Routledge
Taylor & Francis Group

Philosophical Explorations

An International Journal for the Philosophy of Mind and Action

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rpex20

Psychiatric fictionalism and narratives of responsibility

Sam Wilkinson

To cite this article: Sam Wilkinson (2022): Psychiatric fictionalism and narratives of responsibility, Philosophical Explorations, DOI: 10.1080/13869795.2022.2116473

To link to this article: https://doi.org/10.1080/13869795.2022.2116473

Published online: 02 Sep 2022.

PHILOSOPHICAL EXPLORATIONS
https://doi.org/10.1080/13869795.2022.2116473

Routledge
Taylor & Francis Group

Psychiatric fictionalism and narratives of responsibility

Sam Wilkinson

Sociology, Philosophy and Anthropology, University of Exeter, Exeter, UK

ABSTRACT

I explore the relationship between psychiatric fictionalism and the attribution of moral responsibility. My central claim is as follows. If one is a psychiatric fictionalist, one should also strongly consider being a fictionalist about responsibility. This results in the 'intrinsic view', namely, the view that mental illness does not just *happen* to interfere with moral responsibility: that interference is an intrinsic part of the narrative. I end by discussing three illustrative examples.

ARTICLE HISTORY
Received 11 May 2021
Accepted 17 June 2022

KEYWORDS
Fictionalism; psychiatric fictionalism; responsibility; exculpation; narratives

1. Introduction

One of the thorniest questions in philosophy of psychiatry is: When is someone who behaves in socially unacceptable ways afflicted by mental illness (and hence in need of treatment and compassion) and when is it a reflection of who they are (and hence potentially subject to moral evaluation)? A first step towards approaching this question, perhaps rather obviously, involves asking: What is mental illness? The hope has tended to be that an answer to this can shed light on when somebody is suffering from something that might explain her behaviour, and hence requires compassionate treatment, rather than punishment or moral disapproval.

Elsewhere (Wilkinson 2022), I have answered the question 'What is mental illness?' by appealing to fictionalism. Fictionalism has been applied to many domains, including ethics (Joyce 2001; Kalderon 2005), mathematics (Yablo 2002), and the mind (Toon 2016). The idea behind this *psychiatric* fictionalism is that when we engage in psychiatric discourse (most canonically the attribution of mental illness, but other derived forms that tacitly attribute mental illness), we are not attributing a robust, objective property, but are instead engaging with a useful, perhaps even an indispensible, fiction.

In this paper, I want to develop this further by exploring the relationship between psychiatric fictionalism and notions of moral responsibility. My central claim takes the form of a conditional, and it is as follows. If one is a psychiatric fictionalist, then one should also strongly consider being a fictionalist about responsibility.[1] Two things should be noted at this point. First, I don't need to argue for psychiatric fictionalism *per se* since I am mainly interested in the relationship between psychiatric fictionalism *if adopted* and fictionalism

CONTACT Sam Wilkinson S.Wilkinson@exeter.ac.uk Sociology, Philosophy and Anthropology, University of Exeter, St German's Road, Exeter EX4 4QJ, UK

about moral responsibility. Second, (as you will notice from my phrasing in terms of 'should strongly consider') I am not claiming that this relationship is one of strict entailment, but rather of a neat and plausible fitting together, that I hope to illustrate as we progress.

Before moving on, I need to clarify something at the outset. This proliferation of fictions and fictionalisms may strike some as rather extreme, and, worse, as making light of very serious matters. But this would be to misunderstand fictionalism: it does not imply that that these discourses, narratives and judgements don't matter. On the contrary, the postulated fictions are powerful and, it is precisely *because* there is flexibility, and indeed a degree of choice in the form that they take, that it is incumbent on us to make sure they reflect the kind of society that we want to live in.

I proceed as follows. First, I present psychiatric objectivism and how it has interacted with certain views about the nature of responsibility. The standard way of thinking about this (and this is central to much theory and practice in forensic psychiatry (see Eastman et al. 2012)) is as follows. Psychiatry tells us objective facts about a person's condition. In contexts of legal or moral investigation, these facts can become relevant when they shed light on the various ways in which the necessary conditions of moral (or legal) responsibility are interfered with. These conditions can be *circumscribed* and pertain to a particular action (e.g. the person was delusional and didn't know what she was doing), or *generic* and pertain to the capacity to be (fully) responsible for anything at all (e.g. this person cannot be expected to tell right from wrong). Usually (but not always) the former involves psychiatric disorders, whereas the latter usually (but not always) involves neurodivergent/neurodiverse conditions (psychopathy, intellectual disability, and so on). The difference between these two will become relevant later. In any case, I call this 'the incidental view', since mental illness (or another kind of condition) interferes *incidentally* with the conditions of moral responsibility.

Then I present psychiatric fictionalism, distinguishing it not only from objectivism, but also from expressivism (with which it has some important affinities). Then I show how, in contrast to psychiatric objectivism, psychiatric fictionalism accounts for (or should account for) diminished moral responsibility in psychiatric contexts. This is in terms of a negotiation of the boundaries between psychiatric and moral domains of discourse, which are in fact *fictions* or *narratives* in the relevant and very serious sense intended by fictionalists. I call this the 'intrinsic view', since a mental health condition doesn't just *happen* to interfere with moral responsibility: that interference is an intrinsic part of the narrative.

I end by illustrating this with three examples: (i) delusion and hallucination (namely, informational interferences), (ii) addiction (namely, motivational interferences) and (iii) ADHD, anti-social personality disorder and their relationship to naughtiness and youth offending respectively.

2. Psychiatric objectivism and the incidental view

I start by presenting the broad and mainstream position of psychiatric objectivism, and then show how it is taken to have a bearing on questions of moral (and by extension legal) responsibility.

2.1. Psychiatric objectivism

Psychiatric objectivism can be seen as a broad family of answers to the question 'What is mental illness?' Anyone who answers this by saying that it is an *objective* phenomenon subscribes to psychiatric objectivism, regardless of the finer details of their view. In other words, this includes any view according to which there is a fact of the matter about whether someone is suffering from mental illness and that this is independent of human perspective (whether individual, collective, societal). Crucially, this is not to be confused with facts *about* or *involving* human perspectives. In the relevant sense, it is an objective fact that I am not in excruciating pain at the moment. This is not the kind of perspective-dependence that we are talking about here. We are talking about whether human perspective brings a fact, or something fact-like, into existence, rather than about facts that, like facts about pain, have human perspective, human subjectivity, as their subject matter.

The standard path to psychiatric objectivism is via medical objectivism, namely, the claim that there are perspective-independent facts about illness and disease more generally, and to claim that this transfers unproblematically to the mental domain (it is this unproblematic transfer that Szasz famously takes issue with in his 'The Myth of Mental Illness' (1960)). Medical objectivism is supported by the intuition that, in a possible world devoid of humans (or indeed any theorising, or cognizant, beings), pathological processes take place, plants and animals get diseases, and have their lives negatively impacted upon by them and may even die from them. Psychiatric objectivism is based on the idea that something similar can be said about the mind and mental illness.

Some might be tempted to equate psychiatric objectivism with biomedical, or, even more stringently, bio-reductive approaches to mental illness, but this would be a mistake. The latter are certainly a species of psychiatric objectivism, but there are many non-reductive, even socially and environmentally distributed forms of objectivism (see Davies 2016, for example). According to such views, there is an objective fact of the matter about whether someone has a mental illness, but that fact obtains in virtue of, e.g. organism/environment coupling. Indeed, while this seems attractive for mental illness (the attraction being boosted, no doubt, by fashionable 4E approaches to cognition (Newen, De Bruin, and Gallagher 2018)), it is arguably just as applicable to somatic medicine (see Glackin 2017).

2.2. The incidental view and the requirements for responsibility

Psychiatric objectivism is a broad church, and I need it to be for the following application of it. Whatever the finer details of your preferred account of mental illness, the question arises: how does this interact with questions about moral responsibility?

The natural way of thinking about this from an objectivist perspective is to think about what requirements an agent needs to fulfil in order to count as morally responsible (both in general and for a specific action), and to reflect on how a given instance of psychiatric disorder may interfere with them. Note that these requirements for responsibility are presented quite independently of psychiatric disorder. After all, the nature of moral responsibility is a core philosophical question quite generally.

So, putting mental illness to one side for now, what are the requirements for responsibility more generally? I present perhaps the two most influential accounts: 'The Deep Self View', associated with Harry Frankfurt, and others, and 'The Sane Deep Self View', which is from Susan Wolf, and builds on the former in the face of a concern raised.

2.2.1. The deep self view

The Deep Self View starts with the observation that to deem someone morally responsible for an action is to attribute it to them. In classic examples of diminished responsibility, we aren't comfortable saying that the actions in question, as they appear on the surface, come from, or are a reflection of, the '*real them*'. As Will Cartwright (2006) puts it:

> The familiar excuses of accident, ignorance, coercion, and so on, are to be explained on this view as cases where the action does not reflect, and so cannot be attributed to, the agent's real self. (145)

This immediately raises the question: when *are* actions attributable to the *real* or *deep self*? Perhaps the best-known presentation of this is found in Harry Frankfurt's (1971) classic paper, *Freedom of the Will and the Concept of a Person* (although relevantly similar views are also found in the work of Garry Watson (1975) and Charles Taylor (1976)). What they all have in common is that, in order to be responsible agents, our actions cannot be the product of desires that are just found '*in* us, but expressions of characters that come *from* us, or that, at any rate are acknowledged and affirmed *by* us' (Wolf 1987, 365).

As Frankfurt puts it:

> Besides wanting and choosing and being moved to do this or that, [persons] may also want to have (or not to have) certain desires and motives. They are capable of wanting to be different, in their preferences and purposes, from what they are. (1971, 7)

A cornerstone of Frankfurt's position is this introduction of a hierarchy of motivational states. Human beings have the capacity to not only want certain things, but also to want to want certain things. For example, I might want to smoke a cigarette, but want to not want to, because I'm trying to quit. Conversely, I might not have an appreciation of, and hence appetite for, classical music, but wish that I did have such an appreciation, because I feel that it might enrich my life.

So, here is a clear statement of the Deep Self View:

> An agent is fully morally responsible for an action if it comes from their Deep Self, which means that the first-order desires, that proximally give rise to the action are endorsed with second-order desires.

With this distinction between first-order and second-order motivational states, we get a clear account of how mental health conditions might interfere with the full and proper attribution of moral responsibility. In particular we can distinguish motivational interferences, like addiction, or depression, or social anxiety, from informational interferences, like delusion and hallucination.

A gambling addict has a very strong first-order desire to gamble, but has a competing second-order desire: a desire to not desire to gamble, a desire ultimately to not be a gambling addict and all of the negative consequences that ensue. If they steal in order to

gamble, that is their first-order desire winning out over their second-order desire. They didn't really want to steal 'deep down'. Similarly, the depressed or socially anxious person wants to want to go to the party, but is too depressed or anxious, and so at a primitive first-order level, doesn't want to go to the party, even though, 'deep down' they want to go, they want the benefits that party-going affords.

With informational interferences, like delusion or hallucination, the account is quite different, but still works. If the desire to commit the crime was endorsed by a second-order desire, it was only a desire *de re*. For example, the delusional individual who kills her husband because she thinks he is a malevolent android has not endorsed the desire *to kill her husband*, but rather to protect herself and her family from a malevolent android.

So far, we've looked at circumscribed cases of diminished responsibility pertaining to relevant interferences to specific actions, but the Deep Self View also has the resources to explain a general lack of capacity for responsibility. Just as Frankfurt highlights that non-human animals may lack 'the capacity for reflective self-evaluation that is manifested in the formation of second-order desires' (1971, 7), the same might be said of some humans. It would be an empirical issue as to precisely where this line may be drawn, but obvious candidates are young children and those with severe intellectual disabilities.

2.2.2. The sane deep self view

In presenting the 'Sane Deep Self View' (Wolf 1987), Susan Wolf takes the Deep Self View to be too strict, namely, to attribute the full force of moral responsibility to cases where it should not apply. In particular, there are cases where there is second-order endorsement of action, but we still feel the pull of exculpatory intuitions. Wolf illustrates this with the well-known case of JoJo:

> JoJo is the favorite son of Jo the First, an evil and sadistic dictator of a small, undeveloped country. Because of his father's special feelings for the boy, JoJo is given a special education and is allowed to accompany his father and observe his daily routine. In light of this treatment, it is not surprising that little JoJo takes his father as a role model and develops values very much like Dad's. As an adult, he does many of the same sorts of things his father did, including sending people to prison or to death or to torture chambers on the basis of whim. He is not coerced to do these things, he acts according to his own desires. Moreover, these are desires he wholly wants to have. When he steps back and asks, "Do I really want to be this sort of person?" his answer is resoundingly "Yes," for this way of life expresses a crazy sort of power that forms part of his deepest ideal. (Wolf 1987, 367–368)

Wolf goes on to claim that our intuitions pull us in the direction of denying that JoJo is responsible for his actions. This is because it 'is unclear whether anyone with a childhood such as his could have developed into anything but the twisted and perverse sort of person that he has become' (Wolf 1987, 368). She rehearses two explanations for this reduced responsibility – the first she rejects in favour of the second.

The first is that JoJo had no say in who his Deep Self might become. Quite rightly, Wolf dismisses this explanation: viewed 'from the outside' none of us have! Indeed, as Wolf memorably puts it, 'self-creation is not just empirically but logically impossible' (368). The second explanation is in terms of what Wolf calls 'sanity' (in an admittedly 'special sense' (369, 370)). An agent is 'sane', in this sense, if she is able 'cognitively and normatively to understand and appreciate the world for what it is' (1987, 387).

This notion of understanding the world for what it is both cognitively *and normatively*, opens up the possibility that two forms of exculpation are the result of agents 'not knowing what they do', but in different senses. Some agents don't know what they are doing in a factual sense, either in a direct sense like in delusion (or more short-lived delirium), or because they are ignorant of the true nature or significance of their actions (like Oedipus not knowing that Jocasta was his mother), or because they simply don't appreciate the consequences of their actions. Other agents don't know what they are doing in the normative sense, namely, they know exactly what they are doing, cognitively, but don't know that it is *wrong*. So, JoJo is let off the hook, so to speak, because he is 'insane', in the sense of not knowing right from wrong. To paraphrase Wolf, JoJo's actions are driven by *mistaken values*.

Regardless of whether this account appeals, note that it is firmly grounded in a realist meta-ethics. It presupposes that there are objective moral facts that we can either know and understand, or fail to know and understand (like JoJo). In contrast, however, many meta-ethicists (namely, any proponent of any of the many forms of moral anti-realism) would balk at the idea that you can intelligibly talk of 'appreciating normatively the world for what it is'. The reason this is relevant for our purposes is as follows: the sorts of considerations that can lead one away from objectivist psychiatry are somewhat similar to those that lead one away from realist meta-ethics. The possibility, and plausibility, of this double moving away from both objectivist psychiatry and realist meta-ethics is what we explore in the rest of the paper.

3. Psychiatric fictionalism

I present psychiatric fictionalism in the following way. I present two ways of opposing objectivism. The first amounts to psychiatric expressivism. Then I find problems with expressivism, which then get addressed through the adoption of fictionalism instead. Since this version of fictionalism retains some of the motivational aspects of expressivism, I call it motivational fictionalism.

3.1. Two ways of opposing objectivism

Since my central claim is of a conditional form, I don't need to argue convincingly for a rejection of objectivism. However, it is important to illustrate the content of these views, what these views look like, as it were, via their opposition to objectivism. If you are a psychiatric objectivist, you adhere to (at least) the following two claims, one linguistic, and the other metaphysical.

(1) Discourse about mental illness is fact-stating.
(2) Those facts about mental illness obtain objectively.

Let us clarify each in turn. What (1) means is that claims about mental illness are in the business of describing – of making factual statements about – the world. Thus, saying 'James has a mental illness' is like saying 'James has blue eyes' in the following sense. It is trying to describe the world, and is true if and only if James really has a mental illness or has blue eyes. One standard way of thinking about this, which will

serve our purposes, is that fact-stating assertions that *p*, express (or purport to express) beliefs that *p*.[2]

In order to better understand what it means for discourse to be fact-stating, it is helpful to reflect briefly on discourse that is not.[3] Some utterances, rather obviously, don't express (or purport to express) beliefs. Requests, for examples, express desires ('Please pass the salt' expresses something like the desire to have the salt, or for you to pass it to me), promises, for example, might express committed intentions ('I promise I'll mow the lawn tomorrow'), exclamations of 'Ouch!' might express pain, 'Yuck!' might express disgust. Claim (1) is in part saying that statements about mental illness are not like these utterances. They are, as indeed there appear to be, like straightforward, belief-expressing assertions like 'James is 6ft tall'.

Claim (2) can be seen as building on (1) in that, not only are these statements expressions (or purported expressions) of beliefs; these statements and corresponding beliefs can be true, and what makes them true is the obtaining of objective, perspective-independent, facts. The two ways of not being an objectivist, in the relevant sense, about a domain of discourse, are remarkably simple. One way is to deny (1): the domain of discourse may *look* fact-stating, but it's actually not. This does not logically *entail*, but is often motivated by, a denial of (2). In particular, denying (1) obviates the need to give account of facts that the discourse seems to be referring to, since it is not fact stating discourse after all.

The other way is to *accept* (1), but focus on denying (2): the discourse is fact stating (or at least purportedly), but that discourse is not true in virtue of facts that obtain objectively. This could be either because it's not true at all, or is true in some sense, but not true *objectively*. To simplify, let's call the first strategy (denying both 1 and 2) *expressivism*, and the second strategy (accepting 1, but denying 2) *fictionalism*. Both expressivism and fictionalism have significant pedigree when it comes to ethics (see, Ayer 1952; Blackburn 1998, etc., for expressivist views, and Joyce 2001; Kalderon 2005, for fictionalist views). What I think the most promising position is, at least when it comes to psychiatry, is a sort of expressivism/fictionalism hybrid; what we might call *motivational fictionalism*. Since the clearest path to this is via criticism of certain aspects of expressivism, let's start with that.

3.2. Exploring psychiatric expressivism

Psychiatric expressivism takes psychiatric discourse, most canonically, mental illness attributions, in spite of surface appearances to the contrary, to not be in the business of stating facts, where this can be unpacked in terms of being expressions of something other than belief (Hare 1986 holds this view for illness and health in general, and Fulford 1989 applies this to psychiatry more specifically). Why would anyone hold this view, and what's wrong with it?

3.2.1. Motivations for psychiatric expressivism

Expressivism is best known as a position in meta-ethics, namely, about the nature of ethical discourse. Interestingly, the three main motivations for expressivism about psychiatry mirror those for expressivism about ethics. They are; (i) ontological unease, (ii)

arguments from disagreement and (iii) arguments from motivation. Let's take these in turn.

What I'm calling 'ontological unease' is a reluctance to posit objective 'mental illness properties' in a way that is somewhat analogous to how expressivists in meta-ethics are reluctant to posit moral properties of 'rightness' or 'wrongness'.[4] It is worth mentioning that this somewhat begs the question of the objectivist, since the whole point of objectivism is to tell you what mental illness properties are! But the expressivists' point is not that objectivism happens to fail, but that its endeavour is fundamentally misguided. We can explain our discourse and behaviour without appeal to these properties that we seem to posit in daily life. The next two motivations for expressivism don't beg the question in this way. What's more, I'm not in the business of convincingly criticising objectivism here, in any case.

With 'the argument from disagreement' there is again a useful parallel with the moral case. When people disagree morally, they can be in agreement about all of the relevant facts concerning a given case. The disagreement must be down to something else, namely, a conflict of *values*. Expressivists take this to be evidence that moral judgment, at its core, is not factual judgment, and that correlated moral claims are not factual claims (they express the sorts of states that reflect our values, like emotions, prescriptions or pro-attitudes). The same can be said about the cases of psychiatric disagreement where the facts are agreed, but there might be disagreement about whether something counts as a mental illness. In psychiatry, we can especially think about changes in views over time as simply a form of diachronic disagreement. For example, few people, and certainly not the Western psychiatric establishment, these days view homosexuality as a mental disorder. And yet it was only removed from the DSM in 1973. This removal, although it may have been informed by some reflection on the facts of the matter, is not fundamentally a factual move: it signals a change in values.[5]

Finally, there is 'the argument from motivation', where, yet again, comparison with the moral case is useful. Built into the judgment that 'Stealing is wrong', is a (*ceteris paribus*) reluctance to steal. This could only make sense if the mental state that the assertion expresses (namely, reveals you as having) is not a straightforward factual belief, since (at least granting a Humean account of motivation) beliefs do not motivate in and of themselves. Even the belief that I'm about to be hit by a train will only motivate me to get off the tracks if it is paired with the desire not to die. Psychiatric judgments, the expressivists might argue, have similar built-in motivational force. On such a view, deeming someone to be mentally ill enjoins society to help said person, rather than, say, punish them.

3.2.2. Problems with psychiatric expressivism

In spite of being an attractive and interesting position, there are a number of problems with psychiatric expressivism. I'll focus on two. These challenges are as follows. First, psychiatric discourse just doesn't seem to express non-beliefs: desires, emotions, 'hot' intrinsically motivating states. Unlike ethics, psychiatry seems cold and institutionally regimented. The second is related: is psychiatric discourse really *intrinsically* motivating, in anything like the way ethics is? Let us look at these more closely.

Our reactions to things we deem to be morally wrong are emotionally inflected. Indeed, we might even be tempted to say that someone who doesn't have the usual

emotional reactions to things (e.g. think about the canonical representation of a psychopath) cannot make authentic moral judgments (e.g. Shoemaker 2011).

It is this sort of intuition about moral thinking as emotion-laden that led Ayer to come up with his emotivist theory. Although in its original form it has few, if any, adherents, subsequent versions of expressivism can be seen as modifications to that core emotivist insight. This raises the question: is a similar account of psychiatric discourse even remotely plausible? Another related concern is that, whereas there are pages and pages of scientifically-grounded psychiatric theorising, some of which gets canonised in manuals like the DSM and ICD (not to mention numerous textbooks for medical students), the same cannot be said for ethics. In short, even if we grant that there may be expressive elements to psychiatric discourse (as I have argued elsewhere (Wilkinson 2020), the attribution of delusion is least partially expressive) that can't be the whole story.

Furthermore, it seems like we might here be committing what Kalderon (2005) calls the 'pragmatic fallacy'. Whereas Kalderon charges meta-ethical expressivists of committing this fallacy, it is even more clear in the psychiatric case. The pragmatic fallacy involves an unwarranted jump from uses to meanings. In other words, just because a term can be used expressively, doesn't mean it has an expressivist semantics. We might talk about things being 'sick' or 'crazy' in a deeply expressive way, but that doesn't consign the whole of medical or psychiatric discourse to an anti-descriptive, expressive semantics.

The second problem is to do with intrinsic motivation. Ethical norms arguably function primarily to regulate behaviour. As a result, the idea that there is an intrinsic motivational force to ethical thought is very plausible. If psychiatry is to be thought about along these lines it needs, at best, some serious modification. For a start, it is not obvious that psychiatry is fundamentally about regulating behaviour: it may have behaviour regulating effects (indeed, these are well-documented, and not just within the anti-psychiatry movement) but that's a different matter. Perhaps the most plausible path for a psychiatric expressivist to take would be to say that mental illness attribution enjoins others to help the individual deemed ill. The big question is, is this *constitutive* of mental illness discourse, or merely a causal feature of it? For this to work as a distinctly expressivist argument, it needs to be the former, and that might be an implausible claim to make. This is especially so, given that the fictionalist can account for the motivational aspects of psychiatric discourse in causal, non-intrinsic ways, both in terms of what motivated us to adopt the fictions in the first place, and the motivational roles that they play once adopted.

3.3. From expressivism to (motivational) fictionalism

Remember the two ways of opposing objectivism. The fictionalist accepts (1), but rejects (2). In other words, the fictionalist accepts that psychiatric discourse is (purportedly) fact-stating, but denies that these facts obtain objectively. In other words, psychiatric statements (attribution of mental illness) are not straightforwardly true, but it misses something important to dismiss them as straightforwardly false (or meaningless). To accommodate this idea, fictionalism will take psychiatric discourse to be akin to a useful, or perhaps even indispensable, fiction. Why hold this view, and what's good about it?

3.3.1. Motivations for psychiatric fictionalism

One way of presenting the motivations for fictionalism is in terms of its differences with, and advantages over, *expressivism*. Fictionalism shares with expressivism an ontological unease with countenancing objective facts pertaining to the domain in question (in this case mental illness). However, the appearance of fact-stating discourse (e.g. the fact that people sincerely attribute mental illness to people) will be accounted for differently. It is not that what is being expressed is, contrary to appearances, not really belief, but rather that, although belief (or something like it) is involved, the reality that is being engaged with is in some sense fictional. There are different ways of unpacking this idea, and they yield different forms of fictionalism, but, first, let's look at why we might be broadly tempted towards fictionalism in general, as opposed to expressivism.

When we engage with fictions, these fictions may *trigger* emotions, and 'hot' cognition more generally, but it is not *fundamentally* an emotional enterprise: it is very cognitive, and this might fit nicely with psychiatric discourse. It can be complex, detailed and discursive, and while some of it may be causally emotive, it is not intrinsically so.

In a similar vein, fictionalism might provide a welcome corrective to the strong motivational internalism that the expressivists need to support their argument. Yes, fictions can be used to elicit various motives, but it is not intrinsic to their meaning. Fictions need to be understood first, and then reacted to, whereas the expressivist picture is one where to understand *just is* to appropriately react. Another way of putting this is that fictions are *causally* motivational, rather than *constitutively* motivational.

Another place where the fictionalist may agree with the expressivist, and unsurprisingly so since it is derived from the point about ontological unease (anti-realism) about objective psychiatric facts, concerns the argument from disagreement. Both fictionalists and expressivists would hold that there could be cases where all of the facts of a case are nailed down and agreed upon, but there is disagreement about whether it counts as pathological. However, how this is explained will be different. The expressivist, as we've seen, will claim that the difference is one about values as reflected in non-cognitive reactive attitudes to the case in point. The fictionalist will instead claim that the disagreement is about two different, but superficially similar, fictions. Again, the diachronic version of this, exemplified by the removal of homosexuality from the DSM (and similar changes, like the punishable cowardice of shell shock giving way to the pitiable illness of PTSD) reveals a difference: the expressivist will say that it's about our reactive attitudes changing, the fictionalist will claim instead that it's a change in our fictions and narratives (which in turn engender a change in our reactive attitudes).

3.3.2. Varieties of fictionalism

So far, I have spoken about fictionalism in the broadest possible terms, but, of course, it comes in a number of different varieties, within the constraints of what I have so far presented. These constraints are as follows: the discourse that is a candidate for the fictionalist treatment is purportedly fact-stating (i.e. it is representational and cognitive, not expressive) and although the facts don't robustly, objectively obtain, it misses something very important to claim that the discourse is routinely false or mistaken.

The first distinction to be made is between hermeneutic and revolutionary fictionalism. Hermeneutic fictionalism constitutes a descriptive claim about the actual nature of a

domain of discourse. Revolutionary fictionalism, in contrast, is a normative claim about what a domain of discourse ought to be like, that it *ought* to be fictional in the relevant sense. We are interested in the former.

Within hermeneutic fictionalism, there is the distinction between content fictionalism and attitude (or force) fictionalism (for a similar, or perhaps identical, distinction see prefix-fictionalism vs. pretence-fictionalism (Toon 2016)). The content fictionalist locates the fictional operator in the content. So the attitude (or force, at the level of discourse) is belief (or assertion), but the content that is believed (asserted) is fictional. In contrast, force fictionalism retains the content, but the attitude towards that content is something that falls short of belief, so it might be (following Toon 2016) pretence, or pretend belief (or pretend assertion).

This difference is nicely illustrated with an example of actual discourse about fiction. Suppose I say: 'Sherlock Holmes lives at 221b Baker Street'. The content fictionalist will say that this is an assertion, but there is an implicit constituent (a 'prefix') '[In the Conan Doyle Stories] Sherlock Holmes lives at 221b Baker Street'. The force fictionalist will say that it's not really assertion, it's something else, like a pretend assertion.

One thing that distinguishes between these two kinds of fictionalism is whether what we are talking about is an assertion or not. One way of thinking about this is whether, if sincere, it expresses *belief*. The force fictionalist will claim that, whatever the apparent assertion is doing, it is not expressing belief, whereas the content fictionalist will. So, any domain of discourse that seems belief-expressing will, at first blink, seem to favour content fictionalism. However, another distinction between the two is that the content fictionalist needs an account of the modified content, the prefix as it were, whereas the force fictionalist doesn't. In the Sherlock Holmes example, content fictionalism has an easy task: we have the Conan Doyle stories. But for the domains of discourse that the fictionalist is interested in giving an account of, what plays the equivalent role of prefix 'In the Conan Doyle stories'? For moral fictionalism, or mental fictionalism, there is no explicit fiction to be appealed to, and so many fictionalists in these domains have tended to opt for force or pretense fictionalism (see, Toon 2016, for example). I think this emphasis on explicitly shared fictions is a mistake, and overlooks the extent to which narratives are co-constructed in a socially distributed manner.

In any case, my sense is that psychiatry has a more explicit, codified 'fiction' than these domains, via various institutionalised practices (let alone the DSM and ICD). Furthermore, we take these fictions very seriously, and act on them as we would if they were beliefs, so I'd be strongly tempted to say that they *are* beliefs. So, I would suggest that an attractive form of psychiatric fictionalism is hermeneutic content fictionalism. However, I am also a motivational fictionalist. This means that the fictions emerge because of how they *motivate* society to behave or react to things that are described in terms of those fictions. Note that this does not simply collapse into expressivism because these fictions are causally, not intrinsically motivating (and this is related to the relevant mental states being beliefs rather than desire-like states).

4. Psychiatric fictionalism and the instrinsic view

I would like to start my presentation of the relationship between psychiatric fictionalism and moral responsibility, with the following quote from Craig Edwards:

> Mental illness is not a label that picks out a set of consistent qualities in a mental condition; it is a label that stipulates how people *should* respond to the condition, and in particular whether they should respond *morally* or *medically* (Edwards 2009, 75)

This strikes me as an insightful remark (and similar ones have been made by others) and, under a particular interpretation, it says exactly what I want to say, and brings together precisely the two elements that I want to unite here. First ('not a label that picks out a set of consistent qualities') there is a natural interpretation according to which this is a statement of opposition to psychiatric objectivism (specifically realism). Second ('it is a label that stipulates how people should respond') we have something that, though at first glance looks like expressivism, more closely resembles my proposed motivational fictionalism. The reason for this is that, the emphasis is on the label itself, something public and external, rather than on the mental state that a speaker is in. What is more, the motivational aspect is causal, and strengthened by social convention, but not constitutive: i.e. it is about how people should respond, rather than how they *have to*, if the discourse is to count as the kind of discourse that it is (as goes the argument for expressivism). Third, the opposition of '*morally* or *medically*' is exactly what I am looking to explore with the intrinsic view. Mental illness attribution, on this view, is like drawing a boundary. As you increase the psychiatric territory, you *ipso facto* shrink the moral territory.

Let me unpack this view a bit. First, I want to introduce the intuitive idea that certain labels and narratives have certain *pragmatic profiles*, namely, they have stable tendencies to elicit particular responses or courses of action on the part of societies, their institutions and individuals within these societies and institutions, including, importantly, those individuals themselves who are being labelled (i.e. there are self-directed responses to being labelled – indeed these are sometimes the most important since they often have the most direct impact on positive outcomes). Mental illness discourse (explicit attribution of mental illness, and implicit presupposition of mental illness) has a particular pragmatic profile. These profiles are:

(1) Variable
(2) Malleable
(3) Exclusive (i.e. defining) in whole
(4) Non-exclusive (i.e. non-defining) in part

Variability (1) means that different people, in different contexts, react differently. This doesn't imply that the discourse is in any way faulty. It's just a fact that emerges from the variability across individuals within human societies, and the way in which discourses and narratives function in those societies. They regiment, but not with total uniformity, human reactions. Although there will be case-by-case variation, there will be a peak in the probability-density distribution, a centre of gravity, as it were, that represents the 'collective response'.

Malleability (2) means that over time, societal attitudes can change, or the fictions themselves can change, thereby shifting the pragmatic profile (moving the peak of the distribution, the collective response). This is simply the basic idea that societal change is possible. Point 3 means that, taken as a whole, the pragmatic profile of the discourse

is what defines it. Something cannot have exactly the same pragmatic profile and yet somehow fail to be that discourse. If it looks like (e.g.) psychiatric discourse, sounds like psychiatric discourse, acts like psychiatric discourse, then it *is* psychiatric discourse. However, 4 clarifies that there will be overlap between parts of the pragmatic profiles of different discourses, for example, between illness, disability, divergeance, etc.

This observation, if it is correct, says nothing about how this negotiation between the moral and medical normative realms plays out, nor does it explicitly say anything about the metaphysical or epistemic status of these negotiations. A realist ontology, which would be antithetical to fictionalism, is logically compatible with it. According to such a view, you discover that someone objectively has a mental illness, and that always interferes with the (again, objective) conditions of responsibility. Though a coherent position, it seems like it would be an unlikely coincidence that the medical and moral happen to align in this way. A much neater and less implausibly coincidental explanation is that neither moral nor medical narratives are robustly objective, that they interact with each other, and are left with plenty of 'wiggle room' once the relevant facts are settled. This is not to say that people are never truly responsible for their actions, at least to the extent that that there are many cases that we can all agree upon. But this is a result of our robust societal alignment, rather than objective conditions of responsibility (a 'set of consistent qualities', to echo Edwards) being met or failing to be met. This will become much clearer by looking at some examples.

5. Examples

I want to illustrate the position with three examples that are standardly thought to involve a degree of diminished responsibility: (i) delusion and hallucination (namely, informational interferences), (ii) addiction (namely, motivational interference) (iii) ADHD, antisocial personality disorder and their relationship to narratives of naughtiness and youth offending respectively.

5.1. Delusion and hallucination

The incidental view, and the objectivism on which it is built, works especially well for cases like this, and these are indeed paradigmatic cases of diminished responsibility. If someone has a radically inaccurate take on the world, they cannot be held responsible for performing an action that they were not aware they were performing. The inaccuracies of the world-view seem to be an objective fact: either somebody took the world to be a way it's not, or they didn't. Either I knew that my spouse was my spouse, or I was deeply panicked and confused and convinced that they were an evil and malevolent android. In the latter situation, I cannot be held responsible for the action described as 'Harming my spouse' since when I harmed my spouse, I was convinced that it wasn't my spouse that I was harming.

If this fits the objectivist incidental view so well, then how does it fit the fictionalist intrinsic view? And, by inference to the best account, if it doesn't fit so well, isn't this bad news for the fictionalist?

The first thing to point out is that, when delusions and hallucinations exculpate, they exculpate in virtue of their epistemic properties, not their psychiatric properties. Indeed, if

somebody without any psychiatric issues acts in profound ignorance or confusion, they are similarly exculpated. In a sense, psychiatric status is a red herring, and so this is not relevant to psychiatric fictionalism: it is simply a feature of how we individuate actions, whether this is under psychiatric influences or something else.

In addition to being misinformed, however, this individual may feel the force of moral and legal disapproval if there are forms of epistemic negligence: namely, they have failed to know something that they should have known. This becomes more relevant to the fictionalist position since it seems like attributing a psychiatric condition makes us (quite rightly) view the person as a victim of, rather than an agent in, their state of misinformation.

Indeed, as I've argued elsewhere (Wilkinson 2020), this is related to the understandability of particular beliefs and epistemic positions, and to the extent to which we are tempted to deem these problematic. For example, someone might be highly evidence-resistant for very understandable reasons. Consider a mother whose son is being charged with murder, and has evidence mounting against him. Is her resistance to this evidence delusional? No. She doesn't elicit in us the folk-epistemic bafflement that is a hallmark of delusion (and I would say constitutive of delusion attribution (Wilkinson 2020)). Is she mentally ill? Again: No. She is not, because she is not a victim of her irrationality: she is the agent of her own evidence-resistance, which is not to say that she controls it, but rather that it comes *from* her. It is not happening *in spite of* her. Having said that, we do not blame her for it either, but not because there is anything wrong with her medically: indeed, there is nothing wrong with her at all. She is entirely normal, both morally and medically. (Indeed, it is excessive rationality and impartiality in such a condition that would raise alarm in us.)

So, delusions and hallucinations involve two levels of diminished responsibility. One involves acting in ignorance, which is not specific to a psychiatric condition at all. The second involves deeming the person to lack autonomy, to be a victim of their condition, rather than an agent in, their inaccuracies. And this, I think, is best understood within fictionalism, as a narrative construction. This is not to say that anything goes: some narratives are more appropriate than others, but it is not a brute fact that someone is or isn't responsible for their irrationality.

5.2. Addiction

How much autonomy do we attribute to an addict? My view is that there are relevant facts about an addict's state, but they inform, rather than determine, the choices we make about autonomy.

A contemporary classic in this literature, which I take to be highly compatible with the framework I am putting forward here, is Hanna Pickard's 'Responsibility without blame' proposal which is found in various papers (Pickard 2011, 2013), including 'Responsibility without blame for addiction' (Pickard 2017).

In the latter she begins by contrasting two different ways of approaching addiction. One, which she calls *the moral model*, is broken down into two parts: (1) the claim that drug use is a choice, and (2) a critical moral stance towards that choice and the person who has made that choice. In response to this, *the disease model*, which denies both (1) and (2) has typically been presented as a welcome, and scientifically grounded, ideological corrective (see Leshner 1997).

However, Pickard argues, it's not a good model *either*. This is because it removes the autonomy and responsibility of the addict. According to the disease model, drug use is, for the addict, no longer a choice. It is an involuntary compulsion. But this seems neither accurate, nor helpful. As Pickard (2017) puts it: 'self-conceiving as a helpless victim of a disease […] risks placing addicts in a position whereby they view themselves as dependent on medical and associated professionals for a cure' (171). No, what we need to instil in addicts for therapeutic success is 'a sense of agency and empowerment, alongside the fashioning and enacting of a life narrative that makes sense of the past while telling the story of a different future' (Pickard 2017, 171). Pickard suggests that this involves treading a middle ground between the moral model and the disease model. But how do we do this?

Drawing on her own clinical experience in a Therapeutic Community, Pickard suggests that the optimal stance to take towards addicts is one of *responsibility without blame*. In effect, we need to accept the first part of the moral model (the choice part), but reject the second (the moral disapproval part).

Why does this apparently simple solution seem like a conundrum? (As Pickard herself admits: 'I initially had no idea how this stance was conceptually possible, let alone achievable for myself within my own clinical practice' (174).) It is because there is a deep-rooted, but ultimately misguided, tendency to tie responsibility and blame together. A particularly strong expression of this is to be found in the quote by Gary Watson (quoted in Pickard 2017, 174–175): 'to regard oneself or another as responsible just is the proneness to react to them in these kinds of ways'. However, Pickard contends, the judgement that someone is responsible for something and 'our practice of responding to […that someone] with what is in effect an affective form of blame – a set of hostile feelings typically accompanied by equally hostile thoughts and actions' ought, both in principle and at times in practice, to come apart.

Addicts, in clinical contexts, are most usefully thought of as responsible, (after all, 'people will only change what they believe is in their power to change' (175)) but the negative affective response of blame is both unhelpful and inappropriate. What is more, this separation of responsibility and blame gives rise to positive societal consequences. Or rather, the tying of responsibility and blame together, forces us to adopt one of two extremes, either the moral model, or the disease model.

> [And,] in placing blame squarely on addicts or their disease respectively, both models are united in enabling us to keep the focus of our attention away from ourselves and our society, avoiding the question of whether we, as a society, also collectively bear some responsibility for drug use and addiction and their consequent harms. (Pickard 2017, 177)

This, when we reflect on the sort of societal problems that give rise to addiction, is a very welcome corrective.

Ultimately, from the perspective of what I am suggesting here, to adopt this framework, this attribution of responsibility without blame, is not to discover a particular fact about addicts, but rather to *choose* a particular narrative, one that is enabled by the facts about addicts (you can't overlay any narrative onto any set of facts), but which is not mandated by it. Its adoption is motivated by recognition that it promotes outcomes that are helpful and humane.

5.3. ADHD, anti-social personality disorder, naughtiness and youth offending

The diagnostic category of 'Attention Deficit Hyperactivity Disorder' (ADHD) has come under significant criticism on charges of (often financially-motivated) medicalisation and diagnostic creep (see, e.g. Kazda et al. 2021). I'm not interested in adding my voice to these criticisms here, although I am broadly sympathetic with them. What interests me here is the kind of negotiations that occur between different kinds of discourses.

Let us grant that there are canonical cases of ADHD. However, let us also grant that there are borderline cases, where a behaviourally problematic child is on what looks like an indeterminate boundary between being deemed 'naughty' (or similar), and getting an ADHD diagnosis. Now, the psychiatric objectivist (who is also on board with the legitimacy of ADHD diagnosis, which, of course, an objectivist might not be) will take the applicability of the diagnosis to rest on a particular fact. Thus we might ascertain the true status of the child by discovering more scientific facts about the child. The fictionalist will say that there might be something to be gained from that, but, ultimately, the classification is a societal decision.

From the fictionalist perspective, this becomes a hugely rich area for reflecting on different pragmatic profiles. I can't do justice to this here, but consider the concept of naughtiness. This is a kind of proto-moral attribution: naughtiness attributes responsibility to the child, but not full responsibility.[6] The naughty child lacks full moral understanding, not in the objective, factual sense of Wolf (1987), but in the social sense that they are not yet fully-fledged partakers of social norms, of the economy of morality. However, because we are also trying to encourage them towards that point, calling a child naughty serves to tell the child that the way that they behaved was not okay, that they can control their behaviour and do better next time. Naughtiness partially implicates the child (but with a view to a forward-looking, corrective perspective), but it also partially implicates the parent(s), and indeed society as a whole. We all, to some extent, have responsibility to shape the minors of our society into 'responsible adults'.

At its extremes, challenging behaviour among young people can cross into the legal territory of 'youth offending' (what was formerly called 'juvenile delinquency'), and again, this is an interesting concept to reflect on. The individual doesn't have full legal responsibility, but has a high enough degree of it that some punitive rather than merely rehabilitative measures are taken. However, the emphasis remains squarely on rehabilitation. Something is wrong with you, and it needs to be fixed. But it is not something *medically* wrong with you. And yet, there are behaviourally problematic minors who are placed in a medical category, in particular the diagnostic category of personality disorders, most typically anti-social personality disorder. Such a diagnosis immediately shifts the societal response towards this individual, decreasing the focus on punishment, increasing the focus on rehabilitation and changing the kind of rehabilitation at play: disciplinary issues (partially) give way to medical ones, the child psychologist gives way to the child psychiatrist, and so on.

In light of this, what does a medical diagnosis achieve, whether it is the naughty or inattentive child at school getting a diagnosis of ADHD, or a more problematic child getting a diagnosis of anti-social personality disorder? It doesn't only say: you need to be treated, not just helped (and certainly not punished). It also absolves, and not only the children themselves but also parents, teachers, society at large, of responsibility

(which, for the reasons outlined above, by Pickard (2017) may turn out to be a demotivating double-edged sword). To repeat and repurpose Pickard's point it enables 'us to keep the focus of our attention away from ourselves and our society, avoiding the question of whether we, as a society, also collectively bear some responsibility' (177).

Just as with the first two examples, the narratives about illness and disorder serve to encroach on narratives of responsibility. We end up with an 'intrinsic view' of the relationship between mental illness (institutionally branded as 'psychiatric conditions') and moral responsibility, and this relationship is neatly explained by a fictionalism about both domains. One difference between the informational and motivational interferences of the first two examples and this third example is that the fictions here attach to a certain category of person, rather than to a particular action. This tracks the two different uses of 'responsibility', and two different modes of exculpation, where one applies to the short-term circumstances of a particular action, whereas the other applies to a general long-term capacities to be 'responsible' for any action.

6. Conclusion

What I have managed to say here is rather preliminary, and there are several objections that I haven't had time to rehearse and respond to. However, I have attempted to present a view about the relationship between mental illness and moral responsibility, according to which attributing mental illness (which is usually tacit, rather than explicit) is intrinsically (not incidentally) about the removal of responsibility. In other words, the normative discourses of psychiatry and morality interact. Furthermore, this can most clearly be accounted for by being a fictionalist about both.

Recognising this fact, if it is indeed a fact, matters since we need to work to make our fictions and narratives reflect and help to create the world we want to live in. As the expressivist (and quasi-realist) Simon Blackburn puts it:

> What we need to do is to make our responses mature, imaginative, cultured, sympathetic, and coherent, and we can accept what help we can from people who have thought more deeply about human life – people who have climbed further up the mountain. (Blackburn 1998, 310)

The only thing I would add is that these responses are societal, and shaped by and reflected in the narratives and fictions that we, most often unknowingly, weave.

Notes

1. This fictionalism about responsibility turns out to be a subspecies of moral fictionalism (Joyce 2001; Kalderon 2005) in that it is not the more usual version, since it is primarily about attributions of responsibility to persons, rather than properties of moral right and wrong to actions, although the two issues are related and the two forms of fictionalism are compatible.
2. I use 'expression' in the sense that a piece of behaviour expresses a state of mind if it is designed to reveal that the agent is in that state of mind. Thus, if I want to lie or otherwise mislead I might assert something that I don't believe, but since I can't express a state of mind I do not have, my assertion merely purports to express that state of mind. If you believe that James is 6ft tall (or merely intend to get me to believe that) then you might say, 'James is 6ft tall'. Similarly, if you believe that James is ill, then you might say, 'James is ill'.

3. Different frameworks might put this in different ways, either in terms of the social norms that govern different speech acts, or in terms of the mental states expressed. These details are unimportant for now. I will use the simplest framework for illustrative purposes.
4. Ontological unease is thus a generalised version of Mackie's 'argument from queerness' for moral properties. Mackie thought that moral properties were peculiar because they have a 'to-be-doneness', but there may be other reasons to find certain properties peculiar.
5. Indeed, within yet another different value system, homosexuality is thought of as a *sin*. More on this kind of medico-moral interplay later.
6. Note how calling an adult 'naughty' sounds very odd – something Monty Python exploited to great comic effect: 'He's not the Messiah, he's a very naughty boy!'

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on the contributor

Sam Wilkinson is Senior Lecturer in Philosophy at the University of Exeter. His research is mainly in the philosophy of mental health and wellbeing, with a specific focus on hallucinations, delusions, psychosis and trauma. He is author of the forthcoming *Routledge Contemporary Introduction to Philosophy of Psychiatry*.

References

Ayer, A. J. 1952. *Language, Truth and Logic*. New York: Dover Publications, first Dover edition.
Blackburn, S. 1998. *Ruling Passions*. Oxford: Clarendon Press.
Cartwright, W. 2006. "Reasons and Selves: Two Accounts of Responsibility in Theory and Practice." *Philosophy, Psychiatry, and Psychology* 13 (2): 143–155.
Davies, W. 2016. "Externalist Psychiatry." *Analysis* 76 (3): 290–296.
Eastman, N., G. Adshead, S. Fox, R. Latham, and S. Whyte. 2012. *Forensic Psychiatry*. Oxford University Press.
Edwards, C. 2009. "Ethical Decisions in the Classification of Mental Conditions as Mental Illness." *Philosophy, Psychiatry, and Psychology* 16 (1): 73–90.
Frankfurt, H. 1971. "Freedom of the Will and the Concept of a Person." *Journal of Philosophy* 68 (1): 5–20.
Fulford, K. W. M. 1989. *Moral Theory and Medical Practice*. Cambridge: Cambridge University Press.
Glackin, S. 2017. "Individualism and the Medical: What about Somatic Externalism?" *Analysis* 77 (2): 287–293.
Hare, R. M. 1986. "Health." *Journal of Medical Ethics* 12 (4): 174–181.
Joyce, R. 2001. *The Myth of Morality*. Cambridge: Cambridge University Press.
Kalderon, M. 2005. *Moral Fictionalism*. Oxford: Clarendon Press.
Kazda, L., K. Bell, R. Thomas, K. McGeechan, R. Sims, and A. Barratt. 2021. "Overdiagnosis of Attention-Deficit/Hyperactivity Disorder in Children and Adolescents: A Systematic Scoping Review." *JAMA Network Open* 4 (4): e215335.
Leshner A. 1997. "Addiction is a Brain Disease, and It Matters". *Science* 278 (5335): 45–47.
Newen, A., L. De Bruin, and S. Gallagher. 2018. *The Oxford Handbook of 4E Cognition*. Oxford: Oxford University Press.
Pickard, H. 2011. "Responsibility Without Blame: Empathy and the Effective Treatment of Personality Disorder." *Philosophy, Psychiatry, and Psychology* 18 (3): 209–224.
Pickard, H. 2013. "Responsibility Without Blame: Philosophical Reflections on Clinical Practice." In *The Oxford Handbook of Philosophy and Psychiatry*, edited by Bill Fulford, Martin Davies,

Richard Gipps, George Graham, John Sadler, Giovanni Stanghellini, and Tim Thornton, 1134–1152. Oxford University Press.
Pickard, H. 2017. “Responsibility Without Blame for Addiction.” *Neuroethics* 10 (1): 169–180.
Shoemaker, D. 2011. "Psychopathy, Responsibility, and the Moral/Conventional Distinction." *Southern Journal of Philosophy* 49 (s1): 99–124.
Szasz, T. 1960. “The Myth of Mental Illness.” *American Psychologist* 15 (2): 113–118.
Taylor, C. 1976. “Responsibility for Self.” In *The Identities of Persons*, edited by Amelie Oksenberg Rorty, 281–299. Berkeley: University of California Press.
Toon, A. 2016. “Fictionalism and the Folk.” *The Monist* 99 (3): 280–295.
Watson, G. 1975. “Free Agency.” *Journal of Philosophy* 72 (April): 205–220.
Wilkinson, S. 2020. “Expressivism About Delusion Attribution.” *European Journal of Analytic Philosophy* 16 (2): 59–77.
Wilkinson, S. 2022. “Psychiatric Fictionalism.” In *Mental Fictionalism: Philosophical Explorations*, edited by Tamas Demeter, Adam Toon, and T. Parent, 235–252. Abingdon: Routledge.
Wolf, S. 1987. “Sanity and the Metaphysics of Responsibility.” In *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*, edited by Ferdinand David Schoeman, 46–62. Cambridge: Cambridge University Press.
Yablo, S. 2002. “Abstract Objects: A Case Study.” *Philosophical Issues* 12: 220–240.