The Language of Mental Illness

Renée Jorgensen Bolinger[1]
bolinger@princeton.edu

To appear in *The Routledge Handbook of Social & Political Philosophy of Language*, edited by Justin Khoo and Rachel Sterken. Please refer to the published version for citation purposes.

1 Introduction

Battles over how to use our language are fought over all kinds of terms, but often in scattered skirmishes across a wide range of social platforms and contexts, and so are difficult to trace. But since the *Diagnostic and Statistical Manual of Mental Disorders* (DSM) is used to standardize the meanings of psychiatric terms, it acts as a sort of lightning rod concentrating the disputes over the language of mental illness, making it easier to see the relationships between issues, the interests of varying stakeholders, and background beliefs about the nature of underlying phenomena. This is one reason why the language in and around psychiatry is particularly rich terrain for political philosophy of language. It also resists a sort of naïve-realist first response—that the way to settle these disputes is simply to *find out* what these terms 'really' mean, and then use them that way. The language of mental illness is pervasive, and everywhere political.

There are two primary clusters of politically charged questions about the language of mental illness, which I'll explore in this chapter. The first concerns their straightforwardly descriptive meanings in diagnostic contexts. I devote §2 to exploring these, ultimately contending (though this should not be taken as the orthodox or consensus view) that disputes over these literal meanings must ultimately be considered a problem in conceptual ethics, to be settled by sorting out which are the most important *functions* for these terms. The second cluster of questions center on metaphoric or slang uses of terms that invoke mental illness ('crazy', 'bonkers', etc.), when they are applied pejoratively to people who do not have mental health

[1] Draft updated January 2020. I am indebted to several people for illuminating discussions on these topics; my particular thanks to Colin Klein and Peter Clutton for written comments that greatly improved this paper, as well to August Gorman, Cassie Herbert, Matthew McKeever, Elyn Saks, Rachel Sterken, and Gary Watson. Thanks also to Stephanie Collins, Nicola Kemp, and Eric Wiland, and audiences at the New York Philosophy of Language Workshop, USC Speculative Society, and the 2017 meeting of the Central APA for helpful discussion of much of this material.

conditions. §3 considers whether we should think of these uses as *slurs*, and §4-5 analyze whether we have harm-based reason to censor or avoid using the terms this way.

First, a note about my decision to use the language of 'mental illness' and 'disorders' rather than the more positive 'mental health challenges' and 'neuroatypicalities' to frame the issue. Mental health challenges are a fairly expansive class of obstacles to one's mental well-being (including for example workplace pressures, financial stress, and the loss of relationships), that vary widely in their severity and impact on an individual. By contrast, 'mental illness' refers generally to a much narrower set of conditions that acutely interfere with an individual's day-to-day functioning and relationships, and which the DSM and portions of the World Health Organization's *International Classification of Disease* (ICD) attempt to codify. Everyone faces mental health challenges; but lumping mental illnesses in with these risks trivializing the differences they make to the lives of people who have them. 'Neuroatypicality' is slightly better, in that it picks out only large variations from the 'typical', but because it presupposes that the relevant dimension of difference is only neurological, it is too narrow for our purposes.

Still, the terminology I've decided to adopt isn't ideal. 'Illness' and 'disorder' suggest an unequivocally bad disfunction internal to an agent, and this can be misleading in a few ways. There is a strong case to be made that some conditions classified as mental disorders are simply neuroatypicalities, harmful only insofar as social structures are inimical and unaccommodating. Additionally, even having a detrimental mental disorder does not necessarily diminish one's mental well-being; one might be in circumstances that do not aggravate it, or have found an effective way to manage the condition (Keller 2019; Wakefield 2007). With the caveat that classifying a mental condition as a 'disfunction' is an ethically fraught and contested enterprise, I will defer to standard use in this chapter and speak throughout of 'mental illnesses' and 'disorders' in order to focus narrowly on the conditions that generally pose significant challenges for those who have them.

2 Diagnostic Applications

Debates over what the literal meanings of psychiatric terms are or should be find expression in controversies over revising the DSM. Now in its fifth major revision, the DSM is a manual of the diagnostic criteria for recognized mental disorders, published by the American Psychiatric Association and mainly designed to assist clinicians, but also used for a wide variety of other purposes. The manual's purpose, animating framework, and importance has changed significantly over the years. The first edition (published in 1952) was little more than a tool for standardizing the use of psychiatric labels applied in state institutions, to facilitate more accurate tracking of public health statistics. By the advent of the DSM-III (1980), however, the manual had both a more articulate self-understanding and a wider audience. While the first edition used language that was unabashedly Freudian to define some conditions, the DSM-III was explicitly

atheoretical, individuating conditions at "the lowest order of inference necessary to describe the characteristic features of the disorder". It was similarly cautious about ontological commitments, making "no assumption that each mental disorder is a discrete entity with sharp boundaries (discontinuity) between it and other mental disorders, as well as between it and No Mental Disorder." It also included a cautionary note that the manual was not designed "for non-clinical purposes, such as determination of legal responsibility, competency or insanity, or justification for third-party payment" ([APA] 1980).

Despite this disclaimer, DSM-III had an immense impact on how mental illnesses were treated, in every domain: in addition to guiding private clinical practice, its categories were used to define research programs, determine insurance coverage, establish claims to accommodations under the Americans with Disabilities Act, as well as to make various legal determinations of responsibility. These uses revealed inconsistencies in the diagnostic criteria, which were corrected by the DSM-III-R (1987). As psychiatric research progressed, pressure grew to provide empirical bases for diagnostic categories, and to align more closely with the categories in the ICD, and so after a six-year revision process, the DSM-IV was published in 1994. The next major revision started in 2000, and culminated in 2013 with the release of the DSM-5. The process began with academic conferences, thorough reviews of the relevant scientific literature published since the release of the previous edition, and the authorization of thirteen work groups tasked with drafting revisions to the classifications and diagnostic criteria. Proposals were field-tested and opened to public comment before being finalized. The resulting document is self-consciously a reflection both of professional consensus (where it exists), and of the limits of our understanding of the phenomena labeled pathological. The introduction emphasizes the provisional and instrumental nature of its categorizations:

> "the current diagnostic criteria are the best available description of how mental disorders are expressed and can be recognized by trained clinicians. DSM is intended to serve as a practical, functional, and flexible guide for organizing information that can aid in the accurate diagnosis and treatment of mental disorders. It is a tool for clinicians, an essential educational resource for students and practitioners, and a reference for researchers in the field." ([APA] 2013: 20)

Even so, the changes made---and even the revisions process itself---were intensely controversial. Many were frustrated that the revisions were too conservative, complaining that the DSM-5 does not provide an adequate framework for research design (Hengartner and Lehmann 2017), and suffers from being locked in to the path-dependent decisions of earlier versions of the DSM (Cooper 2018). Some of the restructuring decisions that *were* made—for instance merging a number of previously separate diagnoses into the single category 'autism-spectrum disorder'—were met with hostility and incredulity. The former because changing the available diagnoses can have outsize effects on the lives of those living with mental illness (service-users); the latter

because attempts to mitigate these disruptive effects of revisions yields diagnostic categories that are disjunctive and make little sense etiologically.

At first glance, it is not obvious why these revisions should be so fraught. There is obviously significant value in coordinating how the various stakeholders use the terms. The DSM reads as an index of recognized mental disorders, giving the reference and truth-conditions for sentences like "x is a mental disorder" or "x is schizophrenia," much as the periodic table lists the elements and gives conditions for sentences like "x is gold." But the project is not as simple as an index of chemical compounds would be; it does not simply report our most advanced understanding of some practice-independent object of inquiry out in the world. Nor is it a project of descriptive semantics, simply recording shifts in our application of mental illness labels over time. The DSM *constructs* a classification of mental disorders, but some are controversial (to varying degrees) within psychiatry, and many others are defined with reference to particular diagnostic tools or treatment methods. As these tools are refined and new methods trialed, our understanding of the conditions—and our concept of mental illness—evolves too.

Many of the fraught controversies over particular revisions trace back to more fundamental disagreements about how to think of the meanings of mental illness terms—what *kinds of things* they refer to, and what we are doing in using the terms. There is a rich debate about psychiatric classifications and kinds, but its primary focus is on questions (e.g. the scientific and metaphysical status of mental disorders) most at home in philosophy of science, and orthogonal to our linguistic concerns. In some ways, though, these disagreements re-enact classic disputes about meaning, and it can be illuminating to map the dominant positions in debates over mental illness terms onto standard (and perhaps more familiar) viewpoints in philosophy of language. I'll spend a little time doing that, but while I think this is a good way to better understand the relationship between the main positions in the debate, I am not optimistic that this mapping will be of much help in guiding future revisions. Even if we think that the terms should 'carve nature at its joints', and thus that our revisions should always aim to more closely track underlying divisions between conditions, there are too many joints—non-equivalent, competing ways of carving the conceptual space—and which is best depends on what we are trying to do. While in some sense the DSM simply articulates the appropriate application conditions of diagnostic labels, a lot hangs on these semantic questions, and the standards for the project are neither straightforward nor widely agreed.

What kind of thing does 'Mental Illness' refer to?

The question the DSM answers seems simple: *which conditions are mental disorders*? But to know how to revise its answers, we first need to settle on *what sort of thing* 'mental disorder' picks out. One might think that it refers to (i) the specific underlying processes causally responsible for the clusters of diagnostic symptoms, or (ii) just the collections of symptoms, or that (iii) it doesn't refer at all, but is a covert evaluative term. The first of these suggests that mental disorder labels

are natural or social kind terms, properly referring to all and only the conditions that share the relevant kind-individuating property, reminiscent of the causal theory of reference (Kripke 1972). The second takes the labels to be short-hand descriptions, referring to the condition of *having* the diagnostic symptoms, and might be akin to a sort of contemporary descriptivism (Jackson 1998). The third interprets the labels as thick terms, expressing social disapproval of the properties enumerated in the diagnostic criteria; in this it resembles various treatments of thick terms discussed in metaethics (Väyrynen 2019).

A wide range of views fall into the first of these options. Advocates of the biomedical model urge that "mental disorders be understood and treated as brain disorders" (Insel and Quirion 2005) and characterized in primarily biological terms (e.g. with reference to genetics, organ functioning, abnormalities in blood tests, MRIs, etc.).[2] This approach has substantial currency among organizations like the American Psychiatric Association, reflected in its reassurance that "mental illness is […] a medical condition, *just like heart disease or diabetes*" ([APA] 2018). If we take mental disorders to be individuated by the specific neuropathologies causally responsible for (most of) the behaviors and psychological symptoms in the diagnostic criteria, then if a condition that fits the criteria for 'schizophrenia' is *not* caused by the relevant underlying neuropathology, we should revise to exclude it, much as we should stop calling iron pyrite 'gold' after learning that it has a different chemical structure.

But one need not take psychiatric kinds to be biochemical to think that they are *real* kinds. One could instead individuate conditions by etiology, or by the affected mental mechanisms, or by responsiveness to particular treatments, and allow that psychiatric kind-terms denote partially socially constructed categories, much like 'weed'. The extension of such kinds transparently depends on contingent social facts. To be a weed is to be a thing with certain physical properties (a plant) that occupies a particular place in social practices – e.g. being unwanted, in competition with cultivated plants, etc. Weeds are a real, meaningful category of thing, but whether a specific plant is a weed depends on how it fits with local botanical preferences, *and* on the plant's intrinsic properties. Similarly, which conditions are mental disorders depends on how they interact with our local social practices, structures, and values, as well as on the underlying etiological/functional/causal properties of the conditions themselves.

This is the characterization implicit in the biopsychosocial model, which takes mental illnesses to be constituted by a nexus of irreducible, overlapping causal forces at biological, psychological, and social levels (Engel 1977; Davies and Roache 2017). It is explicit in the Harmful Disfunction model, which defines a mental disorder as a "failure of an internal

[2] The revisions that produced the DSM-5 were at least partially guided by this view, as the American Psychiatric Association explained their reorganization of the diagnostic categories by writing in the introduction that "It is hoped that this organization will encourage further study of underlying pathophysiological processes that give rise to diagnostic comorbidity and symptom heterogeneity" ([APA] 2013: 13).

mechanism to perform one of its naturally selected functions" in a way that, given social values, is injurious to the subject (Wakefield 2007: 150). It also underlies models on which mental illnesses are primarily socially-constructed kinds (Murphy 2001; Hacking 1999), but partly constituted by the underlying pathologies and expressed behavioral symptoms. On these ways of understanding psychiatric kinds, how we carve the term's extension must be sensitive to our interests and practices, instead of deferring wholly to the presence or absence of certain underlying biochemical features.

Alternatively, we might take the labels simply to abbreviate the description *having the properties listed in the diagnostic criteria* rather than refer to kinds. On this approach, revisions to the DSM aim not to more accurately track some underlying concept or phenomena, but rather to sort classes of symptoms into useful clusters, given our aims. This analysis can be appealing as an approach to clinical categories even if we believe that mental disorders are real kinds, but beyond our present epistemic reach. For treatment purposes, we may do better to take the labels to refer to symptom clusters rather than kinds. This approach was explicitly adopted for many entries in the DSM-III, but is less prominent in subsequent revisions.

Finally, one could accept the biomedical model's characterization of mental disorders, but take it to encode a conceptual mistake and therefore determine a necessarily empty extension. Szasz (1960: 118)'s famous declaration that "mental illness is a myth" takes this route. He contends that an illness is conceptually a disfunction in a physical system, but the sorts of disfunctions 'mental illness' is meant to pick out—unlike diabetes—distinctively concern the *contents* of one's mental representations; they involve *getting things wrong* in a sense that simply isn't applicable to physical disorders.[3] Nothing can satisfy the description of being a neuropathology that causes symptoms that can only be specified normatively. It would follow that 'mental illness' fails to refer, but that need not imply that the concept is contentless. Rather, like the predicate 'witch', mental illness labels could be covertly evaluative: they express and aim to elicit normative condemnation of the object of predication, in this case of mental states or conditions that deviate from accepted social norms.

Negotiating the Meanings of Mental Illness

Each of these different ways of characterizing the referents of 'mental illness' yields a different prescription for revisions, and we have hardly even scratched the surface of (much debated) questions about the metaphysical nature of natural kinds in psychiatry. But I want to suggest that none of these disagreements, as deep and divisive as they are, is actually the core of the issue we face when revising the DSM. Because of the tight connection between the categorization and

[3] Arpaly (2005) articulates this difference in depth, though she does not endorse the antirealist model.

diagnostic criteria given in the DSM and our social practices concerning what and how we recognize or treat as mental illness, making these revisions is radically unlike refining our criteria for distinguishing beech trees from elms. We cannot uncontroversially frame the project as simply trying to find the best way to characterize the semantic content of an expression for some static object of inquiry. Nor even can we hold the concept fixed and simply attempt to more accurately capture its extension. When revising the DSM, we confront problems in conceptual ethics, not conceptual analysis: the issue is not primarily what we *do* mean by the terms, but rather what we *should* mean by them, given their roles in our social, political, and scientific practices.[4] There is much more at stake than reference and truth-conditions.

Identifying the debate as a case of conceptual ethics names the issue, and offers a framework for understanding the problem we face, but does little to help resolve it. As with *disease*, there is no goal-independent right answer to how *mental disorder* should be conceptualized, precisely, to which we can simply defer. Whether the condition we're concerned with is physical, like *infertility*, or mental, like *depression*, labelling it an illness or disorder does much more than merely classify the condition. It provides an interpretive framework, sets social scripts in motion, and has implications for agents' political rights.[5] Controversies over the definitions of terms in the DSM are proxy battles, the linguistic site of the underlying struggle over the priorities of psychiatry. The terms matter so much because the DSM's articulation of mental disorders serves simultaneously as dictionary, clinician's handbook, research framework, standardization metric, and insurance guide, and consequently any revision affects many stakeholders with competing interests.

When DSM categories are used to set up randomized controlled trials, researchers have an interest in using the labels to refer to all and only conditions that are the 'same' in the sense causally relevant to the experiment—which varies depending on the nature of the intervention being tested. Some, like selective serotonin reuptake inhibitors (SSRIs), are effective across a wide range of diagnostic categories, while more targeted interventions require fine-grained biochemical divisions. Pharmaceutical research may have reason to generally prefer variations of the biomedical model, while other psychiatric research, aimed at discovering a wider range of social risk factors, prognosis, and cognitive/behavioral treatment options, have reason to

[4] (Burgess and Plunkett 2013). Conceptual ethics is also called 'ameliorative analysis' (Haslanger 2004), or 'conceptual engineering' (Cappelen and Plunkett 2020).

[5] Kukla (2017) and Cooper (2016) make related points for application of the terms 'disease' and 'psychopath', respectively.

individuate conditions according to factors including not only genetics and biomarkers, but also individual personality, social conditions, and specific etiology.[6]

Other stakeholders, including social workers and clinicians, have interests that are application- rather than research-oriented. Because they address the effects of mental conditions at the social level, these practices have an interest in individuating disorders by their effects on a person's ability to behave in various socially important ways, or by their responsiveness to particular types of treatment. Clinicians, whose primary focus is on diagnosis and treatment, need a nosology that will enable them to reach a verdict on whether their client has a disorder, and indicate how it can be treated. For other applications, including decisions about resource distribution (evaluating claims for insurance, special accommodations, or other entitlements), or individuals' legal standing (e.g. assessing fitness to stand trial, to be released after serving a sentence, to be licensed for certain professions, or to fulfil parental responsibilities), it is crucial not only that the diagnoses closely track relative fitness for such treatment, but also that they be *reliable*, in the sense that they do not vary substantially across different clinicians.

Service-users, meanwhile, have a wide mix of interests at stake in the DSM's nosology. Some support a relatively expansive, symptom-based individuation of mental disorders: insurance coverage, and thus access to services, accommodation, and treatment, is in many cases conditional on having a recognized diagnosis. More informally, a diagnosis can offer reassuring confirmation that one's condition is *real*, and help friends and family recognize that it is neither something one can "just snap out of", nor a character flaw for which one should be blamed (Arpaly 2005). These interests can come dramatically apart from research aims: even conditions that are not psychiatric disorders can sometimes be alleviated by psychiatric treatment (Clutton and Gadsby 2017).

But service-users also have strong interests in construing mental disorders narrowly and individuating them finely. If diagnostic categories are too broad, they undermine service-user's chances of receiving an informative prognosis and being offered the most effective treatment options.[7] And being diagnosed as 'mentally ill', or sharing a diagnosis with more stigmatized conditions, exposes subjects to significant social risks. Many mental disorders have symptoms

[6] For more detailed discussion, see generally Tekin and Mosko (2015); Hengartner and Lehmann (2017); Cooper (2018)

[7] The interest in effective treatment does not always cut the same way. It can take the form of emphasizing different aspects within a model, as the Positive Psychiatry movement does in urging researchers and clinicians to devote more attention to a wide range of factors (neurological, as well as psychological and social) associated with positive prognosis and recovery for patients with schizophrenia (Jeste, Palmer, and Saks 2017). Or it can militate against classifying a pathology as a *mental* disorder—as opposed to a physical disease. For example, some patient advocacy groups opposed the classification of Chronic Fatigue Syndrome/Myalgic Encephalomyelitis as a psychiatric disorder in the DSM-5, citing fears that this would prompt treating CFS as a delusion, rather than continuing to seek a biological explanation of the symptoms that constitute the syndrome (Hawkes 2011).

that temporarily compromise some aspect of agency, for which the recommended treatment involves curtailing autonomy. There is great risk of this extending too far, dismissing a patient's credible testimony and unjustly violating her autonomy. These concerns are particularly urgent because the DSM is *de facto* authoritative: by classifying a condition as a mental disorder (or not), the DSM *makes it so* socially. Regardless of whether the DSM's classifications attempt to track an underlying kind, they determine, by stipulation, which conditions are included in the socially created category 'mental illness', and thus also who gets labeled 'mentally ill.'

These risks may be exacerbated by a biomedical model, as its emphasis on controlled trials and clinical evidence devalues patient testimony as anecdotal (Crichton, Carel, and Kidd 2017; Faulkner 2017). Similarly, defining disorders by reference to departures from typical functioning—rather than explicitly requiring that they also be harmful—risks pathologizing the merely different. This is especially treacherous because it implies that the only rational response to neuroatypicality—a *disfunction*—is to desire a cure. This inclines us to reject service-user's testimony about the positives of her condition, taking it as evidence that she is epistemically compromised rather than evidence that we are mistaken in classifying her condition as a disorder. The danger is not merely theoretical: homosexual attraction was infamously included in the DSM as a psychiatric disorder until 1973, and several conditions currently classified as disorders are contested, as those who have the conditions urge that they are neuroatypical, not dysfunctional.[8]

There are deep similarities between these and the issues that arise for the language and social attitudes concerning physical disabilities. Many service users emphasize that though they benefit, sometimes dramatically, from psychiatric treatments, it is the socially created harms—stigma and the lack of accommodation—not the disorder itself, that most significantly negatively impact their wellbeing (Thornicroft, Rose, and Kassam 2007). The particulars vary between conditions: while those with depression or anxiety may be viewed as fragile or incompetent, they may escape the suspicion and fear of violence often directed at those with diagnoses involving psychosis or personality disorders.

While our linguistic practices aren't wholly responsible for these social attitudes, they do reflect and can influence them. The difficulty is deciding which tradeoffs to make. Characterizing mental disorders as neurobiological kinds "just like diabetes" discourages blaming agents for their conditions, but encourages thinking of agents as having a permanent *disposition* to manifest the disorder, and casts the condition both as purely bad, and as the domain of expert medical practitioners. A metastudy of public reactions to explanations of psychiatric disorders found that "People who are given a neurobiological explanation of a psychiatric condition tend to see

[8] These include diagnoses of Autism Spectrum Disorder, Attention Deficit-Hyperactivity Disorder, Bipolar Disorder, Epilepsy, Tourettes, among others (Fenton and Krahn 2007).

sufferers as more dangerous and less likely to recover, and are more likely to distance themselves from them, than people who are not" (Loughman and Haslam 2018). Taking 'mental illness' to be partially socially constructed is somewhat less likely to encourage essentializing the conditions, as the very contingency of social categories underscores that they could have been otherwise. It also leaves more space, at least in principle, to acknowledge the ways some service users experience their conditions as enriching their lives. But it has other downsides: it cannot be expected to provide categories that would be useful for finely-tuned research, and leaves service-users more vulnerable to blame, as well as doubts about the reality of their conditions.

The various uses give mutually incompatible prescriptions for revising the DSM's specifications, and while there are some reasons in favor of retaining a single framework for all these uses—mutual intelligibility and a common language across disciplines chief among them—they may be outweighed. The National Institute of Mental Health, serving the interests of psychiatric pharmaceutical researchers, has already abandoned the DSM and launched its own research framework, the *Research Domain Criteria* (RDoC).[9] The *Hierarchical Taxonomy Of Psychopathology* (*HiTOP)* may serve a similar role for other psychiatric research (Hengartner and Lehmann 2017), leaving future revisions of the DSM answerable only to application-based interests.

For their part, service users have launched several 'Mad-Positive' campaigns to reshape at least the social meanings of mental illness, including Mad Pride (Poole et al. 2012), the Icarus Project (DuBrul 2014), and the Neurodiversity movement (McWade, Milton, and Beresford 2014). Modeled explicitly on similar LGBTQ and disabilities anti-stigma campaigns, they aim (among other things) to draw attention to the ways in which mental conditions can enhance the lives of those who have them, and to advocate for accommodations and treatments developed in close consultation with those with lived experience with the conditions.

The controversies over revisions to the literal meanings assigned to mental illness terms in the DSM highlight the political dimensions of linguistic decisions that may have looked straightforward at first sight. In addition to the important questions about the ontology of mental disorders—which have been pursued vigorously in the philosophy of psychiatry—we must also ask what mental illness terms *should* mean; which of the several available concepts we have most reason to use them to express. Whatever answers we arrive at to the first class of questions, this second set must be faced before we can give univocal recommendations to guide future revisions to the DSM. Because the interests of the affected groups pull in starkly different directions, the language of mental illness nicely underscores the fact that once we identify a linguistic dispute as

[9] NIMH officially withdrew support from the DSM in 2013, citing the conflict between these aims and the symptom-driven DSM categories. They have introduced the RDoC as a research framework instead, with the goal of characterizing normal functioning in biogenetic terms, so that deviations can be identified for treatment. This faction see the future of psychiatry as "clinical neuroscience" (Insel 2014).

a problem in conceptual ethics, we have named but not solved the issue; there's difficult work to be done to determine how to balance the competing interests to identify which set of concepts we ought to be using. While what our terms *do* mean---the semantic question---is often underdetermined or indeterminate, what they *should* mean is a question squarely in political philosophy.

3 Metaphoric Applications

As mad-positive movements gained social visibility,[10] people began to ask more transparently political questions about the use of terms invoking or referring to mental illness in non-diagnostic contexts. As a case in point: in 2014 Smith College hosted a panel discussion on free speech and hate speech. The transcript of the event replaced occurrences of familiar slurs with '[n-word]' and '[c-word]'—and of 'crazy' in the sentence "We're just wild and crazy, aren't we?" with '[ableist slur]'. That last replacement sparked a flurry of Op-Eds over whether the term is properly a slur, and whether we should censor or attempt to restrict its use.[11] The rest of this chapter takes up these questions about broadly metaphoric—or at least not literally descriptive—uses of terms that invoke or refer to mental disorders (e.g. 'crazy', 'nuts', 'paranoid', 'schizo', 'psychopath', 'klepto', 'depressed', 'OCD', 'mental', etc.). The interest in whether terms like these are slurs stems, I think, from a deeper question: whether we have moral reason to change our use of the terms. The underlying assumption on both sides of the dispute is that if a term is a slur, it follows that we should avoid using it. That's well and good. But the inference running the opposite direction—from a term's *not* being a slur to the *absence* of a reason to avoid it—lurks beneath the surface, and is treacherous.

Slurs and harm

Slurs are most readily recognizable in what Jeshion (2013b) calls *weaponized uses*, in which they are applied directly to members of the target group in order to belittle, derogate, intimidate, and insult them *qua group member*. Slurs are *dysphemisms* (the inverse of euphemisms): they are markedly impolite terms for a group, and are often subject to social taboos of varying strength. They call to mind negative stereotypes of the group, are identifying and totalizing: they imply that the individual's group membership is the most significant aspect of their identity, and that it is worthy of contempt.[12]

[10] See Abraham (2016); Glaser (2008) for more background on 'mad positive' social movements.

[11] See for example (Silvergate 2014; Kaminer 2015).

[12] These are frequently offered as characteristic of slurring terms (see, e.g. Anderson and Lepore 2013; Bach 2018; Blakemore 2015; Croom 2015; Camp 2013; Jeshion 2013a; Nunberg 2018). Bach distinguishes what he calls

The rationale for restricting slurring speech through legal or social sanctions is that slurs are actively harmful, not merely offensive. Some of these harms are indirect, mediated via the cognitive effects on slur users: sustained use of slurs foregrounds contempt and frames targets as moral inferiors for whom subordinate social roles and oppressive treatment (or worse) are appropriate.[13] The more immediate, direct harms fit Waldron (2009)'s regimentation of the types of direct harms imposed by hate speech. First, by expressing aggressive and contemptuous attitudes for the target *qua* group member, a slur harms the immediate target by intimidating, threatening, or humiliating her in the context of utterance. Second, slurs affront the dignity and social standing of all members of the target group by conveying some variant of *people of that group are contemptible,* and raising derogatory stereotypes to salience.

Third, an aggregative harm arises from widespread use of slurs. Using a slur identifies the speaker as someone who endorses the associated negative attitudes toward the target group, and encourages others to do so as well. Targets learn that the speakers not only are not committed to democratic tolerance, but in fact *aim to exclude* their targets from full participation in the political community, and are willing to incur social costs in pursuit of this aim. The larger the set of slurring speakers, the more likely it is that they will succeed in gaining political power, and thus the more they undermine the targets' assurance of full participation in the political community. Importantly, and unlike simple insults, slurs can cause these harms even if the present speaker has only good attitudes, and does not even mean to be offensive, let alone harmful. Theorists give widely varying accounts of how these effects relate to the semantics of slurs; for the purposes of this article, I'll stay neutral.

Mental Illness Pejoratives

So, we have reason to avoid using slurs because they express derogatory attitudes in harmful ways. And we have reason to censor and restrict others' use of slurs in order to protect targets from the intended harms. Do these reasons generalize to mental illness terms? In some cases, yes: weaponized uses of terms like 'psycho' and 'crazy', in which the speaker intentionally targets people with mental illnesses, can be readily analyzed alongside paradigmatic race- or gender-based slurs. But the terms are also used pejoratively in an entirely different way, in constructions like "don't be so paranoid", "don't listen to him, he's mental", "the opposition party is full of psychos who want to drive the country into the ground", or even "we're just wild and crazy."

---

personal slurs (terms like 'jerk' and 'dweeb') from *group slurs*, and proceeds to treat group slurs as the central case. This distinction roughly tracks the distinction between general pejoratives and slurs that I appeal to in this chapter.

[13] Theorists vary in characterizing this set of effects. For example, Camp (2013) holds that slurs propagate a contemptuous *perspective* on their targets, eroding the targets' social standing over time. Tirrell (2017) emphasizes that slurs lead users to re-conceptualize the targets as less than human, the first step in committing worse crimes against them. Jeshion (2016) glosses contempt expressed with slurs as a sort of emotional contagion, which can be expected to spread to naive members (i.e. children) of communities with slur users.

These are pejorative uses of group-referencing terms, but seem to demand a different analysis than the one above.

Even when used pejoratively, many of these lack the hallmarks of paradigmatic slurs. Most are not dysphemistic or subject to social taboos. Though some are clearly the slang variant of more technical terms (e.g., 'schizo' and 'schizophrenic'), others either *are* the diagnostic term (e.g. 'paranoid') or have no non-pejorative equivalent (e.g. 'crazy/bonkers'). They are frequently used without hesitation even in polite contexts, by educated and socially conscientious speakers. And they are not stably derogatory: formulations like 'that's crazy/insane/nuts' can express praise or surprise as well as derision, and which content they express appears highly dependent on speaker intent and context.

When speakers do use these terms derogatorily, they often target individuals whom they believe do *not* have the relevant mental condition, and carefully refrain from applying the terms to someone they know to have it. These pejorative uses cannot be intended to denigrate their target *qua group member*, since the target is believed not to be a member of the group at all. Nor are they characterized by vitriol or hateful contempt for those with mental illnesses, but rather by a casual dismissal of the immediate target and refusal to treat them as a full member of the epistemic or political community. Speakers apply terms like 'crazy', 'loony', or 'paranoid' to convey that the immediate target is surprising, irrational, or out of control, but often neither speaker nor hearer spares a thought for genuine mental disorders.

These uses behave more like *general pejoratives* (e.g. 'asshole', 'prick', 'jerk', 'pussy', etc.) than slurs. While slurs insult *qua group member*, general pejoratives are insults used to condemn a particular person for behavior at a given time, and may or may not reference a group in doing so.[14] Though pejorative terms vary in impoliteness, they are not generally marked dysphemisms. Whether their use is offensive is highly contextually variant: constructions like 'you asshole!' range in expressive force from gentle teasing to severe insult, depending on speaker intent and other contextual features (Culpeper 2011). General pejoratives apply to their targets in a loose, metaphoric way. You will find it difficult to give precise application conditions for 'prick' or 'shithead', but if a target has none of a cluster of rude or self-centered properties, the application is inapt. Similarly, 'paranoid' is pejoratively applicable to a target only if he displays some level of unjustified fear, but the degree, frequency, and extent is under-determined. 'Crazy' is even less constraining.

[14] For a thorough discussion of the differences between slurs and general pejoratives, see Hay (2013) sections II and III; Camp (2013) also discusses some reasons to separate slurs from pejoratives in our theorizing. The distinction between slurs and general pejoratives isn't absolute, and is occasionally clouded by terms like 'bitch' which have both a slurring and a general pejorative use, but it does capture something generally distinctive and important about the terms and is close enough for government work, as they say. Nunberg (2018) examines some such crossover slurs, explaining how a slur can drift into use as a general pejorative and visa-versa.

The application patterns of general pejoratives reveal an interesting, and for our purposes important, subdivision in the category. Some, like 'asshole', do not invoke a social group: there is no stable set of people, the assholes, who are descriptively invoked in each use. Others, like 'pussy' and 'girl' in constructions like "Don't be such a _____", do implicate a stable group of people: women. Call the second type 'derived pejoratives', to distinguish them from unfocused general pejoratives. When derived pejoratives are used in a weaponized or derogating way, they tend to be directed at *non-group members*: the typical target of "don't be such a girl" is male. Pejorative uses of mental illness terms pattern together with pejorative uses of 'girl' and 'retarded' in this respect (Siperstein, Pociask, and Collins 2010).

These observations suggest that mental illness-based pejoratives (or 'MI-pejoratives') are derived general pejoratives, not slurs. But if our motivation in asking whether they are slurs is that we want to know whether we should change our use of the terms, the real question is not whether the term fits in this or that semantic category, but rather whether it is harmful. Even when people with mental illnesses aren't the direct target of MI-pejoratives, they are nevertheless implicated in uses of the term. But do these uses harm people with mental illness, either individually or as a group, such that we have reason to change our use of the terms? I think it is very plausible that they do, albeit indirectly. But precisely because the harm is indirect, and the uses are not *aimed* at harming them, we cannot explain the harm done using the hate speech model. And while the existence of a derived pejorative often co-occurs with stigmatization or devaluation of the implicated group, the fact that a group faces harmful stigma or devaluation in a society does not show that the *terms* used to refer to them are causing harm. To do that, we will need a model articulating the role of the terms in producing these harms.

4 Beyond Hate-Based Harm

Even when not directed at a target who actually has a mental illness, using MI-pejoratives metaphorically to divest targets of practical or epistemic authority can increase the marginalization and stigmatization of those who do. To fill in this general sketch, it'll be useful to look in detail at a closely related form of wronging: epistemic injustice. Obviously, some MI-pejoratives principally target non-epistemic agential capacities, but let's fill in this profile before generalizing it. As Fricker (2007) argues, being taken seriously as a reliable source of information on at least some topics, some of the time, is essential to an individual's ability to function socially as an autonomous moral agent. On her original model, agents are owed credibility in proportion to their actual reliability; whenever they are treated with less trust than they merit, they suffer testimonial injustice via this 'credibility deficit'. When this happens in an identity-tracking way, such that an agent frequently and disproportionately suffers testimonial injustice on a variety of topics and in a wide array of contexts, it undermines her ability to interact as a full member of the

epistemic community. Particularly in a knowledge economy, being denied appropriate credibility both devalues the agent and hinders her pursuit of other social goods.

There are many ways to articulate the mechanics of these wrongs; I find Pynn (forthcoming)'s proposal particularly illuminating. On his analysis, testimonial injustice occurs when hearers publicly violate their obligation to accept a speaker's assertion that P when the speaker *knows that P*. The key element in explaining the harm of this injustice is its *publicity*: when unjustly dismissed to an audience, the speaker is *degraded*, represented to others as having lower status as a knower than she in fact deserves. This is "an intrinsically social wrong with a complex structure. Its victim is violated—which is a wrong itself—but in addition, the violation is of such a kind as to diminish the victim in the eyes of others." When a wrongful rejection makes salient harmful stereotypes of the speaker's group, this harm is compounded, representing "the victim as a non-knower who is debased in all the ways encoded by the stereotype" (Pynn forthcoming: 13). So, unfairly rejecting a speaker's contribution on account of her social identity represents (i) her as a non-knower, and (ii) the negative stereotype of her group is true in general, and of her in particular. This is significantly more diminishing than merely being represented as not knowing P in the immediate context, and hence the degradation involved is more severe.

If this is a plausible understanding of the harm of stereotype-based testimonial injustice, it goes a long way toward explaining the felt similarity between MI-pejoratives and paradigmatic slurs. One of the central ways that weaponized slurs harm their victims is by portraying them as worthy of lower status than they in fact deserve, and as being such that negative stereotypes of their group are true in general, and of them in particular. That said, Pynn's gloss cannot be used to straightforwardly explain how MI-pejoratives wrong *people with mental illness*, since often they are not the direct target. So how, if the immediate victim of unfair dismissal with an MI-pejorative *isn't* someone in the term's descriptive extension, does it nevertheless visit harm on actual members of the group?

Here's my proposal. Rejections or dismissals made by invoking a stereotyped social identity involve two representations: one of the individual who is the immediate target, and one of the group that is the basis of the stereotype. By calling someone's fears 'paranoid' or her suggestion 'crazy', the speaker implies that the target ought to receive the treatment appropriate to someone presently in the grip of psychosis or paranoid delusions: her claims should be dismissed.[15] Prescriptive metaphors like these set up a feedback loop. When our linguistic practice reinforces the idea that people who are described as 'crazy', 'schizo', or 'paranoid' are to be dismissed as reliable sources of testimony, it bleeds over into our treatment of individuals who

[15] Actually this overstates the appropriate treatment: even during a psychotic episode, people with delusional or paranoid disorders are not unreliable about all topics, but just those involved in the disorder, and their knowledge claims on unaffected topics *ought not* be dismissed.

actually have schizophrenia or a paranoid disorder. We must go carefully here; plausibly it is not unjustly diminishing to represent an assertion made while in the grip of psychosis as unreliable. So if derogatory uses of 'paranoid' invoked only the property of *being symptomatic of an occurrent delusion*, it might be harmless. But we quickly dispositionalize properties, and this tendency is even stronger when there is some condition underlying the manifest behavior, such as having been diagnosed as vulnerable to paranoid delusions.[16] Even if we do not wrong an individual in discounting her testimony during a delusion, we clearly would wrong her in dismissing her testimony more generally—regardless of how lucid she is—simply because she is occasionally vulnerable to such delusions.

So, pejorative uses of mental illness terms potentially inflict two wrongs: if the immediate target in fact deserves more credibility, she is directly wronged. But even if her comment deserves dismissal, employing an MI-pejorative raises to salience a stereotype of mental illness as paradigmatically untrustworthy, and by implicit endorsement reinforces the assumption that people with mental illnesses ought to be generally dismissed as epistemic agents. This entrenches the social stigma of mental illness and represents a class of people as deserving treatment far below what they deserve, wronging the group thus indirectly invoked.[17] Of course, some mental illness pejoratives indict one's practical authority, or moral rather than epistemic agency; see for example 'klepto', 'narcissist', 'sociopath', etc.[18] Generalizing to accommodate these cases, we may offer the following as a final analysis of how derived pejoratives function:

> A derived general pejorative invokes stereotyped (often negative) properties of a group as appropriate grounds for exclusion from or subordination within the relevant moral, political, or epistemic community. It (i) represents the immediate target as presently (though temporarily) deserving this treatment because they manifest similar properties, and (ii) represents the invoked group as stably possessing such properties.

There are a variety of questions we could ask about the semantics of these pejoratives, including how these terms 'invoke' the relevant stereotyped properties. A similar question arises for slurs, and theorists disagree: some build them directly into the truth-conditional semantics (Hom 2008), or incorporate them as lexical presupposition (Lycan 2015), while others fold them into the pragmatics, appealing to perspectives (Camp 2013) or cognitive salience (Jeshion 2013b;

---

[16] See Leslie (2017) for a discussion of some of the relevant psychological literature on this tendency. It is an unfortunate consequence of the biomedical model of mental illness that it had the unintended side-effect of strengthening the perception that mental illness is a stable dispositional trait (Corrigan and Kosyluk 2013).

[17] Notice that none of these harms depend on the speaker's *intending* to represent the invoked group as disposed to be untrustworthy; the harms of social stigma, like the harms of slurs, arise from how the terms are typically *understood*, rather than the speaker's intended meaning on an occasion of use. (Bolinger 2020).

[18] My thanks to several discussants, but especially Katheryn Lindeman and Eric Wiland for suggesting this extension.

Nunberg 2018). But as interesting as it is, I want to set this question aside in order to focus on the political questions raised by *uses* of derived pejoratives. Participation in the moral, political, and epistemic community is important for agents' well-being and necessary for functioning as a moral equal. Whether brought about intentionally or not, identity-tracking exclusion or subordination in these communities threatens to undermine valuable agency, and when undeserved, is a serious moral wrong. These wrongs are, I believe, significant enough to give well-intentioned speakers sufficient reason to avoid using mental illness terms pejoratively, and to try to persuade others to do likewise. The natural next question is whether we ought to do more. Should we attempt to introduce social sanctions, to compel those whom we cannot convince?

5 Should we stop saying 'crazy'?

An argument for such a policy might proceed as follows: the use of MI-pejoratives harms those with mental illness by representing them in degrading ways and reinforcing attitudes that marginalize and exclude them from full participation in the epistemic, moral, and political community. These harms are similar in kind to those caused by potent racial and ethnic slurs, so if we are justified in censoring and restricting the use of slurs, we are likely also justified in similarly censoring and restricting use of MI-pejoratives.

We *could* say this, but I think we shouldn't. The differences in what speakers *aim* to do with derogatory uses of slurs and derived pejoratives—and the resultant differences in the ways each eventuate harm—have significant implications for how best to address the harms they cause. A weaponized slur aims to harm *both* the individual target *and* the group; it is an intentional attack on both targets. Such attacks can be made in many ways, of course, but the sting of using a slur derives from the fact that *that term* is socially marked as a weapon of this kind. New terms can be introduced, but lack the currency and expressive force of marked and tabooed slurs; dysphemisms are uniquely well-suited to inflicting the intended harms. Consequently, harms from slurs can be effectively addressed by controlling uses of the specific terms.

Derived pejoratives are different. Unlike slurs, we have no strong reason to avoid the terms *altogether*; they aren't necessarily pejorative on every occasion of use. One can presently describe a roadtrip as 'insanely cool!' without obviously eventuating any harm, even indirectly; these uses are simply different in kind from pejorative uses of the term. Also unlike slurs, the harm from a derived pejorative isn't anchored to a particular dysphemistic term: it comes from the metaphor. So long as the characteristic or stereotypical properties of a group are socially accepted as appropriate grounds for exclusion from the community, any expression referring to that group can be used as a derived pejorative. Experiences in other similar cases suggests that

without a shift in the underlying social attitudes, social or legal restriction of the terms will at best effect a merely linguistic shift while leaving the underlying harms undiminished.

Our linguistic practices and the attitudes expressed by MI-pejoratives are in a feedback loop much like pejorative uses of 'girl' and sexist or misogynist attitudes, and 'retard' and dismissive attitudes towards those with intellectual disabilities. The terms in their descriptive and literal application apply to individuals with straightforward characteristics, without prescription or evaluation: 'girl' is a descriptive predicate with an extension that includes young females;[19] 'paranoid' and (until 2005) 'retarded' are descriptive predicates with an extension that includes individuals who satisfy certain diagnostic criteria. But in their extended, pejorative applications ("don't be such a girl", "that's crazy", "that's retarded"), the terms reflect the social appraisal and stereotypes of the individuals in the descriptive extension: they are applied metaphorically to indicate that someone is properly subject to certain treatment or displaying certain disvaluable properties.[20] Initially, there is nothing objectionable or markedly negative about using the *terms* themselves, non-pejoratively. This changes over time: as the pejorative use's emphasis on the devalued properties entrenches an increasingly negative stereotype, the term takes on a marked negative aspect.

In lieu of a taboo on using the terms, we might attempt to insulate the group from the harms of pejorative uses by rejecting the terms as descriptive predicates for the group. If the pejoratives survive, they will be dead metaphors, no longer raising stereotypes of the group to salience in delivering the intended insult. Plausibly this has occurred with 'idiot' and 'anal', at least, insofar as 'anal retentive' is no longer taken to denote a genuine mental illness. But if the negative social attitudes about the group persist, whichever term is chosen as the replacement group-referencing term will, over time, acquire a pejorative use. Early efforts to escape the group harms resulting from the pejorative use of 'idiot', 'moron', and 'imbecile'—which had been the descriptive terms for varying levels of intellectual disabilities—resulted in reclassification in the 1980s as 'retardation', which remained the official diagnostic term for the group until replaced by 'intellectual disability' between 2005 and 2013.[21] The rapid and thorough conversion of 'retard' and its cognates into a pejorative shows that ameliorative attempts that focus exclusively on the language used are unlikely to have much success in the long run. In the face of persistent negative social evaluations, if there is to be any hope of protecting group-members from the harms associated with derived pejoratives, attempts to reform the linguistic practices must be paired

[19] Or young people who express a feminine gender, depending on the use.

[20] These terms bear some family-resemblance to what Sarah-Jane Leslie calls 'dual character concepts' (Leslie 2015). However, whereas those terms (e.g. 'real man') work by invoking normative *ideals* encoded in a social role or kind, derived pejoratives leverage the negative associations. Cousens (2020) leverages this fact to argue that terms like 'moron' are in fact slurs, in that they are terms used to perform oppressive speech acts.

[21] (Salvador-Carulla and Bertelli 2008; Schalock 2002). Terminology changed at the US federal level in 2010.

with efforts to uproot the social stigma. The goal must be to undercut the devaluing social attitudes, thereby blocking the inference from being a group member to being an appropriate target of epistemic, moral, or political exclusion.

In both literal diagnostic contexts and extended metaphoric uses, one of the principal ways our current linguistic practices harm those with mental illnesses is by combining with widespread, unjust prejudice to exacerbate group members' exclusion from the relevant (epistemic, moral, and political) communities. While changing our linguistic practices is unlikely, on its own, to resolve this in either context, an intentional shift in attitude---accompanied by structural changes and using a shift of language to signal sincerity---may well be the best way to address both. Since the harm in derived pejoratives stems from the connection between the metaphoric insult and the reference to the group, there are two ways to address it. One option, if the pejorative use of the term is not yet entrenched, or can be dislodged, is to implore speakers to refrain from using the terms pejoratively. This strategy, or something like it, eventually worked for 'queer' and 'gay'. This may be the best strategy for useful diagnostic terms, and is the self-proclaimed strategy the of Mad Pride movement for terms like 'nuts' and 'mad'.

The alternative is to sever the referential connection between the term and the group, by rejecting the term in much the same way that 'retard' and its cognates have been rejected as a term for intellectual disabilities. It's unclear what we should expect of pejoratives like 'mental', 'delusional', and 'insane' if this strategy is successfully applied. Divorced from their metaphorical significance, it's possible they'll simply fade from use (much as the term 'feeble-minded' has). It's more likely they'll follow the trajectory of 'idiot' and 'imbecile', losing their literal reference completely and retaining only a pejorative content. The words might then still be used to express that someone's testimony is unreliable, unreasonable, not appropriately responsive to evidence, but without any active connection to mental illness, and thus without precipitating the indirect group harms that such uses currently cause.

References

[APA], American Psychiatric Association. 1980. *Diagnostic and Statistical Manual of Mental Disorders, 3rd edition*.

———. 2013. *Diagnostic and Statistical Manual of Mental Disorders, 5th edition* (Washington, DC).

———. 2018. 'What Is Mental Illness?', Accessed January 29. *https://www.psychiatry.org/patients-families/what-is-mental-illness*.

Abraham, Amelia. 2016. "Remembering Mad Pride, The Movement that Celebrated Mental Illness." In *VICE*.

Anderson, Luvell, and Ernie Lepore. 2013. 'Slurring Words', *Nous*, 47: 25-48.

Arpaly, Nomy. 2005. 'How it is not "just like diabetes": Mental Disorders and the Moral Psychologist', *Philosophical Issues*, 15: 282-98.

Bach, Kent. 2018. 'Loaded Words: On the Semantics and Pragmatics of Slurs.' in David Sosa (ed.), *Bad Words* (Oxford University Press).

Blakemore, Diane. 2015. 'Slurs and Expletives: A Case Against a General Account of Expressive Meaning', *Language Sciences*, 52: 22-35.
Bolinger, Renée Jorgensen. 2020. 'Contested Slurs: Delimiting the Linguistic Community', *Grazer Philosophische Studien*, 97: 1-22.
Burgess, Alexis, and David Plunkett. 2013. 'Conceptual Ethics', *Philosophy Compass*, 8: 1091-101.
Camp, Elizabeth. 2013. 'Slurring Perspectives', *Analytic Philosophy*, 54: 330-49.
Cappelen, Herman, and David Plunkett. 2020. 'A Guided Tour of Conceptual Enineering and Conceptual Ethics.' in Alexis Burgess; Herman Cappelen; David Plunkett (ed.), *Conceptual Engineering and Conceptual Ethics* (Oxford University Press: Oxford).
Clutton, Peter, and Stephen Gadsby. 2017. 'Delusions, Harmful Dysfunctions, and Treatable Conditions', *Neuroethics*.
Cooper, Rachel. 2016. 'First Do No Harm?: What Role Should Considerations of Potential Harm Play in Revising the DSM?', *Philosophy, Psychiatry, & Psychology*, 23: 103-13.
———. 2018. 'Understanding the DSM-5: stasis and change', *History of Psychiatry*, 29: 49-65.
Corrigan, Patrick, and Kristin Kosyluk. 2013. 'Mental Illness Stigma: Types, Constructs, and Vehicles for Change.' in Patrick Corrigan (ed.), *The Stigma of Disease and Disability: Understanding and Overcoming Injustices* (American Psychological Association).
Cousens, Chris. 2020. 'Are ableist insults secretly slurs?', *Language Sciences*, 77.
Crichton, Paul, Havi Carel, and Ian James Kidd. 2017. 'Epistemic Injustice in Psychiatry', *BJPsych Bulletin*, 41: 65-70.
Croom, Adam. 2015. 'The Semantics of Slurs: A Refutation of Coreferentialism', *Ampersand*, 2: 30-38.
Culpeper, Jonathan. 2011. *Impoliteness: Using Language to Cause Offense* (Cambridge University Press: Cambridge).
Davies, Will, and Rebecca Roache. 2017. 'Reassessing Biopsychosocial psychiatry', *The British Journal of Psychiatry*, 210: 3-4.
DuBrul, Sascha Altman. 2014. 'The Icarus Project: A counter narrative for psychic diversity', *Journal of Medical Humanities*, 35: 257-71.
Engel, G. 1977. 'The Need for a New Medical Model: A Challenge for Biomedicine', *Science*, 196: 129-36.
Faulkner, Alison. 2017. 'Survivor research and Mad studies: the role and value of experiential knowledge in mental health research', *Disability & Society*, 32: 500-20.
Fenton, Andrew, and Tim Krahn. 2007. 'Autism, Neurodiversity, and Equality Beyond the 'Normal'', *Journal of Ethics in Mental Health*, 2: 1-6.
Fricker, Miranda. 2007. *Epistemic Injustice: Power and the Ethics of Knowing* (Oxford University Press: Oxford).
Glaser, Gabrielle. 2008. '`Mad Pride' Fights a Stigma', *The New York Times,* May 11 2008, section Fashion.
Hacking, Ian. 1999. *The Social Construction of What?* (Harvard University Press: Cambridge, MA).
Haslanger, Sally. 2004. 'What are we talking about? The Semantics and Politics of Social Kinds', *Hypatia*, 20: 10-26.
Hawkes, Nigel. 2011. 'Dangers of Research into Chronic Fatigue Syndrome', *BMJ*, 342.
Hay, Ryan. 2013. 'Hybrid Expressivism and the Analogy Between Pejoratives and Moral Language', *European Journal of Philosophy*, 21: 450-74.
Hengartner, Michael, and Sandrine Lehmann. 2017. 'Why Psychiatric Research Must Abandon Traditional Diagnostic Classification and Adopt a Fully Dimensional Scope: Two Solutions to a Persistent Problem', *Frontiers in Psychiatry*, 8.
Hom, Christopher. 2008. 'The Semantics of Racial Epithets', *Journal of Philosophy*, 105: 416-40.
Insel, Thomas. 2014. 'The NIMH Research Domain Criteria (RDoC) Project: Precision Medicine for Psychiatry', *American Journal of Psychiatry*, 171: 395-97.
Insel, Thomas, and Remi Quirion. 2005. 'Psychiatry as a Clinical Neuroscience Discipline', *JAMA*, 294: 2221-24.
Jackson, Frank. 1998. 'Reference and Description Revisited', *Philosophical Perspectives*, 12: 201-18.
Jeshion, Robin. 2013a. 'Expressivism and the Offensiveness of Slurs', *Philosophical Perspectives*, 27: 231-59.

———. 2013b. 'Slurs and Stereotypes', *Analytic Philosophy*, 54: 239-314.
———. 2016. 'Slur Creation, Bigotry Formation, and the Power of Expressivism', *Phenomenology and Mind*, 11: 130-39.
Jeste, Dilip, Barton Palmer, and Elyn Saks. 2017. 'Why We Need Positive Psychiatry for Schizophrenia and Other Psychotic Disorders', *Schizophrenia Bulletin*, 43: 227-29.
Kaminer, Wendy. 2015. 'The Progressive Ideas Behind the Lack of Free Speech on Campus', *Washington Post,* Feb 20 2015, section Op-Ed.
Keller, Simon. 2019. "Mental Health is Not Well-Being." In. unpublished manuscript.
Kripke, Saul. 1972. *Naming and Necessity* (Harvard University Press: Cambridge, MA).
Kukla, Rebecca. 2017. 'Infertility, epistemic risk, and disease definitions', *Synthese*.
Leslie, Sarah-Jane. 2015. "Hillary Clinton is the only man in the Obama administration': Dual Character Concepts, Generics, and Gender', *Analytic Philosophy*, 56: 111-41.
———. 2017. 'The Original Sin of Cognition: Fear, Prejudice, and Generalization', *Journal of Philosophy*, 114: 393-421.
Loughman, Amy, and Nick Haslam. 2018. 'Neuroscientific explanations and the stigma of mental disorder: a meta-analytic study', *Cognitive Research: Principles and Implications*, 3.
Lycan, William. 2015. 'Slurs and Lexical Presumption', *Language Sciences*, 51.
McWade, Brigit, Damian Milton, and Peter Beresford. 2014. 'Mad Studies and Neurodiversity: a Dialogue', *Disability & Society*, 30: 305-09.
Murphy, Dominic. 2001. 'Hacking's Reconciliation: Putting the Biological and Sociological Together in the Explanation of Mental Illness', *Philosophy of the Social Sciences*, 31: 139-62.
Nunberg, Geoffrey. 2018. 'The Social Life of Slurs.' in Daniel Fogal, Daniel Harris and Matt Moss (eds.), *New Work on Speech Acts* (Oxford University Press: Oxford).
Poole, Jennifer, Tania Jivraj, Araxi Arslanian, Kristen Bellows, Shiela Chiasson, Husnia Hakimy, Jessica Pasini, and Jenna Reid. 2012. 'Sanism, 'Mental Health', and Social Work/Education: A Review and Call to Action', *Intersectionalities: A Global Journal of Social Work Analysis, Research, Polity, and Practice*, 1: 20-36.
Pynn, Geoffrey. forthcoming. 'Testimonial Injustice and Epistemic Degradation.' in Jennifer Lackey (ed.), *New Essays in Applied Epistemology*.
Salvador-Carulla, Luis, and Marco Bertelli. 2008. "'Mental Retardation' or 'Intellectual Disability': Time for a Conceptual Change', *Psychopathology*, 41: 10-16.
Schalock, Robert. 2002. 'What's in a Name?', *Mental Retardation*, 40: 59-61.
Silvergate, Harvey. 2014. 'Liberals are Killing the Liberal Arts', *Wall Street Journal Online,* Nov 09 2014.
Siperstein, Gary, Sarah Pociask, and Melissa Collins. 2010. 'Sticks, Stones, and Stigma: A Study of Students' Use of the Derogatory Term "Retard"', *Intellectual and Developmental Disabilities*, 48: 126-34.
Szasz, Thomas S. 1960. 'The Myth of Mental Illness', *American Psychologist*, 15: 113-18.
Tekin, Serife, and Melissa Mosko. 2015. 'Hyponarrativity and Context-Specific Limitations of the DSM-5', *Public Affairs Quarterly*, 29: 109-34.
Thornicroft, Graham, Diana Rose, and Aliya Kassam. 2007. 'Discrimination in health care against people with mental illness', *International Review of Psychiatry*, 19: 113-22.
Tirrell, Lynne. 2017. 'Toxic Speech: Toward an Epidemiology of Discursive Harm', *Philosophical Topics*, 45: 139-61.
Väyrynen, Pekka. 2019. 'Thick Ethical Concepts', *Stanford Encyclopedia of Philosophy*.
Wakefield, Jerome. 2007. 'The concept of a mental disorder: diagnostic implications of the harmful dysfunction analysis', *World Psychiatry*, 6: 149-56.
Waldron, Jeremy. 2009. 'Dignity and Defamation: The Visibility of Hate', *Harvard Law Review*, 123: 1596-657.