types of languages or modes of representation" (1961a, p. 78): the language of the mind, which for him was inevitably tied to the language of meaning, agency, and the view of individuals as *persons*; and the language of the body, linked to the analysis of individuals as biological systems. This conflation wasn't innocent, in his view, but constituted a "self-serving psychiatric rationalization" (2000, p. 11) whose purpose was to protect psychiatry's self-image as a medical and scientific discipline. Most importantly, this 'myth' subserved important social control functions: to disguise state coercion as medical treatment to enforce norm-conformity in the socially deviant, and to justify the "therapeutic state" (1977) and its totalitarian practices of "psychiatric slavery" and "psychiatric rape" (2009, p. ix) -practices which the author often compared to medieval witch-hunt (1960, 1961a) and "statist" systems like the former Soviet Union or Nazi Germany (2001).

Contemporary endorsement of Szasz's core claims has increased in recent years following the reliability and validity crises of traditional nosological tools like the DSM (Johnstone & Boyle, 2018; Kinderman et al., 2013; Middleton & Moncrieff, 2019; Moncrieff, 2020). Albeit in more balanced tone, critics of the "language of disorder" (Kinderman et al., 2013) reject the medicalizing and pathologizing of mental distress, advocating instead for its *normalizing*, i.e., for viewing it as part of the normal range of human experiences. Szasz's attack on the analogy between mental and somatic disorders takes central stage in this criticism: mental disorders, unlike somatic ones, cannot be understood as deviations from bodily anatomy and function; rather, they involve a deviation from psychosocial norms and expectations (Johnstone & Boyle, 2018; Moncrieff, 2020; see also Chapman, 2023; Morgan, 2023). Explanation for them must be cast out in the language of agency and meaning, not the language of disorder. We need psychosocial, rather than medical, tools for understanding and addressing them properly. The idea is not that we need to consider psychosocial factors to fully explain psychopathology -as psychiatric integrationism demands (de Haan, 2020; Engel, 1977); rather, Szaszian approaches entail some form of eliminativism about "mental disorder" and other medical ways of describing madness (Kingma, 2013).

2 Responses to Szasz

Szasz's views have faced considerable opposition from many different frameworks and standpoints (see Schaler, 2004). Much counterargument has focused on the Virchowian, "materialist-scientific" notion of disorder underlying his critique. A common understanding of his position views it as entailing that what disorder is can be captured in purely descriptive, value-free terms (e.g., Thornton, 2007); mental disorder diagnoses, by contrast, would always involve a socionormative, value-laden element, and thus would not count as real disorders. To put it in contemporary terms: while Szasz is a *naturalist* about disorder, he is also a *normativist* about mental disorder (Fulford, 1989; Thornton, 2007). Likewise, we can distinguish two main responses to Szasz, both of which contest his Virchowian view of disorder. Naturalist responses, on the one hand, agree with Szasz that disorders or some of their core components can be fully characterized in purely value-free terms; however, they reject his Virchowian definition in favor of alternative naturalistic accounts that make room for mental disorders as well. By contrast, normativist responses agree with Szasz that what mental disorder is cannot be specified without reference to social norms and values; however, they contend that the same applies to *any* kind of disorder -or any of the central elements that define the notion. Thus, these approaches "have it both ways" (Varga, 2015): both somatic and mental disorders are value-laden. Although this paper focuses primarily on this second kind of counterargument, I will first briefly discuss naturalist responses, whose problems motivate having-it-both-ways positions.

2.1 Naturalism: mental disorders as mental dysfunctions

It is important to note that most naturalist accounts also admit that there is always a socionormative component involved in disorder attribution practices; their key commitment, however, is that at least a central element in the theoretical definition of notions like "pathology", "disorder", "illness", or "disease" can be specified in purely descriptive terms. In the two most influential naturalist accounts, Boorse's (1975, 2014) *biostatistical theory* and Wakefield's (1992, 2007) *harmful dysfunction analysis,* the key descriptive element is that of *dysfunction*. These views replace Szasz's Virchowian definition of disorder by a functional analysis of the notion, which appeals to the "natural" or "normal function" of an organism's mechanisms as the ultimate criterion to distinguish (theoretical) health from pathology or disorder. On the one hand, Boorse's biostatistical theory is based on a goal-based account of function (Boorse, 1976b). This view emphasizes how the functioning of an organism's internal mechanisms typically contribute to their survival and reproduction. Boorse focuses on the statistically typical functioning of an organism's internal mechanisms relative to a relevant "reference class", i.e., "a natural class of organisms of uniform functional design; specifically, an age group of a sex of a species" (2014, p. 684). 'Health' is thus identified with 'normal function' -i.e., "a statistically typical contribution by [a part or process within members of the reference class] to their individual survival [or] reproduction" (2014, p. 684); likewise, 'pathology' is defined as the "statistically species-subnormal" functioning of such mechanisms within the individual, relative to a relevant reference class.

By contrast, Wakefield's harmful dysfunction analysis adopts an etiological, selectionist account of function. On his hybrid approach, disorders are *harmful dysfunctions*: although the 'harmful' component is indeed an evaluative one, dependent on social norms and conventions, 'dysfunction' is supposed to be strictly value-free. This descriptive component refers to "the failure of a mechanism to perform its natural function" (1992, p. 383), whereby this natural function is defined in terms of the effects of these mechanisms that explain their selection throughout the species' evolutionary history. In the case of mental disorders, the relevant dysfunctions are those that affect *mental* mechanisms (e.g., perceptual or motivational processes); it is these which allow us to distinguish what counts as mental disorder from mere social deviance.

Despite their differences, both accounts understand disorders as primarily the result of dysfunctions of internal mechanisms involved in organismic fitness and survival.

Both admit that disorder ascriptions may involve an evaluative element (i.e., the negative valuing of the effect of these dysfunctions, e.g., as harmful)[2]; however, at least the key dysfunction element in the definition of disorder remains normatively neutral.

This claim has received substantial criticism. A central worry here concerns naturalists' failure to show how disorders or their alleged objective dysfunction component are in fact 'value-free'. This is shown in the process of fixing the relevant reference class or evolutionary time-frame against which to assess a mechanism's proper function (Chapman, 2021; Cooper, 2002; Fulford, 1999; Kingma, 2007, 2013). In Boorse's biostatistical theory, reference classes help us determine whether a certain process within a member of the target population displays subnormal performance; otherwise, processes only found in certain subpopulations at specific developmental stages would count as statistically subnormal functioning, hence as pathological. Boorse circumvents this problem by choosing age and sex cohorts as the relevant reference classes. However, as Kingma (2013) points out, Boorse fails to provide a "non-circular, value-free justification" for assuming that "the reference classes Boorse admits, and thereby the [biostatistical theory], are value-free or interest-independent" (p. 370). If, for instance, we fixed the relevant reference class to only include people experiencing ADHD-related attentional difficulties, neurotypical attention processes would automatically count as pathological. The problem, as Kingma puts it, is that the choice of the relevant reference class is "likely to reflect prior, and possibly value-laden, assumptions about which groups are normal and healthy" (2013, p. 370).[3]

Wakefield's theory faces a similar problem when attempting to fix the relevant timeframe "at which selection pressures should be considered relevant for the attributions of functions" (Cooper, 2002, p. 268). Wakefield defines a mechanism's natural function by focusing on what explains its selection in our remote evolutionary past; however, we could also focus on current selection pressures, or even consider both remote and current environmental conditions in our assessment. As with the fixing of reference classes, we seemingly lack principled, value-free reasons why we should opt for one or another timeframe.

[2] Unlike Wakefield, Boorse does not think that disorder ascriptions *necessarily* involve an evaluative element: he just admits that certain "disease-plus" concepts useful in medical *practice* -not in medical theory-may involve it (2014, pp. 689–690). I would like to thank an anonymous reviewer for pointing this out.

[3] In response to similar criticisms, Boorse (2014, pp. 693, 703, 708) notes that his theory, and specifically his choice of sex/age cohorts as the relevant reference classes, is not evaluative; it just attempts to describe how medical theorists use the concept of pathology. He adds that "if lay people misunderstand medical concepts, that is not my or medicine's fault" (p. 703), and that political contestation of medical uses of these concepts cannot be a reason for de-pathologizing some condition. But this just restates the problem of value-ladenness: while descriptions of how medical experts use pathology concepts might well be value-neutral, whether these uses should be given *precedence* over others when considering the pathological status of some condition surely is not. In cases that highlight the political relevance of demarcating psychopathology from mere social deviance (e.g., homosexuality), medical uses of pathology concepts are *precisely* part of what is at stake.

2.2 Having-it-both-ways: disorder ascriptions as value-laden

These observations have motivated a second kind of rebuttal of Szasz's arguments, which rejects the possibility of defining disorder or any of its core components in purely value-free terms. The upshot of this countercriticism is that, while conceding that mental disorders are value-laden, it rejects Szasz's view that somatic (or 'literal') disorders are not (Fulford, 1989; Graham, 2010; Thornton, 2007, see Varga, 2015). Despite their differences, these having-it-both-ways approaches share the emphasis on the irreducible socionormative character of disorder ascriptions (including its dysfunction component): to ascribe a disorder (or dysfunction) to someone involves *evaluating* their biological or dispositional makeup against a bedrock of social norms and values. On this view, mental disorders are in fact analogous to somatic ones; not because they don't presuppose social norms and values, but precisely because they do. This would allegedly undermine Szasz's challenge, preserving the legitimacy of medical understandings of mental distress.

Fulford's work (1989, 1999; Fulford & van Staden, 2013; see Thornton, 2007) constitutes one of the most influential examples of this kind of strategy. Fulford (1989) attacked what he called the "conventional view" of medicine, according to which its subject matter can or must be defined in purely descriptive terms; particularly, in terms of 'dysfunction', understood as a strictly value-free notion that in turn grounds other medical notions. This conventionalist view constitutes the first conceptual premise shared by both Szasz and his critics; what they disagree on is whether mental health conditions fall under such value-free definition of the scope of medicine. Drawing from the aforementioned criticisms against naturalist approaches, Fulford questions the viability of the conventional view: neither dysfunction nor other derived medical notions are truly value-free. By contrast, all kinds of pathology ascriptions are necessarily evaluative/prescriptive i.e., they constitute a *negative* evaluation of certain physical or psychological phenomena (e.g., wounds, erratic thought and behavior, restlessness, headaches, etc.) as undesirable, and thus, a call to remedy them.

In Fulford's (1989) conception of medicine then, ethics always comes first. This view construes the relation between the notions of illness (i.e., pathology as experienced by patients and other people) and disease (i.e., lesions or alterations in bodily or mental function) as a "reverse" version of the conventional view: it's because we value certain physical and psychological phenomena negatively that we turn to medicine for answers about what may have caused these and how to remedy them. Although medicine is no doubt science-based, it primarily and irreducibly is a values-based discipline.

Social norms and values are thus always present in medicine; however, these are not always equally *salient*. Fulford assumes that, when the relevant norms and values are widely shared, evaluative language can actually look descriptive. The visibility of values however increases as an inverse function of the level of agreement about them; the least our socionormative backgrounds coincide, the most salient our values will be. According to Fulford, this is what really distinguishes mental from physical medicine, and what presumably explains why both supporters and detractors of psychiatry view mental pathology as somewhat problematic. In the case of physical medicine, our values tend to be widely shared, thus becoming invisible and leading us to view pathology ascriptions as purely descriptive; by contrast, as Thornton (2007) puts it, “psychiatry is concerned with areas of human experience and behavior, such as emotion, desire, volition, and belief, where people’s values are particularly highly diverse” (p. 25). Mental health diagnosis and treatment thus bring front-and-center our differing values about what psychological abilities and experiences may make a life worth living.

Different having-it-both-ways approaches diverge from Fulford’s in the specific details about how values are involved in clinical decision-making, how should we address value disagreements in practice, or how much value diversity should be allowed (Graham, 2010; Thornton, 2007; Varga, 2015). However, they all agree with his diagnosis of Szaszianism and naturalism in mental health. This having-it-both-ways perspective has been widely influential, and I think that it’s essentially right in pointing out (i) that pathology ascriptions, and not just mental pathology ones, typically function as evaluative/regulative devices which express our endorsement or rejection of a person’s biological or dispositional makeup and the social norms governing these assessments; and (ii) that such evaluative element is ineliminable. However, my key point here is that recognizing this much still leaves Szasz’s main challenge unanswered. In particular, it fails to undermine his analysis of the notion of mental illness as a metaphorical, rather than literal one. To see this point more clearly, I will recast both Fulford’s and Szasz’s positions as endorsing some kind of *expressivism* about the notion of mental disorder; their main difference is that while the former focuses on the ‘disorder’ aspect, the latter focuses on the ‘mental’ one.

3 Expressivism and the doubly value-laden nature of mental disorder

Fulford’s position and, to some extent, other having-it-both-ways views that adopt a similar framework (e.g., Thornton, 2007) can be understood as an *expressivist* account of the meaning of pathology ascriptions. Expressivism is a family of positions in the philosophy of language that endorse *non-descriptivism* about language or certain regions of it, i.e., the negative thesis that linguistic expressions considered to express one’s evaluative or more broadly normative attitudes (e.g., moral, epistemic, logical vocabulary, etc.) do not describe or represent states of affairs (Frápolli, 2019, p. 1). Fulford draws primarily from Hare’s (1952) prescriptivist analysis of moral language, which can arguably be understood as a *hybrid* or *ecumenical* kind of expressivism (Eriksson, 2009), or at least a precursor of it (Ridge, 2006). Ecumenical expressivism is an elaboration of classical views such as Ayer’s (1936) or Stevenson’s (1944), which established a sharp difference between descriptive/factual and evaluative claims. For classical expressivists, while descriptive claims express cognitive attitudes (e.g., beliefs) and possess truth-evaluable contents, evaluative ones express non-cognitive, conative attitudes (e.g., desires) which supposedly lack truth-aptness. On this view, whereas sentences like “Madrid’s tap water is rich in minerals” convey information about the world, sentences like “Madrid’s tap water is *good*” don’t; rather, the latter expresses the speaker’s positive attitude towards Madrid’s tap water, it’s meaning amounting to something like “Yay to Madrid’s tap water!”.

Hybrid or ecumenical expressivism rejects this sharp dichotomy between descriptive and evaluative claims; by contrast, it assumes that the meaning of evaluative utterances contains *both* evaluative and descriptive elements, i.e., they express both cognitive and conative attitudes (Ridge, 2006). Hare's prescriptivism adopts a similar approach to the analysis of value terms (Eriksson, 2006; see also Ridge, 2006). On this view, value terms are primarily defined by their evaluative and *prescriptive* force. When we use them in a sentence, their particular contribution to its meaning is that they convey the norms and values that the speaker upholds and prescribes their audience to adopt. However, according to Hare, our value judgements are always grounded on descriptive criteria, i.e., we don't value things in the abstract, but on the grounds of certain factual properties that we find valuable. Thus, by using value terms such as "good", speakers not only convey their approval of what they're talking about (e.g., Madrid's tap water) plus a related prescription (e.g., that absolutely everyone should try it); in addition, they express certain beliefs about its properties that makes it a valuable thing (e.g., being rich in certain minerals). On this view, the speakers' standards about what properties are good-making fix the relation between the evaluative and descriptive meanings of an evaluative utterance.

In Fulford's theory, Hare's key insight is that an evaluative utterance may eventually look like a factual or descriptive one when there is widespread agreement about these standards. In other words, the evaluative *appearance* of an evaluative term or utterance is inversely proportional to the amount of agreement regarding the relevant standards; the less agreement there is about what makes something good, the more overtly value-laden the evaluative utterance will be and vice versa. If for most Madrilenians the minerality of Madrid's tap water makes it the savoriest and most satisfying tap water in Spain, something that all Madrilenians should proudly lecture the rest of the world about, then it comes as no surprise that descriptions of its mineral composition are taken as synonymous with what constitutes *good* tap water -the best kind indeed. "Madrid's tap water is good" is therefore likely to sound like a statement of fact, that is, as a plain description of its rich mineral properties. As explained above, Fulford argues that something similar happens with the case of the value term "illness" when applied to somatic conditions. Arguably, we have widely shared criteria about what bodily states should count as "ill", and somatic medicine is therefore relatively "value simple". Psychiatry, by contrast, is "value complex": we hold widely different values regarding human behavior and experience, and thus it's more overtly value-laden. This is intended to accommodate Szasz's claims about the value-ladenness of psychiatry while, at the same time, rejecting that this poses any threat to our understanding of mental pathology as true, literal pathology.

Fulford's expressivism offers a sound analysis of the meaning of pathology ascriptions, which avoids the pitfalls of naturalists' and Szasz's descriptivist analyses in terms of natural functions. In doing so, it provides a more realistic and nuanced view of the role of ethics in medicine, moving beyond the usual scientistic platitudes. The problem, however, is that it does not alleviate Szasz's main anxieties. In a nutshell, the point is that Szasz's critical approach should be understood as primarily tackling the "mental" -not the "disordered"- aspect of mental disorders.

To begin with, Szasz seems to have already advanced some of Fulford's main arguments about the evaluative nature of pathology terms. Drawing from a distinction between "cognitive or information transmitting" and "promotive" uses of language, Szasz suggests that "the phenomenon of calling someone'sick' *–bodily, physically, mentally, emotionally, or in any other way–* constitutes an excellent example of the promotive use of language" (Szasz, 1961b, p. 59, emphasis added). This admission is clearly at odds with Fulford's and others' construal of his position; and, as a matter of fact, Szasz himself rejected such interpretation. In line with Fulford's objection, Bentall (2004) pointed out that Szasz had arrived to the conclusion that mental illness was a myth "on the basis of a false premise—that there is such a thing as 'real pathology', which can be defined in a manner that lies outside anybody's value system" (p. 315). In response, Szasz (2004) firmly states: "Bentall calls my assertion that mental illness is a myth my *conclusion*. That is an error: it is my *premise*" (p. 321); specifically, a premise in a larger argument denying psychiatry's medical status and scientific legitimacy.

But what is this premise then grounded itself on, if not on the idea that "illness" is a purely descriptive concept? Szasz repeatedly insisted that, in saying that mental illness is a myth, a metaphorical or "fictitious illness", he was asserting "an analytical truth, not subject to empirical falsification" (2011, p. 180); one that he viewed as "similar to asserting that bachelors are not married, or that consecrated bread is not the body of Jesus" (2004, p. 315). Specifically, his criticism was based on a conceptual analysis of mental language. For him, the concept of mental illness was outright *nonsense*; not because diagnoses of mental illnesses are value-laden, but because the very idea of thinking of mental states and processes as pathological in literal terms constituted for him a *category mistake,* similar to speaking literally of someone's 'black' or 'rotten' *soul*, i.e., as if souls could literally have these properties. Roughly, the argument is that it doesn't make sense to characterize minds as 'ill' or 'pathological' in a literal sense, just as it doesn't make sense to talk about 'ill wills', 'broken beliefs', or 'diseased desires' other than metaphorically; to think of these expressions in literal terms, as on a par with descriptions of physical pathologies (e.g., broken bones, liver disease, etc.) is for Szasz to misunderstand how mental language works.

Szasz here seems to defend a non-descriptivist analysis of mind, which takes mental language to pertain to a logical category different from descriptions of worldly events, and more akin to the areas of language that expressivists have most often focused on, i.e., logic, morality, aesthetics, and other kinds of evaluative or expressive discourse. To see this, it is useful to consider his arguments in light of those provided by two early analytic philosophers whom Fulford himself relies upon: Ryle and Wittgenstein.[4] These authors' views of mind and mental language can be understood as endorsing two inter-related commitments: a) that mental language is non-descriptive; and b) that it is intrinsically normative, its meaning being connected to attributions of agency and responsibility.

Firstly, Ryle and Wittgenstein share a non-descriptivist analysis of mind, which has later been taken up by different expressivist accounts (see Frápolli, 2019; Frápolli & Villanueva, 2012). In a nutshell, the idea is that minds are not some kind of substance or *res* -neither natural nor non-natural- because mental expressions (e.g., ascriptions of folk-psychological attitudes) do not describe or represent any state of affairs, i.e.,

[4] This connection between Szasz's and Wittgensteinian views on mind has been recently stressed by Szaszian thinkers (see Moncrieff, 2020).

any particular combination of objects, properties, events, or relations among them. Ryle (1949) expresses this idea in his analysis of the Cartesian theory of mind as involving a category mistake; one resulting from representing "the differences between the physical and the mental (…) inside the common framework of the categories of 'thing', 'stuff', 'attribute', 'state', 'process', 'change', 'cause' and 'effect'" (p. 9). For Ryle, "the phrase 'there occur mental processes' does not mean the same sort of thing as 'there occur physical processes', and, therefore (…) it makes no sense to conjoin or disjoin the two" (pp. 11–12). Wittgenstein's remarks on psychological predicates also suggest a similar view of mind (see Frápolli & Villanueva, 2012). For instance, his non-descriptivism about mental language features in his observations about the lack of "genuine duration" of propositional attitudes, i.e., the fact that the duration of one's belief that Madrid's tap water is extremely savory cannot be measured exactly (e.g., by using a stopwatch), nor is interrupted when there is some 'consciousness breakdown', like when sleeping (1980, §§ 45, 51, 178; 1992, §MS169, p. 9). Since worldly events are taken to be spatially and temporally distributed, this lack of genuine temporal properties hints at the non-descriptive nature of mind.

Most importantly, both Ryle and Wittgenstein stress the contrast between the normative properties of mental state ascriptions and the absence of such normative properties in mere descriptions of a person's behavior or its causes, which they also present as an argument for non-descriptivism about the mind. Ryle thinks that mental ascriptions are used when assessing a person's *reasons* for action (1949, p. 75). Unlike scientific explanations of behavior that describe it in mechanistic or probabilistic terms, the primary function of mental ascriptions is to assess actions and reactions in *normative* terms, e.g., as "intelligent", "voluntary", and so on; for Ryle, the foundational mistake of Cartesianism would precisely lie in viewing these normative assessments as a "paramechanical" subtype of causal explanation (1949, pp. 37–38). Similarly, Wittgenstein (1953) stresses the normative and non-descriptive nature of mental language in his argument against private language (§§185–271; see Heras-Escribano & Pinedo-García, 2018). Thinking of mental language as private entails viewing mental self-ascriptions as mere descriptions of inner facts only knowable to oneself, in a manner that others could never be in a position to sanction or correct them. However, for Wittgenstein, this misses the prescriptive force of mental ascriptions, i.e., the idea that they convey *shared* normative expectations concerning how the agent should act or not in different circumstances. When I claim to believe that Madrid's tap water is the savoriest one, I thereby acquire a series of social commitments to pursue certain courses of action (e.g., choosing it over others if given the opportunity); others are entitled to reprimand me or demand explanations if I deviate from these norms or expectations. This normative force is absent in mere descriptions of behavior or its causes: certain patterns of brain activity might consistently predict my choice of Madrid's tap water over that in other regions; however, when I self-ascribe the aforementioned belief, I am *rationally* expected to choose it over others if given the opportunity.

Ryle's and Wittgenstein's analyses of mind thus convey the idea that, unlike descriptions of a person's biological or behavioral profiles, mental state ascriptions are intrinsically tied to concerns about a person's rationality, agency, and responsibility: rather than describing facts about the person, their main function is to rationalize or justify their actions and reactions, i.e., to *evaluate* them in terms of their conformity to social rules or normative standards,[5] e.g., of rationality, morality, etc. (see also Fernández Castro, 2023; Heras-Escribano & Pinedo-García, 2018; Kalis & Ghijsen, 2022).

Szasz's view of mind is strikingly similar in this regard. Compare Ryle's and Wittgenstein's remarks above with Szasz's suggestion that we should think of "the connection between the psychological and the physical not as a relationship between two different types of occurrences or processes, but as a relationship between two different types of languages" (1961a, p. 78). The key idea here is that, like Ryle and Wittgenstein, Szasz associates the physical, and the factual more broadly, with what is temporally and spatially located, and bound to cause-effect relations. Mental language, by contrast, would not refer to facts -neither biological nor psychological, natural nor supernatural.[6] Szasz too views the essential difference between mental and bodily concepts as radically different in their normative properties. Bodily, biological concepts would play a role in causal-mechanistic, descriptive accounts of human affairs; they would characterize the treatment of human beings as *organisms,* i.e., as complex interplays of causally bounded, spatially, and temporally located biological events and dynamics. Mental language, by contrast, would characterize their analysis in irreducibly normative terms, as rational, autonomous, and responsible *persons*. Crucially, Szasz understood this conceptual distinction as a strict dichotomy: the sole description of human affairs in causal or factual terms -whether physical or psychological- "ought to be recognized as metaphorical rather than literal" (1961a, p. 8).

This is why Fulford's having-it-both-ways approach and similar views fail to come to grips with Szasz's critique of the analogy. In Fulford's view, pathology ascriptions (as all evaluations in Hare's prescriptivism) always convey an evaluation of some *fact* -just like "Madrid's tap water is good" involves a descriptive component concerning the properties of Madrid's tap water that the speaker values positively. Szasz would share this much: even if *all* pathology ascriptions are evaluative, they would necessarily involve an evaluation of *facts* about a human being, i.e., causally bounded, spatially and temporally structured facts. But in mental pathology, there would just be no such fact: mental state ascriptions just don't describe spatial-temporally located properties of an organism whose pathological nature could be assessed. Rather, they constitute themselves *evaluations* of a person's behavior, cognition, and experiences as conforming to socionormative standards of rationality, autonomy, morality, and so on. Mental pathology, understood literally, would thus be an absurdity. As he puts it: "Mind is not matter, hence mental illness is a figure of speech" (2008, III, paragraph 8); just like morality, aesthetics, humor, and other presumably non-descriptive realms of discourse, minds could only be "sick" in metaphorical terms (Szasz, 1961a, p. x).

[5] In a similar vein, contemporary regulative views of mind have recently drawn from this Rylean and Wittgensteinian perspective to argue that folk-psychological interpretation is not primarily about mindreading (i.e., describing and causally explaining one another), but about *mindshaping* (i.e., reciprocally regulating our actions and reactions in norm-conforming ways) (Fernández Castro, 2020; Kalis & Ghijsen, 2022; McGeer, 2007; Zawidzki, 2008).

[6] This analysis of Szasz as a non-descriptivist radically stands against his usual interpretation as a *dualist* about the mind–body relation (e.g., Chapman, 2023b). For Szasz, it was not merely physicalism about the mind that was misguided, but *factualism* more broadly (i.e., the idea that minds are causally bounded entities of *any* sort; see Heras-Escribano & Pinedo-García, 2018).

Thus, even if all pathology ascriptions are evaluative, and even if the norms governing these evaluations partially overlap (e.g., even if there are "family resemblances" between all kinds of pathology ascriptions) mental disorders could not be understood a "disorders like any other". In short: the issue with mental disorders is not that these, unlike somatic ones, are value-laden, but that they are *doubly* value-laden, as they are both 'mental' and 'disordered'.

To better grasp what Szasz's views imply and how exactly Fulford's remarks fail to counteract his main criticisms, it is useful to also consider Szasz's views in the light of expressivism; not about pathology though, but about the mind. In the following section, I illustrate this difference by focusing on the case of delusions, comparing two recent expressivist analyses of delusion ascriptions.

4 Two expressivist analyses of delusion ascriptions

Delusions are typically defined as irrational, bizarre, or somehow *pathological beliefs*[7] (APA, 2022; Coltheart et al., 2011). As I view it, an expressivist analysis of delusions *à la* Fulford's would focus on the 'pathological' or 'wrong' nature of such beliefs. Recently, Wilkinson (2020) has offered an expressivist approach to delusions in a somewhat similar direction. In the face of longstanding worries about the prospect of arriving at any widely shared definition of delusions, Wilkinson adopts a common expressivist strategy: to move from the question "what is X" (a delusion, in this case), to the question "what do we do when we attribute X to someone?". Once we do this, so Wilkinson argues, we are in a better position to realize that the main function of our delusion ascriptions is not to describe a person's (neuro)psychological profile; rather, it is to evaluate it as non-conforming to presumably shared norms governing our *folk epistemology*; i.e., our daily practices of assessing each other's doings in terms of epistemic merit or demerit. His argument draws primarily from similar observations to those that motivate epistemic expressivism (Field, 2009). Roughly, for epistemic expressivists, knowledge attributions do not describe some factual properties or criteria that make of a certain belief an epistemically good one, but rather express the attributor's approval of the attributee's beliefs -more specifically, of the epistemic standards that entitle them to those beliefs. Wilkinson argues that delusion ascriptions play the exact opposite role: to express one's 'folk-epistemological disgust' or disapproval for another person's beliefs.

What is important to note here is that the main focus of this kind of expressivist analysis is on the 'pathological' (or at least 'wrong') character of delusional beliefs.[8] By contrast, an expressivist analysis of delusions *à la* Szasz would not primarily focus on what makes certain beliefs *delusional*, but rather on these *beliefs* themselves (see

[7] Below I comment on the controversies surrounding this definition.

[8] Although he is explicit that "delusion and pathology should not be conceptually tied to one another" (p. 73), Wilkinson considers the possibility that, if we were to adopt a similar non-factualist analysis of terms like 'illness' or 'pathology', delusion and pathology ascriptions could turn out to partially overlap. From this perspective, a delusion would indeed amount to a *pathological* belief "in the simple folk sense that it can't be 'understood', is weird, alien, flies in the face of how human beings ought to be, and needs correcting" (p. 74).

also Núñez de Prado Gordillo, 2022). My point here is that it helps to understand Szasz's remarks on the non-factual character of the mind (and thus on the necessarily metaphorical nature of mental pathology) as a (proto-)expressivist view of the *mental states* whose pathological character is under examination. Such kind of position has been recently offered by Pérez-Navarro et al. (2019), focusing precisely on belief ascriptions.[9] Like Wilkinson (2020), Pérez-Navarro et al. (2019) draw primarily from epistemic expressivism; this time, however, the authors point out that this position is not "expressivist enough". Roughly, their argument is that, while epistemic expressivists are right in pointing out that knowledge ascriptions are non-descriptive and rather express the attributor's approval of the (epistemic standards entitling the) attributtee's beliefs, they still implicitly assume that the belief ascriptions presupposed by these knowledge ascriptions are descriptive; in other words, they still assume that there is a fact of the matter about whether the attributtee actually has the beliefs whose epistemic merit is being assessed.

Wilkinson's (2020) expressivist analysis of delusions as folk-epistemically *bad* beliefs likewise remains neutral about the relevant belief ascriptions. By contrast, Pérez-Navarro et al. (2019) contend that similar reasons to those recommending an expressivist analysis of knowledge ascriptions (or delusion ascriptions for that matter) also motivate an expressivist approach to belief ascriptions. A core motivation of epistemic expressivism is the possibility of *normative disagreement* about knowledge ascriptions, i.e., the possibility that disagreements about whether some agent knows that *p* or not rationally and faultlessly persist despite all participants agreeing about all the relevant facts and despite having made their different epistemic standards explicit. According to Pérez-Navarro et al. (2019), this kind of disagreements can also occur in the case of belief ascriptions.

In fact, as the authors point out, delusions provide a case in point. For more than two decades, there has been a longstanding debate about the doxastic status of delusions -i.e., about whether they should count as beliefs or not (see Bortolotti, 2010). The dividing line lies on the attitude-attitude and attitude-behavior inconsistencies that some people with delusions display, i.e., their failures to reason and act in accordance with their self-professed beliefs. A commonly cited example is the Capgras syndrome, whereby the person claims that another person, typically a loved one, has been replaced by an identically looking impostor. Apparently, many people with Capgras syndrome fail to display the kind of attitudes and behaviors that we would typically expect of someone who believed such content; for instance, they continue to live with the impostor as they did with their abducted loved one, or fail to solve or excuse certain contradictions, such as why the impostor should know certain details about their relationship (Coltheart et al., 2011). For *antidoxasticists*, these attitude-attitude and attitude-behavior inconsistencies preclude an analysis of delusions as beliefs; in other words, they take these inconsistencies as evidence that people with Capgras delusion don't *really* believe that their loved ones have been replaced -they merely *say* so (e.g., Schwitzgebel, 2012). *Doxasticists* disagree. On their view, this kind of inconsistencies are widespread and don't preclude an analysis of other non-clinical instances

[9] For a broader defense of mental expressivism, see Fernández-Castro (2023); for a full-fledged application of this expressivist view to the analysis of belief ascriptions in delusions, see Núñez de Prado-Gordillo (2022).

of irrational cognition and behavior in terms of beliefs; antidoxasticism just hinges on too stringent criteria for belief ascription. Instead, doxasticists stress similar, yet more relaxed criteria (e.g., intelligibility instead of full consistency), and emphasize the agent's first-person authority (Bayne & Pacherie, 2005; Bortolotti, 2010; Clutton, 2018).

What is important to note here is that a) both parties agree about all the relevant facts, i.e., that at least *some* people with certain delusions somewhat fail to conform to social expectations about what follows from believing the relevant contents; b) both parties have made their different standards for belief attribution explicit, i.e., overall consistency vs. intelligibility or first-person authority; yet c) they still (rationally and faultlessly) disagree about whether certain delusions should count or not as beliefs. This disagreement is thus a normative one; one which doesn't just dissolve once each party's standards for belief attribution have been made explicit. Although both concede that the standards endorsed by the opposite party are fairly reasonable ones which guide belief ascription in other contexts, they still disagree about which standards should we employ in *these* cases in particular. What Pérez-Navarro et al. (2019) make of this case is that belief ascriptions are best seen as evaluative, not descriptive devices (see also Fernández-Castro, 2023); there's just no fact of the matter as to which ascription is correct, for what is at stake is which doxastic policies should we endorse.[10]

Although differing in scope, it is important to note that both Wilkinson's (2020) and Pérez-Navarro et al.'s (2019) expressivist analyses are fully compatible. If we combine both approaches, the resulting view of delusion attributions is that these are *doubly evaluative or value-laden*. By means of delusion attributions, we evaluate both a) a person's thoughts, actions, and experiences *as instances of believing* a certain content; and b) such beliefs as 'pathological' -or, at least, 'folk-epistemologically bizarre'.

Now we can see more clearly why Fulford's claim that *all* disorders are value-laden fails to come to grips with Szasz's contention that mental disorders, unlike somatic ones, primarily involve deviations from psychosocial and ethical norms of conduct. As I view it, while having-it-both-ways analyses tackle the evaluative nature of pathology ascriptions, Szasz's criticism is better understood as tackling the evaluative, non-descriptive nature of the mental. From this perspective, only facts about a person's bodily makeup would be literally evaluable as pathological; minds (e.g., beliefs, intentions, desires, etc.), not being spatial-temporally and causally bounded entities, would only be *metaphorically* describable in such medical terms. Even if pathology ascriptions are in both cases evaluative, *what is under evaluation* is radically different; in the case of the mental, there are no facts at all being evaluated.

The persuasiveness of this argument obviously hinges on whether one adopts this view of mind. However, although I personally find parts of it compelling, my point here is not to argue for it. Rather, my point is that analyzing mental disorder ascriptions as doubly value-laden sheds light on the enduring legacy of Szaszian-like psychiatric skepticism. This skepticism doesn't solely arise from conceptual concerns, to be sure;

10 Importantly, this doesn't necessarily mean that belief ascriptions (or mental/mental disorder ascriptions for that matter) lack truth conditions. Doxastic expressivism is compatible with contemporary expressivist views that reject classical and hybrid expressivisms in their assumption that non-descriptive statements are not truth-apt. On these views, non-descriptive statements can be as truth-apt as descriptive ones (e.g., Price et al., 2013).

it is further fueled by sound concerns about the efficacy and validity of mainstream psychiatric approaches, by the vested interests of alternative professional groups (e.g., clinical psychologists), as well as from other practical reasons. Nonetheless, I do think that one conceptual reason for the persistence of Szaszianism is that it taps into a genuine worry about the legitimacy of the analogy between mental and physical conditions; a concern which persists even when we acknowledge the value-laden nature of *all* disorder ascriptions.

Now Szaszianism goes far beyond this, claiming that it then makes no sense whatsoever to speak of mental pathology. In fact, Szasz took it a step further, suggesting that the entire structure of scientific psychiatry, built upon the 'foundational myth' of the analogy, constitutes an illegitimate and intrinsically corrupt endeavor; its object of study amounting to medicalized 'malingering', its methods to 'psychiatric slavery' and 'rape'. These extreme claims were rooted in his libertarian understanding of mind and agency (Chapman, 2023, 2023b; Frazer-Carroll, 2023), according to which *any* possible causal account of human affairs amounted to a mere metaphor, albeit not an innocent one: rather, one that reflected and reinforced totalitarian ideologies aimed at curtailing individual freedom. The analogy between mental and somatic disorders epitomized this illegitimate metaphor. It was not only a way of disguising political coercion as science, but a form of oppression in and by itself[11]; one which obliterated individual freedom and responsibility by the mere act of describing human affairs in the fashion of the natural sciences, and which he viewed as lying at the core of coercive "socialist-statist" practices and institutions conforming the "Therapeutic state" (Szasz, 2001, 2009). Contrastingly, Szasz's underlying vision for psychiatry was "to reintroduce freedom, choice, and responsibility into (its) conceptual framework and vocabulary" (p. 6), the paramount of "consensual psychiatry" being "a buyer–seller relationship between putative equals" (2009, p. 26). Contemporary critical demands to "drop the language of disorder" and re-emphasize "meaning" and "agency" in mental health (Johnstone & Boyle, 2018; Kinderman et al., 2013; Moncrieff, 2020), although far from his radical libertarianism, partially echo these Szaszian claims.

Despite their liberatory ambitions, these claims may also be damaging for those who are supposed to be liberated, as activists have long denounced (Sedgwick, 1982). It is therefore essential to scrutinize Szasz's claims to discern the actual implications of his critique on the analogy between mental and somatic disorders, distinguishing what logically follows from his argument and what does not. In the next and final section, I take issue with certain conclusions that Szasz draws from his rejection of the analogy, and I claim that none necessarily follows from the analysis of mental disorders as doubly value-laden. Although a detailed exploration of this issue lies beyond the scope of this paper, I illustrate my point by drawing from recent socio-normativist approaches to mental health and contemporary mad/neurodiversity activist concerns with the perils of undue normalizing discourses.

[11] I would like to thank Virginia Ballesteros for her helpful comments on this point.

5 Social boundaries and medical language

Szasz considered his critique of the analogy between mental and somatic disorders as the core premise in a larger argument against the medicalization of suffering and the legitimacy of diverse mental healthcare institutions. Here I want to dispute two core claims that are often backed up by Szasz's premise: a) that mental disorders just amount to medicalized social deviance, and therefore no real distinction can be drawn between the two; and b) that medicalizing psychological suffering necessarily amounts to some sort of *wrong-doing* for people from mental health-related collectives. I will contend that neither of these claims follow from the rejection of the analogy between mental and somatic disorders.

5.1 Self-regulation and self-understanding as socio-normative boundaries

Szasz's criticism of the analogy between mental and somatic disorders is usually held in support of the idea that no real boundary can ever be drawn between mental disorder and social deviance -an example of what some refer to as the boundary problem (Rashed, 2021). This idea comes in different flavors. A milder version of it amounts to the *continuity thesis*, i.e., the idea that clinical and non-clinical cases stand on a continuum, with no clear boundaries between the two (Bentall, 2003). A more radical understanding of it views psychiatric diagnoses as mere tools for social control, a means to enforce conformity to prevailing social norms and expectations (Szasz, 1961a; see also Curtis et al., 2000). To be sure, both versions of this idea hold true on many occasions. Psychopathology often involves mere quantitative deviations from similar, non-clinical forms of behavior; the widespread occurrence of psychotic experiences in non-clinical populations provides a case in point (Johns & van Os, 2001). Conversion therapies, on the other hand, grimly exemplify how psychiatric and psychological assessment and treatment methods have indeed been used as methods of social control.

What I take issue with here is that these claims *necessarily* follow from the analysis of mental disorders as doubly value-laden and the concomitant rejection of the analogy between mental and somatic disorders. To claim that no boundaries can *ever* be drawn between social deviance and psychopathology amounts to claiming, as Szasz did, that the presence of psychological distress can never warrant the attribution of the 'sick role' to a person, and that no meaningful distinction can be made between a person's actions and reactions which express who they *really* are and those that are better seen as psychopathological. There are important political, ethical, and even clinical motivations behind this distinction. For instance, it allows for a more caring and nuanced approach to responsibility assessment in scenarios of wrongdoing by people in situations of mental distress (Brandenburg, 2018). Moreover, addressing the so-called *self-illness ambiguity* (see Dings & Glas, 2020) is widely considered to be a crucial step towards recovery, since it allows for the clarification of the person's values and the subsequent identification of idiosyncratic therapeutic goals that fit them. For Szasz, his rejection of the analogy between mental and somatic disorders meant the rejection of these useful distinctions as chimeras at the service of psychiatric coercion and state control.

As I view it, none of this follows from the analysis of mental disorders as doubly value-laden. What follows is just that the notions of mental and physical health are conceptually distinct, i.e., that mental disorders, if *genuinely* mental, are not ‘just like any other disorder’ and that *both* psychopathology and social deviancy involve non-conformity to social norms and expectations. But this is still compatible with defending that there are criteria -irreducibly social ones, for sure- guiding this distinction.

Telling psychopathology and social deviancy apart surely is not a straightforward task; as many have argued, what (and whose) norms and values are relevant to determine what counts as pathological and what as merely deviant probably varies across types of mental conditions and particular cases, and hence no general maxims nor principles will allow us to ascertain, for all and every possible case, what merits a psychopathology ascription (Thornton, 2007). But contemporary approaches to the boundary problem offer some interesting candidates as potential rules of thumb to guide these distinctions. A primary candidate, which lies at the core of contemporary psychotherapeutic approaches, is that mental disorders primarily involve alterations in the person’s self-regulatory abilities (de Haan, 2020; Hayes et al., 2006; Leder, 2019; Leder & Zawidzki, 2023); that is, in their ability to express themselves and act in accordance with their *own* norms and values. Far from a subjectivist understanding of what a person’s “own” norms amount to, recent socio-normative views have set to understand these as inherently social, as embedded in the person’s social and cultural niche. What distinguishes psychopathology from social deviancy then is that the person systematically fails to act in accordance with shared norms and values that they *actually* endorse, vs. those that other people may try to force on them.

Recent enactivist views (de Haan, 2020; Nielsen, 2023) illustrate this point. According to enactivists, the hallmark of psychopathology is precisely its “self-defeating dynamic”, whereby the person displays systematic, ’loopy,’ or ’sticky’ tendencies that run contrary to their own norms and values, thus being experienced as somewhat alien to the self. De Haan’s view is particularly interesting in that she explicitly rejects the analogy between mental and somatic disorders. She describes psychiatric disorders as “structurally disordered patterns of sense-making” (2020, pp. 209–210), that is, as systematic alterations in a person’s ability to flexibly make sense of themselves and of different circumstances. Psychiatric disorders thus differ from physical ones in that while the latter might have ‘secondary effects’ on sense-making, the former are primarily characterized by such alterations; thus, they “dissolve if one succeeds in changing one’s way of interacting with the world” (p. 11). Specifically, she stresses the centrality of *existential* (vs. natural) values in psychopathology, i.e., those involving what it is for the individual to live a *good* life, not just any kind of life. These, for De Haan, aren’t chosen at will, but are “relational realities” stemming from the person’s socio-cultural world and their interaction with it. Psychopathology can thus be differentiated from social deviance in that it primarily involves a systematic departure from the person’s existential values, grounded in their particular history of interactions with their social niches.

Enactive psychiatry thus offers one way to draw the line between psychopathology and social deviance in irreducibly socio-normative terms. A somewhat complementary view has been recently defended by Rashed (2021; Aftab & Rashed, 2021), who emphasizes the role of one's identity and self-understanding in telling apart disorder from social deviance. For the author, debates on the boundary problem are often framed by a misplaced emphasis on the concept of disorder and what should lie *within* its scope; Rashed's alternative strategy is to shift focus to what should lie *without* it. In particular, he focuses on concepts central to mad activism such as the notion of social recognition and its relation to social and personal identity, arguing that "clarifying the boundary problem is not to be achieved by getting a handle on the definition and limits of the concept of mental disorder, but on understanding the addressees and normative limits of recognition" (2021, pp. 298–299). Among these normative limits, Rashed stresses the importance of the consistency and temporal stability of the person's self-conception. What falls outside the scope of recognition, and *may* therefore be understood as disorder, are the person's actions and experiences that are not sufficiently unitary to conform a recognizable identity. Crucially, Rashed views the limits of recognizability as irreducibly socionormative, grounded on our social customs and practices, and thereby open to contestation and modification.

What de Haan's enactive view and Rashed's recognition-based approach illustrate is that there are ways to spell out the differences between psychopathology and social deviance in a manner that honors the deeply socionormative character of this distinction. Like Szaszian critical psychiatry, these views also emphasize the intimate connection between notions of psychopathology and normative issues regarding the person's agency, identity, and meaning-making abilities; unlike Szaszianism, however, this is not taken to rule out the possibility of using "the language of disorder" to make useful distinctions.

5.2 The up and downsides of normalization

Warnings against the use of "the language of disorder" to address psychological distress often respond to legitimate concerns about the often-overstated efficacy of pharmaceutical treatments, how the mechanisms behind this efficacy are often depicted in public discourse, or the often-underplayed risks of secondary effects associated with them. But, above all, the Szaszian "anti-medical" approach is primarily based on the assumption that, since mental disorders do not literally exist, a medical framing of mental distress is straightforwardly unjustified and unavoidably leads to its undue pathologizing. "Normalizing" strategies are instead recommended.

My point here is that neither medicalizing attitudes are senseless or necessarily harmful once we assume that mental disorders are not literal disorders, nor "normalizing" attitudes are always the best way forward; in fact, they can be just as unjust and harmful (Carel, 2023; Chapman, 2023, 2023b). Firstly, the medical understanding of mental distress plays several key functions in current social arrangements. Other than allowing for the useful distinctions pointed out above, diagnostic practices also provide a rationale for granting access to welfare programs and key social and material resources (e.g., social benefits, sick leaves, etc.). As matters stand now, it is not so clear what alternative rationale critical approaches could offer. In this sense, although there is much to praise in the emphasis on meaning, responsibility, and agency shared by many critical views, and although this emphasis answers to legitimate concerns regarding abusive psychiatric practices, it risks undermining the very social and material conditions that are key to the survival and improvement of many. The "recovery movement" (Anthony, 1993), born in the 1990s as the coalescence of antipsychiatry views and the demands of the user/survivor movement, provides a case in point: as mad and neurodiversity activists have long warned (Sedgwick, 1982), the emphasis on agency, responsibility, and self-management can easily be -and in fact has been- co-opted by neoliberal austerity politics, proving to be a key factor in the ongoing dismantling of mental health services (Chapman, 2023, 2023b; Frazer-Carroll, 2023; Thomas, 2016; Woods et al., 2022).

Another major downside of Szaszianism's uncompromised commitment to demedicalization is that it downplays the hermeneutical function of psychiatric diagnoses, i.e., their role in fostering a sense of personal and collective self-understanding. Crucially, this is something that many contemporary mad and neurodiversity activists find valuable, and one of the reasons why many of these diagnostic labels have been systematically reclaimed and politicized in their struggle for recognition (Chapman, 2020b, 2023; Chapman & Carel, 2022; Curtis et al., 2000; Frazer-Carroll, 2023; Wardrope, 2015). In this sense, the phenomenon of epistemic injustice -i.e., the unfair discrimination against someone in their testimonial and hermeneutical abilities based on prejudices about their social identity (Fricker, 2007)- also illustrates why both medicalization and normalization can be equally harmful. Much research on epistemic injustice in psychiatric contexts has targeted how it may be promoted by medical understandings of the person's experiences. Privileging medical perspectives would function as a contributory factor in epistemic injustice, for instance by reinforcing prejudices about the disorder necessarily impairing their judgement and communicative abilities (Chapman & Carel, 2022; Kidd et al., 2022). However, this epistemic wrongdoing may also be inflicted by de-medicalization discourses, which often portray neurodiversity and mad advocates as "naïve victims" that have been tricked by the lies of medical institutions and pharmaceutical companies into believing that psychiatric diagnoses are something other than a plain myth (Chapman, 2023; Wardrope, 2015). By doing so, they straightforwardly neglect mental health collectives' own longstanding epistemic practices of reappropriation and repoliticizing of traditional medical understandings. Furthermore, this neglection, based itself on the unquestioned premise that normalizing attitudes are intrinsically good, may contribute to reinforce the very ableist and sanist norms and structures at the root of epistemic injustice (Chapman, 2023, 2023b).

In sum, both the co-option of recovery programs by austerity politics and the phenomenon of epistemic injustice illustrate why Szasz's recommended de-medicalizing and normalizing strategies aren't necessarily conducive to their self-purported liberatory aims. There may be conceptual reasons to reject the analogy between mental and somatic disorders; however, it doesn't follow that we must then necessarily reject the medical framing of psychiatric conditions, nor that normalizing attitudes will always be more just. Further considerations about the role of medicalization vs. normalization discourses in the broader social and political contexts in which they are embedded are needed to reach that conclusion.

6 Conclusion

Recent years have witnessed a notable increase in the popularity of Szaszian arguments against the medical model of mental distress. Szasz's main point takes issue with the analogy between mental and somatic disorders, i.e., the idea that mental disorders are just "like any other disorder". He viewed this analogy as a 'myth': mental disorders cannot be literal disorders for they primarily involve deviations from psychosocial and ethical norms of conduct, not functional–anatomical anomalies in biological functioning. Usual responses to Szasz's challenge focus on his Virchowian, naturalistic notion of disorder. Here I've mainly focused on having-it-both-ways views, which argue that *both* mental and somatic disorder ascriptions involve social norms and values, and therefore this cannot be a criterion to reject the analogy. The problem with this counterargument is that it fails to come to grips with Szasz's main contention point: that mental disorders can only be metaphorical, for mental ascriptions are primarily evaluative, not descriptive devices. To illustrate this point, I've argued that both Fulford and Szasz seem to endorse some sort of expressivist view of mental disorders; however, the key difference is that while the former targets the 'disorder' aspect, the latter targets the 'mental' one. I have exemplified this difference by contrasting Wilkinson's (2020) and Pérez-Navarro et al.'s (2019) recent expressivist views of delusions. The conclusion is that what Szasz's challenge points to is the *doubly* value-laden nature of mental disorders, i.e., that not only disorder ascriptions are value-laden, but also mental ones in general.

However, recognizing this does not necessarily entail other Szaszian claims about the problems of medicalization; these only make sense from his underlying libertarian view of mental health and human affairs more broadly. The view of mental disorders as doubly value-laden, in itself, doesn't entail Szasz's view of patients as "malingers", mental healthcare institutions as perverse instruments of social engineering, or mental health policy-making as the scheming of the "statist-therapeutic state". I've targeted two key inter-related Szaszian claims: a) that, once we reject the analogy, there's no principled way to distinguish psychopathology from social deviance; and b) that the medicalizing of mental distress necessarily amounts to wrong-doing, normalizing attitudes being always preferrable. On the one hand, recent enactive and recognition-based views provide several ways to spell out the difference between psychopathology and social deviance that stress the irreducibly socio-normative nature of this distinction; on the other hand, as mad and neurodiversity activists have long denounced, *both* medicalizing and normalizing attitudes can be harmful when the perspectives and needs of neurodivergent, mad, and other relevant collectives are not taken into account. Although there might be conceptual reasons to support Szasz's rejection of the analogy, there are also important political, ethical, and clinical reasons to maintain the distinction between psychopathology and social deviance, as well as to preserve medical language in certain cases. Viewing mental disorders as inherently value-laden does not automatically blur these valuable distinctions.

Acknowledgements This work was supported by the Dutch Research Council (NWO) through a postdoctoral fellowship associated to the VIDI project "Shaping Our Action Space: A Situated Perspective on Self-Control" (VI.Vidi.195.116). It also received partial support from the Spanish Ministry of Science and Innovation through the projects "The Social Roots of Mental Health" (PID2021-126826NA-I00), and

"Functional Study of the Clinical Interaction in Patients Diagnosed with Mental Illness" (PSI2016-76551-R), where the author participates as a team member, as well as from the Autonomous University of Madrid through an FPI-UAM Predoctoral Fellowship (2017). I would also like to thank Virginia Ballesteros and Víctor Fernández-Castro for their helpful advice and comments.

Declarations

Competing interests None.

References

Aftab, A., & Rashed, M. A. (2021). Mental disorder and social deviance. *International Review of Psychiatry, 33*(5), 478–485. https://doi.org/10.1080/09540261.2020.1815666

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders: DSM-5™, 5th ed* (pp. xliv, 947). American Psychiatric Publishing, Inc.

American Psychiatric Association. (2022). *Diagnostic and Statistical Manual of Mental Disorders: DSM-5-TR™, revised 5th ed*. American Psychiatric Publishing, Inc.

Anthony, W. A. (1993). Recovery from mental illness: The guiding vision of the mental health service system in the 1990s. *Psychosocial Rehabilitation Journal, 16*(4), 11–23. https://doi.org/10.1037/h0095655

Ayer, A. J. (1936). *Language*. Courier Corporation.

Bayne, T., & Pacherie, E. (2005). In defence of the doxastic conception of delusions. *Mind and Language, 20*(2), 163–188. https://doi.org/10.1111/j.0268-1064.2005.00281.x

Bentall, R. P. (2003). *Madness explained: Psychosis and human nature* (1st ed.). Penguin.

Boorse, C. (1975). On the distinction between disease and illness. *Philosophy & Public Affairs, 5*(1), 49–68.

Boorse, C. (1976a). What a theory of mental health should be. *Journal for the Theory of Social Behaviour, 6*(1), 61–84. https://doi.org/10.1111/j.1468-5914.1976.tb00359.x

Boorse, C. (1976b). Wright on functions. *The Philosophical Review, 85*(1), 70–86. https://doi.org/10.2307/2184255

Boorse, C. (2014). A second rebuttal on health. *Journal of Medicine and Philosophy, 39*(6), 683–724. https://doi.org/10.1093/jmp/jhu035

Bortolotti, L. (2010). *Delusions and other irrational beliefs*. Oxford University Press.

Brandenburg, D. (2018). The nurturing stance: Making sense of responsibility without blame: The nurturing stance. *Pacific Philosophical Quarterly, 99*, 5–22. https://doi.org/10.1111/papq.12210

Carel, H. (2023). Vulnerabilization and de-pathologization: Two philosophical suggestions. *Philosophy, Psychiatry, & Psychology, 30*(1), 73–76. https://doi.org/10.1353/ppp.2023.0013

Chapman, R. (2020a). *Neurodiversity, disability, wellbeing*. Routledge.

Chapman, R. (2020b). The reality of autism: On the metaphysics of disorder and diversity. *Philosophical Psychology, 33*(6), 799–819. https://doi.org/10.1080/09515089.2020.1751103

Chapman, R. (2021). Neurodiversity and the social ecology of mental functions. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science, 16*(6), 1360–1372. https://doi.org/10.1177/1745691620959833

Chapman, R. (2023a). A critique of critical psychiatry. *Philosophy, Psychiatry, & Psychology, 30*(2), 103–119.

Chapman, R. (2023b). *Empire of Normality: Neurodiversity and Capitalism*. Pluto Press.

Chapman, R., & Carel, H. (2022). Neurodiversity, epistemic injustice, and the good human life. *Journal of Social Philosophy*. https://doi.org/10.1111/josp.12456

Clutton, P. (2018). A new defence of doxasticism about delusions: The cognitive phenomenological defence. *Mind & Language, 33*(2), 198–217.

Coltheart, M., Langdon, R., & McKay, R. (2011). Delusional Belief. *Annual Review of Psychology, 62*(1), 271–298. https://doi.org/10.1146/annurev.psych.121208.131622

Cooper, R. (2002). Disease. *Studies in History and Philosophy of Science Part c: Studies in History and Philosophy of Biological and Biomedical Sciences, 33*(2), 263–282. https://doi.org/10.1016/S0039-3681(02)00018-3

Curtis, T., Dellar, R., Leslie, E., & Watson, B. (Eds.). (2000). *Mad pride: A celebration of mad culture*. Spare Change Books.

de Haan, S. (2020). *Enactive psychiatry* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781108685214

Dings, R., & Glas, G. (2020). Self-management in psychiatry as reducing self-illness ambiguity. *Philosophy, Psychiatry, & Psychology, 27*(4), 333–347. https://doi.org/10.1353/ppp.2020.0043

Engel, G. L. (1977). The need for a new medical model: A challenge for biomedicine. *Science, 196*(4286), 129–136.

Eriksson, J. (2009). Homage to hare: Ecumenism and the Frege-Geach problem. *Ethics, 120*(1), 8–35. https://doi.org/10.1086/606161

Fernández Castro, V. (2020). Regulation, normativity and folk psychology. *Topoi, 39*(1), 57–67. https://doi.org/10.1007/s11245-017-9511-7

Fernández Castro, V. (2023). An expressivist approach to folk psychological ascriptions. *Philosophical Explorations, 0*(0), 1–20. https://doi.org/10.1080/13869795.2023.2251491

Frápolli, M. J. (2019). *Expressivisms*. Cambridge University Press.

Frápolli, M. J., & Villanueva, N. (2012). Minimal expressivism. *Dialectica, 66*(4), 471–487. https://doi.org/10.1111/1746-8361.12000

Frazer-Carroll, M. (2023). *Mad world: The politics of mental health*. Pluto Press.

Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press.

Fulford, K. W. M. (1989). *Moral theory and medical practice*. Cambridge University Press.

Fulford, K. W. M. (1999). Nine variations and a coda on the theme of an evolutionary definition of dysfunction. *Journal of Abnormal Psychology, 108*, 412–420. https://doi.org/10.1037/0021-843X.108.3.412

Fulford, K. W. M., & Staden, W. V. (2013). Values-based practice: Topsy-turvy take-home messages from ordinary language philosophy (and a few next steps). In K. W. M. Fulford (Ed.), *The Oxford handbook of philosophy and psychiatry.* Oxford University Press.

Graham, G. (2010). *The disordered mind: An introduction to philosophy of mind and mental illness*. Routledge.

Hare, R. M. (1952). *The language of morals*. Oxford University Press.

Hayes, S. C., Luoma, J. B., Bond, F. W., Masuda, A., & Lillis, J. (2006). Acceptance and commitment therapy: Model, processes and outcomes. *Behaviour Research and Therapy, 44*(1), 1–25. https://doi.org/10.1016/j.brat.2005.06.006

Heras-Escribano, M., & Pinedo-García, M. (2018). Naturalism, non-factualism, and normative situated behaviour. *South African Journal of Philosophy, 37*(1), 80–98. https://doi.org/10.1080/02580136.2017.1422633

Johns, L. C., & van Os, J. (2001). The continuity of psychotic experiences in the general population. *Clinical Psychology Review, 21*(8), 1125–1141. https://doi.org/10.1016/S0272-7358(01)00103-9

Johnstone, L., & Boyle, M. (2018). The power threat meaning framework: An alternative nondiagnostic conceptual system. *Journal of Humanistic Psychology*. https://doi.org/10.1177/0022167818793289

Kalis, A., & Ghijsen, H. (2022). Understanding implicit bias: A case for regulative dispositionalism. *Philosophical Psychology, 35*(8), 1212–1233. https://doi.org/10.1080/09515089.2022.2046261

Kendell, R. E. (1975). The concept of disease and its implications for psychiatry. *British Journal of Psychiatry, 127*(4), 305–315. https://doi.org/10.1192/bjp.127.4.305

Kidd, I. J., Spencer, L., & Carel, H. (2022). Epistemic injustice in psychiatric research and practice. *Philosophical Psychology, 0*(0), 1–29. https://doi.org/10.1080/09515089.2022.2156333

Kinderman, P., Read, J., Moncrieff, J., & Bentall, R. P. (2013). Drop the language of disorder. *Evidence Based Mental Health, 16*(1), 2–3. https://doi.org/10.1136/eb-2012-100987

Kingma, E. (2007). What is it to be healthy? *Analysis, 67*(294), 128–133. https://doi.org/10.1093/analys/67.2.128
Kingma, E. (2013). Naturalist accounts of mental disorder. In K. W. M. Fulford (Ed.), *The Oxford handbook of philosophy and psychiatry* (p. 363). Oxford University Press.
Leder, G. (2019). What does it mean to have a meaning problem? Meaning, skill, and the mechanisms of change in psychotherapy. *Philosophy, Psychiatry, & Psychology, 26*(3), 35–50. https://doi.org/10.1353/ppp.2019.0027
Leder, G., & Zawidzki, T. (2023). The skill of mental health: Towards a new theory of mental health and disorder. *Philosophy and the Mind Sciences*. https://doi.org/10.33735/phimisci.2023.9684
McGeer, V. (2007). The regulative dimension of folk psychology. In D. D. Hutto & M. Ratcliffe (Eds.), *Folk psychology re-assessed* (pp. 137–156). Springer. https://doi.org/10.1007/978-1-4020-5558-4_8
Middleton, H., & Moncrieff, J. (2019). Critical psychiatry: A brief overview. *Bjpsych Advances, 25*(1), 47–54. https://doi.org/10.1192/bja.2018.38
Moncrieff, J. (2020). "It Was the Brain Tumor That Done It!": Szasz and Wittgenstein on the importance of distinguishing disease from behavior and implications for the nature of mental disorder. *Philosophy, Psychiatry, & Psychology, 27*(2), 169–181. https://doi.org/10.1353/ppp.2020.0017
Morgan, A. (2023). Power, threat, meaning framework: A philosophical critique. *Philosophy, Psychiatry, & Psychology, 30*(1), 53–67. https://doi.org/10.1353/ppp.2023.0011
Murphy, D. (2013). The medical model and the philosophy of science. In *The Oxford handbook of philosophy and psychiatry* (pp. 966–986). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199579563.001.0001
Nielsen, K. (2023). Embodied, embedded, and enactive psychopathology: Reimagining mental disorder. *Springer International Publishing*. https://doi.org/10.1007/978-3-031-29164-7
Núñez de Prado Gordillo, M. (2022). *Mental health without mirrors. A non-descriptivist approach to mental health and the intervention with people with delusions* [Doctoral Thesis]. https://repositorio.uam.es/handle/10486/704791
Pérez-Navarro, E., Fernández Castro, V., de Prado, González, Salas, J., & Heras-Escribano, M. (2019). Not expressivist enough: Normative disagreement about belief attribution. *Research Philosophica*. https://doi.org/10.11612/resphil.1794
Pickard, H. (2009). Mental illness is indeed a myth. In M. Broome & L. Bortolotti (Eds.), *Psychiatry as cognitive neuroscience.* Oxford University Press.
Price, H., Blackburn, S., Brandom, R., Horwich, P., & Williams, M. (2013). *Expressivism*. Cambridge University Press.
Rashed, M. A. (2019). *Madness and the Demand for Recognition: A Philosophical Inquiry into Identity and Mental Health Activism*. Oxford University Press.
Rashed, M. A. (2021). An approach to the boundary problem: Mental health activism and the limits of recognition. *Philosophy, Psychiatry, & Psychology, 28*(4), 297–313. https://doi.org/10.1353/ppp.2021.0047
Read, J., & Moncrieff, J. (2022). Depression: Why drugs and electricity are not the answer. *Psychological Medicine, 52*(8), 1401–1410. https://doi.org/10.1017/S0033291721005031
Ridge, M. (2006). Ecumenical expressivism: Finessing frege. *Ethics, 116*(2), 302–336. https://doi.org/10.1086/498462
Ryle, G. (1949). *The concept of mind*. Routledge.
Schaler, J. A. (2004). *Szasz Under Fire: A Psychiatric Abolitionist Faces His Critics*. Open Court.
Schwitzgebel, E. (2012). Mad belief? *Neuroethics, 5*(1), 13–17. https://doi.org/10.1007/s12152-011-9127-3
Sedgwick, P. (1982). *Psycho politics*. Unkant Publishers.
Stevenson, C. L. (1944). *Ethics and language*. Yale University Press.
Szasz, T. S. (1960). The myth of mental illness. *American Psychologist, 15*, 113–118. https://doi.org/10.1037/h0046535
Szasz, T. S. (1961a). *The Myth of Mental Illness: Foundations of a Theory of Personal Conduct* (Rev. ed., reprinted, 27. [print.]). Perennial.
Szasz, T. S. (1961b). The uses of naming and the origin of the myth of mental illness. *American Psychologist, 16*(2), 59–65. https://doi.org/10.1037/h0040842
Szasz, T. S. (1977). *Psychiatric slavery*. Syracuse University Press.
Szasz, T. S. (2000). Second commentary on" Aristotle's function argument". *Philosophy, Psychiatry, & Psychology., 7*(1), 3–16.
Szasz, T. S. (2001). *Pharmacracy: Medicine and Politics in America*. Praeger.

Szasz, T. S. (2008). *Psychiatry: The Science of Lies*. Syracuse University Press.
Szasz, T. S. (2009). *Antipsychiatry: Quackery Squared*. Syracuse University Press.
Szasz, T. S. (2011). The myth of mental illness: 50 years later. *The Psychiatrist, 35*(5), 179–182. https://doi.org/10.1192/pb.bp.110.031310
Thomas, P. (2016). Psycho politics, neoliberal governmentality and austerity. *Self & Society, 44*(4), 382–393. https://doi.org/10.1080/03060497.2016.1192905
Thornton, T. (2007). *Essential philosophy of psychiatry*. Oxford University Press.
Varga, S. (2015). *Naturalism, interpretation, and mental disorder*. Oxford University Press.
Wakefield, J. C. (1992). The concept of mental disorder. *American Psychologist, 16*, 373.
Wakefield, J. C. (2007). The concept of mental disorder: Diagnostic implications of the harmful dysfunction analysis. *World Psychiatry, 6*(3), 149.
Wardrope, A. (2015). Medicalization and epistemic injustice. *Medicine, Health Care, and Philosophy, 18*(3), 341–352. https://doi.org/10.1007/s11019-014-9608-3
Wittgenstein, L. (1953). *Philosophical Investigations*. Wiley-Blackwell.
Wittgenstein, L. (1980). *Remarks on the Philosophy of Psychology, Volume 2* (C. J. Luckhardt & M. A. E. Aue, Trans.). University of Chicago Press.
Wittgenstein, L. (1992). *Last Writings on the Philosophy of Psychology* (Vol. 2). Wiley-Blackwell.
Woods, A., Hart, A., & Spandler, H. (2022). The Recovery Narrative: Politics and Possibilities of a Genre. *Culture, Medicine, and Psychiatry, 46*(2), 221–247. https://doi.org/10.1007/s11013-019-09623-y
Zawidzki, T. W. (2008). The function of folk psychology: Mind reading or mind shaping? *Philosophical Explorations, 11*(3), 193–210. https://doi.org/10.1080/13869790802239235

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.