Neuroethics
https://doi.org/10.1007/s12152-017-9347-2

ORIGINAL PAPER

Delusions, Harmful Dysfunctions, and Treatable Conditions

Peter Clutton · **Stephen Gadsby**

Received: 26 June 2017 / Accepted: 9 October 2017

Abstract It has recently been suggested that delusions be conceived of as symptoms on the harmful dysfunction account of disorder: delusions sometimes arise from dysfunction, but can also arise through normal cognition. Much attention has thus been payed to the question of how we can determine whether a delusion arises from dysfunction as opposed to normal cognition. In this paper, we consider another question, one that remains under-explored: which delusions warrant treatment? On the harmful dysfunction account, this question dissociates from the question about dysfunction—there are a broad range of "treatable conditions" beyond mere harmful dysfunctions. As such, many conditions that arise from normal cognition are also eligible for medical intervention. We argue that some delusions that arise from normal cognition may well fall under the banner of treatable conditions. We examine the practical and ethical questions surrounding such treatment, including the issue of coercive and deceptive treatment options.

Keywords Delusion · Treatment · Symptom · Dysfunction · Disorder

P. Clutton (✉)
Philosophy Department, Macquarie University, Sydney 2109, Australia
e-mail: peterclutton@outlook.com
e-mail: peter.clutton@mq.edu.au

S. Gadsby
Monash University, Wellington Rd, Clayton, VIC 3800, Australia

Introduction

Philosophers analysing delusions have recently made the important step of connecting questions about the nature of delusions to wider questions in the philosophy of medicine. The most common approach in that vein has been to show how delusions can fit into Wakefield's [1] Harmful Dysfunction (HD) account of disorder [2, 3]. In one recent example, Sakakibara [3] argues that delusions should be conceived of as a symptom on the HD account—delusions are a symptom that sometimes signify an underlying dysfunction, but can also arise from normal functioning.

This attempt to link delusions to the HD account of disorder is promising. The idea of delusion as a symptom on the harmful dysfunction account has the potential to explain a range of philosophical issues related to delusions. In this paper, we seek to highlight an important aspect of the HD account as it relates to *treatment*, an aspect that we think has been somewhat overlooked in the current literature: the potential double dissociation between treatment and dysfunction. The dissociation going one way is already readily acknowledged: on the HD account, there can be dysfunctions that are not harmful, and thus do not warrant treatment. But many have pointed out that it is often warranted to treat harmful symptoms *even when they arise from normal functioning* [4, 5].

This is an important consideration for the HD account of delusions, but it has been somewhat overlooked. For example, Sakakibara [3] appears to closely link the idea of treatment to disorder. He states that if a symptom responds to treatment, then that is a

Published online: 18 October 2017

potential indicator that the symptom arose from an underlying dysfunction. In addition, Sakakibara [3, pp. 147,156] states that we need to discern delusions that arise from normal functioning from those that arise from dysfunction, because only the latter "are the potential target of medical intervention."[1]

But on the harmful dysfunction view, that is not always so. Indeed, there is a well-established class of what have been termed "treatable" or "undesirable" conditions [4, 5], which comprises a broader range of harmful conditions than the more restricted class of harmful *dysfunctions*. Some treatable conditions arise from normal functioning, but are harmful and thus eligible for treatment.

In what follows, we consider how this idea applies to delusions. In Delusions and the HD Account, we review the harmful dysfunction account along with the recent attempt to place delusions within the HD account as a symptom, and explore some of the advantages of this approach. In Delusions and Treatable Conditions, we review the idea of treatable conditions, and consider how it applies to delusions. We suggest that the potential for treatment and dysfunction to come apart is particularly important for delusion researchers to address, since we are likely to remain in a situation of epistemic uncertainty about the dysfunctionality of some delusions, at least for the foreseeable future. Moreover, we argue that if some delusions really do turn out to be the result of normal functioning—as it appears they might—they seem to be exactly the kinds of cases that are eligible to be "treatable conditions." Thus, we should at least consider treating certain delusions even if they are the result of normal functioning. Finally, we ask whether the dysfunction question has any role in informing treatment types and modalities. For example, it has been argued that at the very least, coercive treatment should be restricted to cases involving dysfunction [3]. We argue that on the harmful dysfunction view, facts about dysfunction can inform our value perspectives, in a number of complex ways, but questions about treatment remain value-centric.

[1] At various points, Sakakibara seems to be talking only about the question of when coercive treatment is warranted, but at other points he includes any kind of "medical intervention," and on one occasion (p. 155) appears to run together the question of pathology and treatment when discussing the potential dysfunctionality of delusional jealousy. In The Uncertainty Surrounding Dysfunction in Delusions we show why it is particularly problematic to link treatment to dysfunction in the case of delusions, and in Treatment Types: A Role for Dysfunction after all? we discuss how this relates to coercive treatment. Whatever Sakakibara's original intent, the idea raises interesting questions that are worth exploring.

Delusions and the HD Account

The Harmful Dysfunction Account

In the philosophy of medicine, a key question is: what is *disease* (or disorder). Attempts to analyse the concept of disease traditionally break down into those that view disease as involving some objective basis, such as a distinct deviation in biological functioning, and those that view disease as socially constructed: what counts as disease is a result of a value judgment reflecting what people in certain times and places regard as violating norms [6].

Hybrid accounts aim to respect the idea that disease involves both elements: there are objective facts that enter into a determination of what counts as a disease—they are out there in the world, waiting to be found—but which of those objective facts we pick out as being diseases always involves a value judgment. The most prominent hybrid account in the recent literature is Wakefield's account of disorder as a harmful dysfunction.

On Wakefield's [1] conception, a disorder is a harmful dysfunction, where "dysfunction" is an aetiological notion fleshed out in terms of some biological mechanism failing to play the function for which it was shaped by natural selection, and "harmful" is a socially determined concept regarding what is harmful to a person. Of the idea of dysfunction, Wakefield [7, p. 152] states:

> [A] natural function of an organ or other mechanism is an effect of the organ or mechanism that enters into an explanation of the existence, structure, or activity of the organ or mechanism. A "dysfunction" exists when an internal mechanism is unable to perform one of its natural functions.

And of the idea of harm, Wakefield [7, p. 151] elaborates that:

> Medicine in general, and psychiatry in particular, are irrevocably value-based professions. "Harm" is construed broadly here to include all negative conditions.

We can see how the account works by observing how it classifies various examples. The harm condition exists as part of a recognition that medicine is deeply tied to value judgments about what harms people. There can be instances where something is biologically dysfunctional

but causes no harm, and such examples do not count as disorders. Wakefield [7] gives the example of benign angiomas, blood vessels connecting to the skin as a result of abnormal growth processes. These are common and harmless, and thus physicians would not generally consider them to be disorders (even though they are dysfunctions).

Conversely, the account allows for and explains instances of negative conditions that do not involve some underlying deviation from selected biological function, and thus should not be considered disorders. Illiteracy can be harmful, but if it is simply the result of a lack of opportunity rather than some underlying neurological or cognitive dysfunction, then it is not a disorder [7]. Dyslexia arising from a neurological dysfunction, on the other hand, would be considered a disorder (assuming it is harmful).

Finally, note the broad construal of harm on this account. Conditions can harm a person in various ways, and different types and quantities of harm might be disvalued to a greater or lesser extent. They may also be disvalued by different people and different cultures in different degrees. The harm element of the harmful dysfunction account is thus a wide-ranging, culturally-sensitive social judgement, and like all value judgements, harm judgements are not always straightforward.

Delusions on the HD Account

The most recent—and strongest—attempt to situate delusions within the harmful dysfunction account comes from Sakakibara [3].[2] Sakakibara argues that delusions should be thought of as symptoms on this account. That is, they are a surface manifestation that often indicate underlying dysfunction. But like many symptoms, delusions can also arise from normal functioning. In this regard, Sakakibara [3, p. 147] draws a distinction between physical symptoms like fever, and psychological symptoms, like sadness in mood disorders. He states that the former are typically "unambiguously pathological," whereas the latter may arise "independently of any illness." Delusions fit into the latter category on this account: a mental symptom that is often, but not necessarily a result of underlying dysfunction.[3]

Sakakibara goes on to provide an account of when we might infer an underlying dysfunction in particular cases of delusional symptom. In general, we are looking for clues which lead us to suppose that the best explanation for some delusional belief is that it involves an underlying dysfunction. One thing we cannot use, Sakakibara argues, is simple irrationality, since irrationality is a common feature of normal human cognition [8]. Delusions may be defined in clinical manuals like the DSM [9] as being irrational, but irrationality itself is not sufficient for an underlying dysfunction. We need other clues.

One such clue, according to Sakakibara, is the "un-understandability" of a delusional belief: if we are unable to grasp how it is that someone could have come to their delusional belief, given their circumstances, this can be a clue that the delusion is a result of some underling dysfunction. Relatedly, we should look for "uniqueness"—when a delusional belief is totally idiosyncratic, and not shared by others, it is more likely to have arisen through endogenous dysfunction.

In addition to these, Sakakibara adds some further attributes that increase the likelihood that there is some medical explanation for a certain belief. If the belief coexists with psycho-physiological disturbance or decreased levels of functioning, for example, this can lead us in the direction of assuming an underlying dysfunction. Further, if there is a preceding organic disease that is known to be associated with delusional beliefs—like stroke, dementia, etc.—this is strong evidence that the delusional belief has arisen from the underlying dysfunction. Likewise, delusions displaying bizarre content are likely linked to an eliciting abnormal experience caused by some kind of sensory dysfunction. Finally, Sakakibara states that while it is true that many illnesses have no known treatment, if a condition does respond to some form of treatment, then this is a sign that it might be caused by dysfunction.

[2] A previous attempt can be found in the work of Miyazono [2]. We think that Sakakibara's work is an improvement on this attempt, mostly for the explanatory power cited in Explanatory Power. Miyazono tends to downplay the idea that delusions are symptoms, and that symptoms can often arise in the context of normal functioning, which we think is a key observation about delusions.

[3] We should point out that we think this is an unnecessary dichotomy: "physical" symptoms *can* arise in the course of normal functioning, though it is true that this problem is more common and more difficult to resolve for mental symptoms.

Explanatory Power

The above view has some clear advantages, not all of which have been fully appreciated. The conception of delusions as symptoms on the harmful dysfunction account has the potential to provide insight into numerous philosophical questions about delusions.

Firstly, this account brings delusions in line with other psychiatrically-interesting states. Confabulation, for example, is a phenomenon of clinical interest, and was once viewed as a symptom that only occurred in the single neuropsychological condition of amnesia (for a review, see [10]). But confabulation is now considered to be a symptom in numerous neuropsychological conditions, and further, it has been noted that the phenomenon is common even in non-clinical subjects, arising through (presumably) normally functioning mechanisms [10]. In a well-known experiment [11], (non-pathological) subjects are placed in a store and asked to choose between two stockings, which unbeknown to them are completely identical. When asked why they chose as they did, subjects almost invariably offer up some confabulatory answer, citing the superior knit or some other imagined factor. Confabulation is thus a state that arises through multiple aetiologies, as well as through normal cognition. It is simultaneously a phenomenon of clinical interest, while still being the type of state that should not be defined in terms of its general dysfunctionality.

The account of delusion as symptom also helps illuminate the relationship between the different approaches to delusions seen in the clinical setting and in the research (i.e. cognitive science/cognitive neuropsychiatry) setting. On the current account, it seems natural (even if over-simplified) to say that clinicians are interested in harmful symptoms, even when they do not arise from dysfunction, whereas cognitive neuropsychiatrists are interested in dysfunctions, even when the resulting symptoms are not clinically significant. Since not every instance of delusional symptom will involve an underlying dysfunction, making this distinction helps to avoid a common confusion: the two groups are not always interested in addressing and explaining the same phenomena; they might rather be thought of as addressing two overlapping subsets of "delusions."

On a related note, the account of delusion as symptom also gives us a way of clarifying recent debates over the definition of "delusion."[4] Researchers have often debated what—if any—aspect of the definition can mark delusions out as pathological, in ways that either presuppose that delusions necessarily arise from pathology, or at least partly gloss over the potential for delusions to be non-pathological. Such an approach runs into the problem of trying to formulate a definition that is relevant to clinicians as well as cognitive scientists. But as Davies et al. [13, p. 134] have pointed out: "it is a highly non-trivial task to provide a definition [of delusions] that meets the needs of both psychological theory and clinical practice." If we instead accept that delusions as clinically defined are symptoms, and that symptoms do not always indicate an underlying pathology, we then have a framework within which to go about identifying the different aspects of delusions. On this account, the general question "what makes delusions pathological?" looks like it might perhaps be ill formed.

Thus those who attempt to provide a definition of delusions that simultaneously respects clinical interests and current clinical definitions, while also trying to mark out how delusions are (neuro-)psychologically aberrant compared to other states, are perhaps already on the wrong track. We might instead accept that there is a difference of approach between these two fields: they are related, but have distinct interests and perspectives. This fits with what cognitive neuropsychiatrists have said about their practice [14, p. 4], and also with recent pluralistic approaches which encourage the pursuit of multiple nosological categories from different "systems of practice" (e.g, clinical versus research practices) [15].

Along the same lines, the account of delusion as symptom helps illuminate the disputes over the taxonomy of delusions one from another. One common way to taxonomise delusions is by their content. Cotard's delusion involves the belief that one is dead; Capgras delusion involves the belief that a loved one has been replaced by an impostor; and so on. Theorists often advert to such definitions in passing, and sometimes explicitly refer to content in debates about whether a particular belief should be classified as one delusion over another (for an example, see [16, p. 356]).

However, some have argued that a better way to taxonomise delusions is by their aetiology (Clutton, Gadsby, & Klein, Under Review). On this view, the taxonomic principle for distinguishing and grouping different delusions is their aetiology: did the delusion arise from a face processing impairment, or from the abnormal experience of hemiplegia, and so on. Sometimes you find similar content that arose from two different aetiologies:

[4] For discussions of delusions and the relation to pathology, see [2, 8, 12].

in mirrored-self misidentification, for example, reported aetiologies include both mirror agnosia and face-processing impairment [17]. On the aetiological approach, these would count as different delusions.

The account we have considered here provides a framework for understanding this dispute. One way of taking the above is that clinicians taxonomise by content because they are interested in symptoms, especially in an area like delusions where we do not have a full understanding of their various aetiologies. However, cognitive neuropsychiatrists should indeed be interested in taxonomising by aetiology for cases where we *do* have evidence about dysfunction (like the two types of neuropsychological breakdowns seen in mirrored-self misidentification). Seen in this way, the call for an aetiological taxonomy is a call for a taxonomy of delusional *disorders*, (i.e. identifying different underlying dysfunctions that can give rise to delusions), whereas taxonomy-by-content is an attempt at identifying descriptive features in the absence of full aetiological knowledge [18].

In sum, conceiving of delusions as symptoms on the HD account has a number of strengths, and has underappreciated explanatory power. It brings delusions in line with similar phenomena such as confabulation, helps clarify debates regarding the relationship between delusion and pathology, and illuminates the relationship between two distinct contexts in which delusions are investigated. As such, the HD symptom view deserves to be taken seriously and explored by more philosophers.

Delusions and Treatable Conditions

We have argued that the HD symptom account of delusions has numerous strengths, and is worthy of further attention. In this section, we will discuss how this account relates to the question of when it is warranted to treat delusions. In the current literature, there has sometimes been a tendency to tie the treatment question to the question about when delusions arise from dysfunction. Sakakibara [3, pp. 147,156], for example, seems to directly associate the idea of underlying dysfunction with treatment, sometimes claiming that only cases of dysfunctional delusion "are the potential target of medical intervention."

But there are problems with this claim. First, from a pragmatic standpoint, we are likely to remain in a position of epistemic uncertainty as to the dysfunctionality of some delusions, at least for the foreseeable future. The project of discerning the selected function of mental states is only just beginning. In the meantime, we should be thinking about what to say about treatment when we are unsure as to whether a particular delusion arises from an underlying dysfunction.

Moreover, on the HD account, harmful dysfunctions are only one category of a broader class of "treatable" or "undesirable" conditions [4, 5]. Treatable conditions encompass many conditions that arise from normal functioning, an important aspect of the harmful dysfunction account [19, pp. 374,391]. Thus, if we are trying to situate delusions within the HD account, we must acknowledge the possibility that non-dysfunctional delusions may be eligible for treatment.

When we turn our attention to some of the proposed examples of potentially non-dysfunctional delusions in the literature, we think they indeed appear to fit under the heading of "treatable conditions." We show where some non-dysfunctional delusions might fit within the various types of treatable conditions, and suggest that these cases might warrant treatment. Thus, more attention ought to be directed to thinking about when we should treat such cases, and the ethical questions that surround that treatment.

Finally, we consider whether dysfunction has any role in informing treatment types and modalities. One might want to say that even though treatment is not restricted to cases of underlying dysfunction (disorder), perhaps coercive treatment is restricted to such cases [3]. That suggestion has merit, but we do not think the issue is so straightforward. Questions of treatment are always and thoroughly value-laden on the HD account. The recognition of an underlying dysfunction might recommend some approaches to treatment over others, since aetiology can be relevant to treatment choice, but there are still a range of ethical concerns to be dealt with, and the existence of dysfunction does not do all the work. We conclude that the dysfunction question can inform treatment types, although this is not always straightforward.

The Uncertainty Surrounding Dysfunction in Delusions

Before discussing the notion of treatable conditions, we point out a practical issue with too closely associating treatment and dysfunction in the case of delusions.

Deciding where disorder exists on the harmful dysfunction account requires us to discern two things: where harm exists, and where a breakdown in selected

function exists. Recognising cases of breakdown in selected function in turn requires us to have at least some notion of what that selected function is (and be able to differentiate it from related systems) [4, p. 457]. In the case of delusions, the relevant breakdown might be in any number of the upstream systems that lead to belief.[5] For example, there is compelling evidence that a breakdown in face-perception leads to the Capgras delusion [21]. Here we have good evidence of a particular breakdown involved in a disorder, which gives rise to the delusional symptom.

But once we start considering the multi-faceted nature of belief formation, and the potentially diverse range of inputs like motivation, the story becomes more complex. What kinds of motivation are likely to be adaptive inputs? Does the selected function of belief map onto traditional epistemic norms, as some have argued recently [22], or is such a view "deeply mistaken," as has also been argued [23]? How does one go about giving an evolutionarily-informed, modular account of a system as open-ended and wide-ranging as belief formation [24]? There has been progress on these questions, but belief stands out as being uniquely difficult to characterise along functional lines [25].

We should expect complex states like belief to defy superficial attempts at discerning their selected function. Such has been the case with other complex states, like memory. There has been a wholesale shift in memory science away from the view that phenomena such as false, distorted, and forgotten memories are always the result of dysfunction, since "there is little adaptive value in designing a system to recover the veridical past, given that the past can never occur again (at least in exactly the same form)" [26, p. 240]. Some researchers have thus begun the project of specifying the complex set of "adaptive constructive processes": the "processes that play a functional role in memory and cognition but produce distortions, errors, or illusions as a consequence of doing so" [27, p. 2]. And all of these processes seem to be driven by motivational and other aspects of our conception of self: "memories may be altered, distorted, even fabricated, to support current aspects of the self" [28, p. 595]. Thus, even if memory turns out to consist of specific modules [29] whose functions can be given a single description at some higher level of analysis, such functions are clearly highly flexible and context-dependent [30], and it is no easy task to judge the appropriate levels of influence from the various input factors from any one instance of that state.

These difficulties are evidenced in the study of memory conditions like hyperthymesia (cases of highly unusual, near-complete and veridical access to autobiographical memory) [31, 32]. This condition can be highly distressing, and although it appears to involve some kind of dysfunction, perhaps in executive function, episodic memory is not well-enough understood to know for sure. Some have even argued against the idea of dysfunction in hyperthymesia, arguing that it is rather a case of hyper- as opposed to dysfunction, the price to be paid for certain mechanisms shaped by primitive evolution [33]. The difference between dysfunction and hyperfunction is difficult to determine, and assessments about complex states like memory are thus highly problematic.

Problems like these suggest that our understanding of the function-dysfunction boundary of certain mental states might not be mature enough to be of much practical use. If we want to ask questions about when and how to treat something, for example, our knowledge of the function-dysfunction boundary is nowhere near ready to give us much direction on that question.[6]

Perhaps we can return to the list of clues offered by Sakakibara [3] for determining where dysfunction is present in delusions. But we think these clues do not give us enough certainty to overcome the above

[5] For our purposes, we take it that delusions are beliefs, which is the general clinical and scientific conception. For an overview of some philosophical debates, and an argument in favour of the doxastic conception of delusions, see [20].

[6] Similar sentiments have been expressed by others:

> Understanding psychopathology based on an analysis of harmful dysfunction will therefore not be relevant from a practical clinical perspective without an evolutionary psychology of normal functional psychological mechanisms as well as psychopathology—that is, a science of function is necessary to define dysfunction. Whether or not something has an evolved function is currently not a question that may be easily resolved—the mapping of our adaptations has only just begun. [34, p. 453]

> evidence does not yield a clear demarcation between normality and disorder for mood phenomena, and the evolutionary literature does not much aid in delineating the boundary either. We may have to accept that much of what we currently identify as clinical depression cannot be shown to be dysfunction, and moreover that the clear presence of dysfunction cannot be used as the criterion for applying or withholding medical attention. [35, p. 212]

> it might be true one day—though it isn't true at present—that our theories about how normal cognitive processes operate have become so detailed that they make predictions about whether particular treatment techniques will or won't be effective. If that day comes, then treatment studies will be relevant to theories about normal cognitive, and cognitive neuropsychology will be directly concerned with issues involving treatment. [14, p. 4]

concerns. For example, Sakakibara suggests that response to treatment is one potential indicator of an underlying dysfunction. But as Wakefield has repeatedly observed [36–38], successful treatment neither implies an underlying dysfunction, nor reveals the nature of some possible dysfunction: one can treat pain seen in "dyspareunia" without that telling us whether there is really any dysfunction present, and without an indication that the relevant dysfunction is in the pain systems (rather than somewhere else) [38]. Sakakibara's other clues are helpful, but do not seem like they can definitively resolve questions on cases where we are uncertain. For instance, it is true that some diseases are already known to be associated with delusions, and so their presence can indicate an underlying dysfunction, but it is precisely in cases where no such gross disorder is present that the question about dysfunction is most vexing (e.g. a case of monothematic delusional jealousy with no obvious abnormalities). Of course, these clues were not meant to be necessary and sufficient conditions for determining pathology, and considered together, they might make us lean towards a pathology judgment on the basis of an inference to the best explanation, as they were intended to do. But there will likely be cases where we remain too uncertain to base anything of import on this.

At the very least, then, we will sometimes find ourselves in a state of epistemic uncertainty about dysfunction in individual cases of delusion. Given this uncertainty, we should be wary of relying on identifying dysfunction to determine whether a condition should be treated. Given that there is a practical necessity to identifying a condition as warranting treatment, we must at least be willing to consider the treatment question independently of the dysfunction question. Similar approaches have been taken for medical conditions like the various "pain syndromes" [38], where the existence and nature of any underlying dysfunction is either unclear or disputed. Such conditions still require current attention while the dysfunction question is being decided.

Treatable and Undesirable Conditions

So there are pragmatic reasons for not tying treatment solely to cases of (harmful) dysfunction. But that is somewhat by the by, since on the HD account, harmful dysfunctions are not the sole target of treatment anyway. Rather, there is a broad class of "treatable" or "undesirable" conditions, of which harmful dysfunctions are only one instance. We often want to treat symptoms or other conditions that arise from normal functioning, because they involve harm.

Treating cases based on harm is of course intuitively plausible. Patients demand treatment based on how bad things are for them, "not on the basis of whether the mechanisms [are] maximising current reproductive success or fulfilling their ancestral function" [35, p. 208]. That seems natural, and has been explored on the HD account as it relates to a number of mental symptoms. Just because negative emotions like low mood often arise normally rather than from dysfunction, and just because you think we should have an evolutionarily-informed account of disease like the HD account, does not mean that you should only treat dysfunctional cases:

> Some people think that the utility of negative emotions means that they should not be treated. This is a serious mistake. We have much to learn from general medicine, where both the utility and the harm caused by responses such as pain and diarrhea is well recognized, and where relief of suffering by blocking defensive responses is a routine goal of clinical work, *whether the symptom is being aroused normally or arises from a faulty mechanism* [...]
>
> ([39, p. 194], emphasis added)

This perspective on treatment is encompassed by the original harmful dysfunction account [19], and a failure to acknowledge this aspect of the HD account is to do it a disservice. Indeed, some [4] have claimed that a misunderstanding of the HD view as claiming that disorders are the sole target of treatment has led to unnecessary reluctance to adopt Wakefield's otherwise conceptually powerful account.

Cosmides and Tooby [4] thus set out to provide a "taxonomy of treatable conditions." They define treatable conditions according to the following criteria [4, p. 456]:

> (a) a characterizable condition in a person; (b) a person or social decision-making unit whose values and decisions will govern the actions taken with respect to the condition; (c) a valuation by that person or unit that the condition is negative and that it ought to be changed (that is, that the persistence of the condition is "harmful," "undesirable," or "unhealthy"); and (d)

knowledge of a method for changing the condition in the desired direction.

One could raise questions about these criteria, and others have offered their own take on similar phenomena. For example, we agree that it seems reasonable to call something a treatable condition if a treatment is merely conceivable and desirable, the point being that the condition is the kind of thing that warrants treatment, even if none is currently available [40, p. 149]. Additionally, we take it that treatment should be understood in broad terms, not limited to drugs and surgery: counselling, cognitive behavioural therapy, ameliorative care, coping strategies, and so on, are all treatments offered by various clinicians, and any of these might be directed at treatable conditions. For our purposes, though, the main point is that some condition is harming a person in a way that is of clinical interest, even if there is no dysfunction involved, and treatment is thus at least something worthy of consideration.

A common form of such cases is where there is a mismatch between evolved function and current environment, where a condition might be functioning correctly but still cause problems for a person. Cosmides and Tooby class these as "development-environment mismatches." Mechanisms driving male sexual jealousy, for example, were selected to be adaptive in ancestral environments, but these mechanisms might operate normally in a modern society in a way that produces distressing, harmful, and disfavoured behavioural and emotional patterns. Another common class of treatable conditions are evolved defence mechanisms. Pain, fever, and diarrhea are all defence mechanisms that can cause subjective suffering and thus warrant treatment even when functioning normally. In the realm of mental symptoms, it has been suggested that emotions like anxiety are defences that often over-act and cause suffering, even while behaving normally: defensive mechanisms favour the false-positive end of the sensitivity spectrum, to maximize their utility, which can lead to unnecessary suffering [41]. There can also be mechanisms that are adaptive on average across the population, even though they may result in harm in individual instances: risk-taking behaviour might be a generally winning strategy evolutionarily speaking, but of course it sometimes fails.

Along these same lines, Del Guidice and colleagues [5, 42] have recently offered their own take on such conditions, which they refer to as "undesirable conditions." On their taxonomy, Wakefield's harmful dysfunctions are just one type of undesirable condition. Opposed to harmful dysfunctions are the range of *functional mechanisms* that are undesirable. Some of these will be *currently maladaptive across individuals*. On the other hand, there are the functional mechanisms that are *currently adaptive across individuals*. These latter break down at the level of the single individual into those that are *adaptive* and *maladaptive* for a *particular* individual. Some defence mechanisms, for example, are generally adaptive across individuals, but may lead to maladaptive outcomes at the individual level. Interestingly, there can be adaptive individual outcomes of generally adaptive, functional mechanisms, which nevertheless are considered undesirable, like various kinds of antisocial and exploitative strategies: psychopathy has survived in the gene-pool as an adaptive and sometimes successful response to certain environments, but one that we might nevertheless want to treat.

For our purposes, there is an important distinction to be made in these examples. On the one hand, there can be normal functions, like defensive functions, that arise from, or are indicative of, an underlying dysfunction, even though the surface manifestation that is causing harm is not itself dysfunctional. Pain, fever, and diarrhea can cause harm (which warrants treatment) while functioning normally, but importantly, these normally functioning defences often indicate an underlying dysfunction. So in that case, there is a normal surface condition that is the downstream result of an underlying dysfunction. Perhaps it is not so surprising that such cases could warrant treatment. But note that there can be cases of normal functioning that *do not* result from underlying dysfunction*, and that these conditions can also warrant treatment* (based on the harm they cause) [43]. The example of normal-but-overactive anxiety mechanisms would be one such case: on some evolutionary hypotheses of anxiety, a normally functioning anxiety mechanism could produce pronounced harm in the absence of any "deeper" dysfunction, and yet still warrant treatment [41]. These kinds of cases are of interest to us as they pertain to delusions. Sakakibara [3] suggests that some delusions might be the result of normal functioning, and that such cases—because they lack pathology—would not be eligible for medical intervention. But in fact, conditions that lack pathology can still be treatable conditions.

Before moving on to consider how all this applies to delusions, we would like to note an alternative stance

that one might take on these examples. One might think that treatment and disorder should map more closely onto each other [44–46], and thus reject the very idea of non-disordered, treatable conditions. That is, some accounts of disorder deny the separation of treatment and disorder that is implied in the HD/treatable conditions account. We note this possibility in order to set it aside: our interest here is in the HD analysis of delusions, and what that does or does not imply about the potential treatment of non-dysfunctional delusions. One may favour other accounts of disease, in which case it is their task to show how the cases we discuss in this paper fit into those accounts.

On the HD account, then, not only are there a broader range of treatable (undesirable) conditions than just Wakefield's harmful dysfunctions (disorders), but harmful dysfunctions in fact represent "only a fraction of what people regard as diagnosable problems or seek treatment for" [42, p. 11].

Non-Dysfunctional Delusions as Treatable Conditions

We think that there really might be cases of non-dysfunctional delusions that nevertheless are best thought of as treatable conditions. Since some examples of potentially non-dysfunctional delusions have been proposed in the literature [3], and since it has been suggested that such examples might not be eligible for treatment, we will examine these examples in light of the idea of treatable conditions. If these examples seem to fit the description of treatable conditions, then we should reject any over-hasty dismissal of treatment: we must at least be open to the idea of treating such cases.

One example that Sakakibara [3] gives is where delusions might arise from normal motivational and self-deceptive influences. An example might be reverse-Othello's syndrome: falsely believing that a former partner still reciprocates your love (or is faithful, etc.).[7] Although such a case could arise from dysfunction, it is easy to imagine how normal, motivationally-driven, self-protective thoughts might lead a person to this delusion. That would then be an example of a non-dysfunctional delusion.

Another example is where normal mental functions might be exaggerated in a way that leads to delusional belief. Suspicion and jealousy are normal mental processes, but these processes might, in certain circumstances, lead to delusions like delusional jealousy (Othello's syndrome). Again, it seems clear that delusional jealousy sometimes arises from a dysfunction (like dementia [47]), but we can see how a person might possibly arrive at such a delusion through (functionally) normal—or at the borderline of normal variation—mental processes of suspicion and jealousy.

To these examples we would like to add a class of delusions that we suspect might at least sometimes be non-dysfunctional: cases of folie à deux. In folie à deux, a secondary patient adopts the delusional belief of a primary who is typically suffering from a psychotic disorder [48]. Secondaries are often related to and dependent on the primary, and their delusion often resolves if they are separated from that person (though not always [49]). It is plausible that the propensity for a dependant to adopt beliefs from those they are closely reliant on is the normal function of the belief system [48].

It seems to us that these examples would be paradigmatic instances of treatable conditions. The case of reverse-Othello, and other kinds of self-deception, sound like defensive functions that might be functioning normally even while producing harmful, maladaptive results for an individual. Self-deception might be generally useful but still capable of causing harm. If this delusion resulted in distress to the patient (or any other kind of harm), we think that the question of treatment is at least on the table, regardless of the fact that the delusion might be the result of normal functioning.

Similarly, the case of delusional jealousy might be considered an instance of development-environment mismatch. That is, while the development of systems of jealousy and suspicion might have been driven by certain selective pressures to secure resources, mates, and so on, these mental states might lead to distress and harmful behaviour in the current environment (as cases of delusional jealousy often do [50]). We think that once the existence of treatable conditions has been acknowledged, our intuition about treating such cases is to consider the harm involved, regardless of whether there is an underlying dysfunction.

Finally, our suggested non-dysfunctional cases of folie à deux might be best thought of as falling under the branch of functions that are generally adaptive across individuals (willingness to adopt beliefs from those we are dependent on), but which in a specific circumstance (being dependent on someone experiencing psychosis) can lead to maladaptive outcomes at the individual level.

[7] Sakakibara [3, p. 154] does not refer to this delusion by name, but the description matches delusional cases of reverse-Othello.

There is clearly potential for profound harm in this area, where the secondary is reliant on the primary. Folie à deux can involve themes that lead to self-harm [51] as well as harming others [52]. And follie à famille and the larger shared delusions seen in apocalyptic cults might result from similar processes, with similar potential for harm at larger scales [53]. As such, these delusions might be treatable conditions (even if the relevant "treatment" is simply a "geographical intervention" to isolate the secondary and provide psychological care).

So these examples appear to fall within established categories of treatable conditions. They also match the general definition of what it is to be a treatable condition: there is a condition—involving delusional beliefs—which could reasonably be judged harmful, and there are some clear ways that they could be treated (geographical/psychological separation for folie à deux, various types of therapy and cognitive interventions, and so on). These cases, we think, are at least eligible for treatment, regardless of their lack of dysfunction.

But can we safely treat them? It has been pointed out that when deliberating about whether to treat defensive functions, for example, we need to consider the potential complications of knocking out one of the body's normal defence functions [4, 35]. Pain or fever can often be blocked without causing any further problems—or at least any problems that are not outweighed by the relief from the symptom—but there can be cases in which blocking a defence has direr consequences [35, p. 212].

On some hypotheses, these considerations might lead to the conclusion that we should at least exercise caution in treating certain delusional cases. For example, some have argued that delusions might sometimes fill the role of a "doxastic shear-pin," in that they are designed to provide a protective function, preventing further downstream damage [54]. Similarly, the "epistemic innocence" approach to delusions [55–57] frequently emphasises the ways in which the formation of a delusional belief can have certain epistemic benefits: the delusion allows one to continue operating in the world as an epistemic agent, for example. Perhaps this function could be disrupted by treatment in some cases, although of course this has to be balanced against a wider view of the harms of abstaining from treatment.

It would be too hasty to conclude from this that we should avoid treating delusions in such cases. Some of the concerns cited above might be countered by the observation that delusions are already treated in many cases, without, it seems, provoking some further downstream cognitive problem directly related to the alleviation of the delusion (although certain treatment options, like anti-psychotics, have many known side-effects, an important consideration) [58, 59]. Further, treatments like meta-cognitive interventions are increasingly being used to effectively lower delusional severity [60], and these options include cognitive training and therapy of the sort that might help prevent unwanted downstream cognitive effects once the delusion itself is no longer present. At the very least, we think that treatment of non-disordered "evolutionary defence" cases of delusions is not something that should be obviously prohibited. More work is needed.

In sum, there may be cases of non-dysfunctional delusions that warrant treatment. At the very least, we should not simply rule out treatment of such cases. If there really are cases like this, then we need to be asking questions about when we can treat a delusion even when it does not arise from dysfunction (or when we remain uncertain, one way or the other).

Treatment Types: A Role for Dysfunction after all?

We have so far stressed the ways in which treatment and dysfunction come apart on the HD account, and applied this to delusional cases. But do questions about dysfunction have no relevance at all to treatment? Some have argued that at the very least, *coercive* treatment must be restricted to cases of dysfunction [3]. And one might think that since the dysfunction question relates to aetiology, surely aetiology can be relevant to treatment options generally. In this section, we attempt to elucidate the harmful dysfunction perspective on these issues. As we will show, the story about the relation between dysfunction and treatment is somewhat complex: facts about dysfunction *can* be relevant to treatment, though as we have stressed in the preceding section, the treatment question remains explicitly value-laden on the harmful dysfunction account.

We begin with the observation that while treatment ought not be tied *directly* to dysfunction, the existence and nature of dysfunction does have *some* relevance to the types of treatment to be preferred. For one thing, there is often a difference in the prognosis of dysfunctional and non-dysfunctional cases, and that difference can weigh against more aggressive or harmful forms of treatment for non-dysfunctions [61, p. 238; 62, pp. 20-21,102]. That

is, there is generally a better prognosis for untreated non-disordered cases than untreated disordered cases: a dysfunction that is left unfixed will likely continue to cause problems, whereas non-dysfunctional symptoms may be more likely to abate on their own, or to remit when triggering stressors subside [62, p. 20]. This needs to be considered when we are thinking about the type of treatment to be preferred. Drug treatments with harmful side effects might be less favoured for a non-disordered patient—the harm of such treatments might be outweighed by the recovery in the disordered patient, whereas the non-disordered patient may have seen improvement without such treatment anyway.

Similarly, since the *nature* of a dysfunction is ultimately an aetiological matter, this has clear significance for preferred types of treatment. For example, anti-psychotics aimed at modulating salience systems might not be the most useful option for a patient we suspect is not suffering from a dysfunctional salience network. Where we suspect that a person's mental distress is a normal response to their circumstances, we might favour an environmental intervention. If that is not possible, psychologists might aim to aid people in managing their situation and coming to some kind of cognitive-emotional understanding of their symptoms [63, p. 84]. Treatment options along these lines have shown success at alleviating the distress associated with delusions [64].

So dysfunction does have relevance to treatment. However, the story is complex: dysfunction *can* be relevant to preferred treatment options, but it is *not* the directly deciding factor [63, p. 64]. In many cases, helping to alleviate a symptom can be somewhat divorced from underlying causes: aspirin works just as well for a fever of unknown origin as for a person with influenza; anxiolytics can be helpful for disordered as well as normal anxiety [66, p. 840]; environmental interventions can be equally efficacious for people suffering from a disorder as for those with a normal environmental mismatch [67]; and psychotherapy and cognitive behavioural interventions can be helpful in managing disordered symptoms just as they can in managing the mental slings and arrows of everyday life [65, p. 64]. Finally, while drug treatments, with their harmful side effects, might be less favoured in the treatment of non-dysfunctions, they are not ruled out: an informed patient, aware of the side effects, may still prefer such treatment for severe (but normal) symptoms, and clinicians ought to consider these wishes.[8]

So there is a complex relation between dysfunction and treatment: treatment is still value-centric, but values can be informed by *relevant fact*s. Dysfunctions are often *relevant facts* to questions about treatment, though again, the link is complex and indirect.

The same is true on the more specific question of involuntary treatment. In practical terms, involuntary or coercive treatment is often a consideration when dealing with delusions: delusions and other psychotic symptoms sometimes involve the kind of perceptual and thought impairments which can undermine a person's capacity to consent to treatment. Further, by their very nature, delusions are often not seen by patients as a "symptom" in need of treatment at all: they see the delusion as just another belief [3]. An important question, then, is whether and when it might be reasonable to use deception, coercion, or involuntary treatment for non-dysfunctional delusions.

Note that again, facts about dysfunction are relevant to this question, sometimes for the same reasons as for treatment more generally. For example, the difference in prognosis between untreated dysfunctions and untreated non-dysfunctions can weigh against coercive treatment, just as it can weigh against *any* harmful or aggressive treatment option. The harms of such treatments are less justifiable where we think a condition may have remitted anyway, as non-disorders can be more likely to do.

Further, facts about dysfunction can serve to help value perspectives converge. Recall that on the harmful dysfunction account, the notion of harm is societal rather than individual [37]. This leaves open the possibility of great divergence of values about the harm of some condition: is sadness a state to be chemically corrected or

[8] This complex relation between dysfunction and treatment is the reason we earlier (The Uncertainty Surrounding Dysfunction in Delusions) raised issue with Sakakibara's inference from successful "treatment" to the existence of underlying dysfunction: treating a symptom does not indicate an underlying dysfunction, since it is possible to treat normal as well as dysfunctional symptoms, sometimes through the exact same method. The inference from successful "treatment" to underlying dysfunction is considered a complete misstep on the harmful dysfunction account [66, pp. 839-840]. On this note, we should say that we think the "bizarreness" and "un-understandability" clues mentioned by Sakakibara are likely the most useful: symptoms being out of context or inexplicable when placed within an appropriate examination of the patient's life and social environment are the kinds of clues that the HD account sees as generally most useful—though still fallible—in the attempt to discern dysfunctionality [68].

bravely borne? Add to this the question—already so vexed—of coercive treatment, and a kind of radical value-pluralism threatens to undermine the ability to converge on judgements about treatment. But values, as we have said, are often informed by relevant facts, and the facts about a dysfunction typically function to narrow this kind of divergence markedly. As Wakefield [37, p. 89] states: "people may differ in how much they value joy or hate sadness in response to life's vicissitudes [...], but there is much less difference in how they feel about true depressive disorder in which something has gone wrong with the mind so as to continually generate painful sadness unrelated to actual losses." Conversely, where we suspect an absence of dysfunction, there is more room for divergence and controversy, and less certainty in making judgements in favour of treatment, especially if that treatment could itself be harmful [69, p. 93]. Where a person seeks out treatment of their own accord, this concern is less pressing, since then a person is actively seeking help, and the HD account considers it an important part of clinical practice to provide that help. But where we are contemplating undermining someone's autonomy with an imposed treatment, one that might have harmful side-effects, these considerations carry greater weight.

Finally, consider the importance of the distinction between function/dysfunction more generally on the harmful dysfunction account. Wakefield has argued that mislabelling a normal condition as dysfunctional is problematic for a number of reasons. One reason is that we want to avoid enforcing the sick role on someone who is in fact normal [70]. If we were to label some of the normal environmental mismatches described in Treatable and Undesirable Conditions as actual harmful dysfunctions, for example, instead of correctly identifying them as normal-but-societally-disvalued conditions, we label someone with a disorder who is in fact simply mismatched with their current environment/society. Labelling something as a (harmful) dysfunction in the person tends to push us towards the direction of treatment (this was the point made regarding value-convergence above: where we suspect actual disordered sadness, our values tend to push towards favouring treatment and disfavouring the continuance of that symptom/state). But if we get the label incorrect, we can be led towards an over-eagerness to enforce treatment on what are in fact normal conditions that are merely problematic in our current society [71, p. 381, 61, 69, p. 93]. As Wakefield [61, pp. 239-240] emphasises, it is perfectly well to properly acknowledge that non-disordered patients ought to be treated, "when individuals believe it will help them and when effective and safe treatments are available," but this use of psychiatry is also "prone to abuse" if we were to actually label those conditions as disorders. A mislabelling of a condition as disordered can make us less likely to raise questions about our own society and value systems, and more likely to see these problem as being "in the person," thereby threatening to turn psychiatry into a means of social control [61]. As a society, we ought to respect normal human variation, even when that variation is annoying or inconvenient, rather than diagnosing it and insisting on treatment [70, p. 5]. Note that this line of thinking does not necessarily rule out coercive treatment in all cases of non-disordered conditions, but it does imply that we ought not be injudiciously liberal with the dysfunction assignment, lest we end up also being unreasonably liberal with the pressure on patients to have their (normal) conditions treated.

So, as with treatment options more generally, dysfunction does have a role when thinking about coercive treatment. But the question of coercive treatment is still a value question, and to say that dysfunction can inform, constrain, and guide our values is not to say that it is the sole or even decisive factor. There are ethical questions to be explored about the use of coercive treatment in both disorders and non-disorders alike. Clinicians and ethicists have raised questions about the legitimacy of deception and coercion even in cases where we know that there is an underlying dysfunction [72, 73], such as in dementia [74] and traumatic brain injury [75]. And in addiction, where debate continues over the existence of disorder [76, 77], there is room for debate over potential treatment options, including legal, formal, and informal mandates (like family interventions) [78, 79].

Conversely, and returning to delusions, we think there may be instances where it is warranted to coercively intervene in cases of non-dysfunction. For instance, one common justification for undermining a person's autonomy through coercive treatment is where we suspect some pre-existing disruption to autonomy, wherein coercive treatment might be thought to be *restoring* autonomy [79]. One such disruption to autonomy that is often thought to warrant coercive intervention is where a dysfunction interferes with someone's judgement: hence, common justifications for coercion or deception of (disordered) delusional patients [80]. However, dysfunctions are not the only things that can undermine autonomy. Take the example of folie à deux. We hypothesised that the secondary patient in some folie à deux cases could be normal (non-dysfunctional). But we might think

that the relationship between the primary and secondary in such cases—dependent, dominating, and controlling as it often is [81]—can itself undermine the autonomy of the secondary [c.f 82], and that it can be warranted, in the interests of the secondary's autonomy, to coercively intervene to perform some kind of temporary psychological/geographical separation. Given that this particular kind of intervention is relatively safe and effective, and the intrusion limited and temporary, it seems justifiable in this case. At the very least, as we have suggested, there are ethical questions to ask here, beyond the factual dysfunction question.

So, there may be cases where coercive intervention is warranted in non-dysfunctions, and there is plenty of room for ethical debate even in clear cases of dysfunction. As always, the treatment question on the harmful dysfunction account is value-centric, and the ethical questions that surround coercive treatment options ought to be given due consideration. At the very least, we need to think about what kind of "coercion" is under consideration; what kinds of factors are influencing the patient's treatment decisions; the relevant local mental health and capacity laws; and so on. And decisions are still required in cases of uncertainty about dysfunction, which we have suggested is the position that clinicians and researchers might find themselves in.

In sum, it is incorrect to say that treatment options, including the use of coercion, are limited to, or directly based on the existence/nature of dysfunction. The factual question about dysfunction does have an indirect role in our value judgements about such cases, but it doesn't do all the work. There are questions of value to be addressed. We should embrace the value-ladenness of treatment, and address the ethical questions that it raises. There are crucial questions about when to pursue certain types of treatment, what kinds of intervention are warranted, and so on. We have made some attempt to outline what the harmful dysfunction account has to say about these questions, although we think this is an area that ought to be explored further.

Conclusion

In sum, we think that the attempt to link delusions to the harmful dysfunction account is a promising approach. One key aspect of such an account is that there can be treatable conditions, even where no dysfunction exists. This potential has particular importance in the study of delusions, because our understanding of the relevant functions and dysfunctions in the area of beliefs is still developing. Researchers may find themselves in a state of epistemic uncertainty on the question of whether a particular delusion is dysfunctional. Moreover, we think that the proposed cases of non-dysfunctional delusions seem like just the kind of symptoms that might be eligible to be treatable conditions. They fall within established categories of treatable conditions, and they have the potential for harm. Further work is needed to address the many questions that arise from this proposal.

We have suggested that the value-ladenness of questions over treatment should be acknowledged and embraced. We must remember that on the harmful dysfunction account, the question of harm (and thus treatment) is always thoroughly value-laden. We should thus be wary of any attempt to tie treatment directly to the factual question of the function-dysfunction boundary. Facts about dysfunction can inform value perspectives, but there are a range of ethical questions about how and when to treat delusions, including dysfunctional and non-dysfunctional delusions, as well as in cases where we remain uncertain. These questions deserve attention.

References

1. Wakefield, J.C. 1992. Disorder as harmful dysfunction: a conceptual critique of dsm-iii-r's definition of mental disorder. *Psychological Review* 99 (2): 232–247.
2. Miyazono, K. 2015. Delusions as harmful malfunctioning beliefs. *Consciousness and cognition* 33: 561–573.
3. Sakakibara, E. 2016. Irrationality and pathology of beliefs. *Neuroethics* 9 (2): 147–157.
4. Cosmides, L., and J. Tooby. 1999. Toward an evolutionary taxonomy of treatable conditions. *Journal of Abnormal Psychology* 108 (3): 453–464.
5. Del Giudice, M. 2016. The evolutionary future of psychopathology. *Current Opinion in Psychology* 7: 44–50.
6. Varga, S. 2017. Mental disorder between naturalism and normativism. Philosophy Compass 12 (6): 1–9.
7. Wakefield, J.C. 2007. The concept of mental disorder: diagnostic implications of the harmful dysfunction analysis. *World Psychiatry* 6 (3): 149–156.
8. Bortolotti, L. 2010. *Delusions and other irrational beliefs*. Oxford: Oxford University Press.
9. Association, A.P, et al. 2013. Diagnostic and statistical manual of mental disorders (DSM-5 R): American Psychiatric Publishers.
10. Coltheart, M. 2017. Confabulation and conversation. *Cortex* 87: 62–68.
11. Nisbett, R.E., and T.D. Wilson. 1977. Telling more than we can know: Verbal reports on mental processes. *Psychological review* 84 (3): 231.

12. Petrolini, V. 2017. What makes delusions pathological? *Philosophical Psychology* 30 (4): 502–523.
13. Davies, M., M. Coltheart, R. Langdon, and N. Breen. 2001. Monothematic delusions: Towards a two-factor account. *Philosophy, Psychiatry, & Psychology* 8 (2): 133–158.
14. Coltheart, M. 2001. Assumptions and methods in cognitive neuropsychology. In *The handbook of cognitive neuropsychology: What deficits reveal about the human mind,* ed. Brenda Raap, 3–21. New York: Psychology Press.
15. Chang, H. 2017. Epistemic iteration and natural kinds: Realism and pluralism in taxonomy. In *Philosophical Issues in Psychiatry IV: Classification of Psychiatric Illness,* eds. Kenneth S. Kendler and Joseph Parnas, 229-245. Oxford: Oxford University Press.
16. McKay, R., and L. Cipolotti. 2007. Attributional style in a case of cotard delusion. *Consciousness and cognition* 16 (2): 349–359.
17. Breen, N., D. Caine, M. Coltheart, et al. 2001. Mirrored-self misidentification: Two cases of focal onset dementia. *Neurocase* 7 (3): 239–254.
18. Tabb, K., Schaffner, K.F. 2017. Causal pathways, random walks, and tortuous paths: Moving from the descriptive to the etiological in psychiatry. In *Philosophical Issues in Psychiatry IV: Classification of Psychiatric Illness*, eds. Kenneth S. Kendler and Joseph Parnas, 342-360. Oxford: Oxford University Press.
19. Wakefield, J.C. 1999. Evolutionary Versus Prototype Analyses of the Concept of Disorder. *Journal of Abnormal Psychology* 108 (3): 374–399.
20. Clutton, P. Forthcoming. A New Defence of Doxasticism about Delusions: The Cognitive Phenomenological Defence, Mind & Language.
21. Coltheart, M., R. Langdon, and R. McKay. 2011. Delusional belief. *Annual review of psychology* 62: 271–298.
22. Sullivan-Bissett, E. 2017. Biological function and epistemic normativity. *Philosophical Explorations* 20 (1): 94–110.
23. Bayne, T. 2017. Delusions and the norms of rationality. In Rationality: Constraints and Contexts, eds. T-W Hung & T. Lane, 77-94. London: Elsevier.
24. Fodor, J.A. 2001. *The mind doesn't work that way: The scope and limits of computational psychology.* Cambridge: The MIT press.
25. Connors, M.H., Halligan, P.W. 2014. A cognitive account of belief: a tentative road map. *Frontiers in psychology* 5: 1-14.
26. Nairne, J.S., and J.N. Pandeirada. 2008. Adaptive memory remembering with a stone-age brain. *Current Directions in Psychological Science* 17 (4): 239–243.
27. Schacter, D.L. 2012. Adaptive constructive processes and the future of memory. *American Psychologist* 67 (8): 603.
28. Conway, M.A. 2005. Memory and the self. *Journal of memory and language* 53 (4): 594–628.
29. Klein, S.B., L. Cosmides, J. Tooby, and S. Chance. 2002. Decisions and the evolution of memory: multiple systems, multiple functions. *Psychological review* 109 (2): 306.
30. Suddendorf, T., and M.C. Corballis. 2007. The evolution of foresight: What is mental time travel, and is it unique to humans? *Behavioral and Brain Sciences* 30 (03): 299–313.
31. Ally, B.A., E.P. Hussey, and M.J. Donahue. 2013. A case of hyperthymesia: rethinking the role of the amygdala in autobiographical memory. *Neurocase* 19 (2): 166–181.
32. Parker, E.S., L. Cahill, and J.L. McGaugh. 2006. A case of unusual autobiographical remembering. *Neurocase* 12 (1): 35–49.
33. Diamond, D.M., and P.R. Zoladz. 2016. Dysfunctional or hyperfunctional? the amygdala in posttraumatic stress disorder is the bull in the evolutionary china shop. *Journal of Neuroscience Research* 94 (6): 437–444.
34. Kennair, L.E.O. 2010. The problem of defining psychopathology and challenges to evolutionary psychology theory. In *The evolution of personality and individual differences,* eds. David M. Buss and Patricia H. Hawley, 451-480. Oxford: Oxford University Press.
35. Nettle, D. 2011. Normality, disorder and evolved function: The case of depression. In *Maladapting Minds: Philosophy, Psychiatry, and Evolutionary Theory,* eds. Pieter R. Adriaens and Andreas de Block, 198–215. Oxford: Oxford University Press.
36. Wakefield, J.C., K.J. Pottick, and S.A. Kirk. 2002. Should the dsm-iv diagnostic criteria for conduct disorder consider social context? *The American journal of psychiatry* 159 (3): 380–386.
37. Wakefield, J.C. 2005. On winking at the facts, and losing one's hare: value pluralism and the harmful dysfunction analysis. *World psychiatry* 4 (2): 88–89.
38. Wakefield, J.C. 2005. Sexual dysfunction or pain disorder?: Dyspareunia from the perspective of the harmful dysfunction analysis. *Archives of Sexual Behavior* 34 (1): 52–57.
39. Nesse, R.M., Jackson, E.D. 2011. Evolutionary foundations for psychiatric diagnosis: making dsm-v valid. In (eds.) Pieter R. Adriaens and Andreas de Block, Maladapting Minds: Philosophy, Psychiatry, and Evolutionary Theory: 173–197.
40. Bolton, D. 2000. Alternatives to disorder. *Philosophy, Psychiatry, & Psychology* 7 (2): 141–153.
41. Nesse, R.M. 2001. The smoke detector principle. *Annals of the New York Academy of Sciences* 935 (1): 75–85.
42. Giudice, M.D., Ellis, B.J. 2016. Evolutionary foundations of developmental psychopathology. In *Developmental psychopathology Volume Two: Developmental Neuroscience,* ed. Dante Cicchetti, 1-58. New Jersey: John Wiley & Sons, Inc.
43. Wakefield, J. 2017. Taking people as they are: Evolutionary psychopathology, uncomplicated depression, and distinction between normal and disordered sadness. In *The Evolution of Psychopathology,* eds. Todd K. Shackelford and Virgil Zeigler-Hill, 37–72. Cham, Switzerland: Springer International Publishing.
44. Bolton, D. 2008. *What is mental disorder? An essay in philosophy, science, and values*. Oxford: Oxford University Press.
45. Cooper, R. 2005. *Classifying madness: A philosophical examination of the diagnostic and statistical manual of mental disorders*. Dordrecht: Springer.
46. Murphy, D. 2006. *Psychiatry in the scientific image*. Cambridge: The MIT Press.
47. Tsai, S.J., J.P. Hwang, C.H. Yang, and K.M. Liu. 1997. Delusional jealousy in dementia. *Journal of Clinical Psychiatry* 58 (11): 492–494.
48. Langdon, R. 2013. Folie à deux and its lessons for two-factor theorists. *Mind & Language* 28 (1): 72–82.

49. Arnone, D., A. Patel, and G.M.Y. Tan. 2006. The nosological significance of folie à deux: a review of the literature. *Annals of General Psychiatry* 5 (1): 11.
50. Silva, A., M. Ferrari, G. Leong, and G. Penny. 1998. The dangerousness of persons with delusional jealousy. *Journal of the American Academy of Psychiatry and the Law Online* 26 (4): 607–623.
51. Daulatabad, D., Sonthalia, S., Srivastava, A., Bhattacharya, S.N., Kaul, S., Moyal, D. 2016. Folie a deux and delusional disorder by proxy: an atypical presentation. *Australasian Journal of Dermatology* 58 (3): e113-e116.
52. Nielssen, O., R. Langdon, and M. Large. 2013. Folie à deux homicide and the two-factor model of delusions. *Cognitive neuropsychiatry* 18 (5): 390–408.
53. Reznek, L. 2010. *Delusions and the Madness of the Masses*. Plymouth: Rowman & Littlefield Publishers, Inc.
54. McKay, R.T., and D.C. Dennett. 2009. The evolution of misbelief. *Behavioral and Brain Sciences* 32 (06): 493–510.
55. Bortolotti, L., Sullivan-Bissett, E. 2017. How can false or irrational beliefs be useful? *Philosophical Explorations* 20 (S1): S1-S3.
56. Bortolotti, L. 2016. Epistemic benefits of elaborated and systematized delusions in schizophrenia. *The British Journal for the Philosophy of Science* 67 (3): 879–900.
57. Bortolotti, L. 2015. The epistemic innocence of motivated delusions. *Consciousness and cognition* 33: 490–499.
58. Meltzer, H.Y. 2012. Clozapine: balancing safety with superior antipsychotic efficacy. *Clinical schizophrenia & related psychoses* 6 (3): 134–144.
59. Lieberman, J.A., T.S. Stroup, J.P. McEvoy, M.S. Swartz, R.A. Rosenheck, D.O. Perkins, R.S. Keefe, S.M. Davis, C.E. Davis, B.D. Lebowitz, J. Severe, and J.K. Hsiao. 2005. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *New England Journal of Medicine* 353 (12): 1209–1223.
60. Moritz, S., R. Veckenstedt, S. Randjbar, F. Vitzthum, and T. Woodward. 2011. Antipsychotic treatment beyond antipsychotics: metacognitive intervention for schizophrenia patients improves delusional symptoms. *Psychological medicine* 41 (09): 1823–1832.
61. Horwitz, A.V., and J.C. Wakefield. 2012. *All we have to fear: Psychiatry's transformation of natural anxieties into mental disorders*. New York: Oxford University Press.
62. Horwitz, A.V., and J.C. Wakefield. 2007. *The loss of sadness*. New York: Oxford University Press.
63. Zachar, P. 2014. Beyond natural kinds: Toward a relevant scientific taxonomy in psychiatry. Kincaid and Sullivan (eds.)(2014): 75–104.
64. Freeman, D., G. Dunn, H. Startup, K. Pugh, J. Cordwell, H. Mander, E. Cernis, G. Wing-ham, K. Shirvell, and D. Kingdon. 2015. Effects of cognitive behaviour therapy for worry on persecutory delusions in patients with psychosis (wit): a parallel, single-blind, randomised controlled trial with a mediation analysis. *The Lancet Psychiatry* 2 (4): 305–313.
65. Wakefield, J.C. 2017. Addiction and the Concept of Disorder, Part 2: Is every Mental Disorder a Brain Disorder?. *Neuroethics 10 (1)*: 55-67.
66. Wakefield, J.C. 2013. The DSM-5 debate over the bereavement exclusion: Psychiatric diagnosis and the future of empirically supported treatment. *Clinical Psychology Review 33* (7): 825–845.
67. Wakefield, J.C. 2013. Addiction, the concept of disorder, and pathways to harm: comment on Levy. *Frontiers in psychiatry* 4: 1-2.
68. Wakefield, J.C., and M.B. First. 2012. Placing symptoms in context: the role of contextual criteria in reducing false positives in Diagnostic and Statistical Manual of Mental Disorders diagnoses. *Comprehensive psychiatry 53* (2): 130–139.
69. Wakefield, J.C. 2005. Disorders versus problems of living in DSM: Rethinking social work's relationship to psychiatry. In (ed.) Stuart Kirk, Mental disorders in the social environment: Critical perspectives: 83-95.
70. Wakefield, J.C. 2010. False positives in psychiatric diagnosis: implications for human freedom. *Theoretical medicine and bioethics 31* (1): 5–17.
71. Wakefield, J.C. 2008. The perils of dimensionalization: Challenges in distinguishing negative traits from personality disorders. *Psychiatric Clinics of North America 31* (3): 379–393.
72. Benn, P. 2001. Medicine, lies and deceptions. *Journal of Medical Ethics* 27 (2): 130–134.
73. Jackson, J. 2002. *Truth, trust and medicine*. New York: Routledge.
74. Hughes, J.C., Baldwin, C. 2006. *Ethical issues in dementia care: making difficult decisions.* London: Jessica Kingsley Publishers.
75. Matthes, J., and H. Caples. 2013. Ethical issues in using deception to facilitate rehabilitation for a patient with severe traumatic brain injury. *The Journal of head trauma rehabilitation* 28 (2): 126–130.
76. Lewis, M. 2017. Addiction and the brain: development, not disease. *Neuroethics* 10 (1): 7–18.
77. Wakefield, J.C. 2017. Can the harmful dysfunction analysis explain why addiction is a medical disorder?: Reply to marc lewis. *Neuroethics* 10 (2): 313–317.
78. Wild, T.C., A.B. Roberts, and E.L. Cooper. 2002. Compulsory substance abuse treatment: An overview of recent findings and issues. *European Addiction Research* 8 (2): 84–93.
79. Caplan, A. 2008. Denying autonomy in order to create it: the paradox of forcing treatment upon addicts. *Addiction* 103 (12): 1919–1921.
80. Ryan, C.J., G. de Moore, and M. Patfield. 1995. Becoming none but tradesmen: lies, deception and psychotic patients. *Journal of medical ethics* 21 (2): 72–76.
81. Teo, D.C., A.M. Abraham, and A.L. Peh. 2017. Folie à deux and Fregoli syndrome with greater severity in the 'secondary'–A case report. *Asian Journal of Psychiatry* 25: 254–255.
82. Series, L. 2015. Relationships, autonomy and legal capacity: Mental capacity and support paradigms. *International journal of law and psychiatry* 40: 80–91.