European Journal of Analytic Philosophy EuJAP | 2025 | Vol. 21 | No. 2 | 155-180
https://doi.org/10.31820/ejap.21.2.4

RECONCEPTUALIZING DELUSION: STRATEGY, DYSFUNCTION, AND EPISTEMIC INJUSTICE IN PSYCHIATRY

Eleanor Palafox-Harris[1] and Ema Sullivan Bissett[1]

[1] University of Birmingham, United Kingdom

Original scientific paper – Received: 02/09/2024 Accepted: 20/01/2025

This paper is part of a book symposium on Justin Garson's *Madness: A Philosophical Exploration* curated and edited by Elisabetta Lalumera (University of Bologna) and Marko Jurjako (University of Rijeka)

ABSTRACT

In his bold and illuminating book *Madness: A Philosophical Exploration*, Justin Garson makes a case for thinking about madness as strategy, rather than as dysfunction. The reader is invited to take away a better appreciation of the historical provenance of madness as strategy, that is, this is not a new idea, destined for the fringes or of interest only to those of a more radical bent. It is rather an idea which has firm roots in the history of psychiatry. Garson's lens is wide, he is advocating a strategy over dysfunction approach for, at least, anxiety, depression, schizophrenia (and its spectrum disorders), and delusion. In this exploratory paper, we focus on delusion. We discuss what a madness-as-strategy approach might say about delusion, and how that fits with the idea that such beliefs are evolutionarily adaptive. We turn then to explore the implications of this reconceptualization of delusion for epistemic injustice in psychiatry. Our discussions will support the idea that much of the theoretical action lies not in the distinction between dysfunction and strategy, but rather in the distinction between everyday and abnormal dysfunction.

Keywords: delusion; strategy; dysfunction; doxastic dysfunction; abnormality; epistemic injustice.

Correspondence: e.l.sullivan-bissett@bham.ac.uk

Introduction

Garson's vision is a bold one as he works through the history of psychiatry understood as containing two broad approaches to madness (dysfunction and strategy). We follow Garson in understanding the approaches in the following ways. The dysfunction approach has it that

> [W]hen someone is mad, it is because something has gone wrong inside of that person; something in the mind, or in the brain, is not working as it ought. Madness results from the breakdown of a well-ordered system; it is a defect or a dysfunction. It represents the failure of the system to achieve its natural end. (Garson 2022, 1-2)

The strategy approach on the other hand has it that

> [I]n the mad, we have a purpose being fulfilled, a movement toward a goal, a machine operating as it ought. (Garson 2022, 1)

In what follows we'll pull on the threads of this characterisation of the madness-as-strategy approach, in particular, we'll speak to the idea that some ways of being mad might be goal directed, or usefully thought of as *strategies*, without them being the outputs of evolutionary mechanisms *operating as they ought*. We focus on Garson's ideas as they might apply to delusion. Let us begin with the diagnostic criteria to get the phenomenon in sight. In the glossaries of the *DSM-IV* and the *DSM-5*, the definition of delusions is given as:

> A false belief based on incorrect inference about external reality that is firmly sustained despite what almost everyone else believes and despite what constitutes incontrovertible and obvious proof or evidence to the contrary. The belief is not one ordinarily accepted by other members of the person's culture or subculture (e.g., it is not an article of religious faith). When a false belief involves a value judgment, it is regarded as a delusion only when the judgment is so extreme as to defy credibility. (DSM-IV 2000, 765, DSM-5 2013, 819)

We won't go over the various ways in which this definition of delusion is contentious (for an overview see Sullivan-Bissett 2024a, 3-6). What is relevant for our purposes is that although the definition is, strictly speaking, compatible with the idea that delusions are (adaptive) strategies, it more naturally puts one in mind of delusions as outputs of dysfunction.

After all, the beliefs are described as "firmly sustained" (when, presumably, they ought to be revised in the face of "incontrovertible and obvious proof or evidence to the contrary"). It is common to focus on this feature of delusions and seek to explicate it by appeal to an abnormality in belief formation or evaluation (this is the research programme of two-factor theories of delusion). And the idea that delusions are instances of *pathology* also represents the current orthodoxy. Indeed, often this is taken as a datum, with the theoretical interest identified as providing a more precise characterisation of what the pathology looks like (see e.g. Bortolotti 2018; Miyazono 2015; Petrolini 2017, 2024; and Sakakibara 2016, cf. Bortolotti 2022; Ichino and Sullivan-Bissett 2024; Lancellotta and Bortolotti 2020).

Let us narrow our focus to *monothematic* delusions, which are those concerning a single theme which arise in otherwise healthy individuals (Coltheart, Langdon, and McKay 2007, 642). We have in mind examples of the following kind:

> Patrick has the delusion that his wife has been replaced by an imposter (Capgras delusion)
>
> Selina has the delusion that she has ceased existing (Cotard delusion)
>
> Kashmir has the delusion that Beyoncé is in love with him (erotomania)

The contents of these attitudes may strike one as sufficiently bizarre as to suggest that something is going very wrong with the subject. Together then with the diagnostic definition, the idea that delusions result from dysfunction rather than (adaptive) strategy is a natural one and is reflected in much of the contemporary research on the nature of delusion and its formation. We note this not in support of the dysfunction framework, but to give a sense of the terrain. Garson's approach then, at least in the context of delusion, marks a significant and interesting departure from much of the contemporary literature.[1] Let us first then take the terrain as the following (Figure 1):

[1] Garson primarily talks about delusions in schizophrenia, which are often polythematic and elaborated. However, we think that the idea that delusions are strategies rather than dysfunctions is more plausible for monothematic delusions, and, in any case, it is these delusions that have been discussed in the context of empiricism, a natural ally of the madness-as-strategy approach.

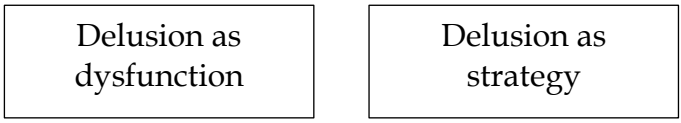

Fig. 1 Dysfunction versus strategy

1. Delusions as strategy

The idea of madness-as-dysfunction is, at least at first blush, natural, plausible, and if one is attracted to a Wakefield-style (1992) account of mental disorder, *conceptually necessary* (cf. Garson 2022, 248-9, 265). And, notwithstanding the clear strands of the madness-as-strategy approach in the history of psychiatry (for which Garson makes a magnificent case), the book, at the same time, leaves some of the finer details to the reader. Take the idea of delusions as strategy. We might well ask strategy *for what*? Garson gives some example answers from the history of psychiatry; from Griesinger (ibid., 90, 158, 184) and Freud (ibid., 4, 8) on wish fulfilment, to Johann Christian August Heinroth on delusions as coping strategies (ibid., 8), to the idea that a delusion is a "more or less deliberate diversion or escape from the unremitting tragedy of everyday life" (ibid., 125, see also 158, 169, 190), or that delusions help a person make sense of unusual feelings or experiences (ibid., 10, 226). We find this last suggestion the most promising, indeed, the idea of delusions as strategies to make sense of unusual feelings or experiences finds a home in some recent work on delusion formation.

Let us begin with predictive coding approaches (mentioned briefly by Garson, ibid., 260). According to these approaches, perceptual processing involves generating predictions about sensory input based on hypotheses about the world. Delusions are said to result from abnormalities in prediction error signalling. On what Eugenia Lancellotta has called "standard" predictive coding accounts, delusions are straightforward byproducts of a single dysfunction in prediction error minimization (2021, 56). There is no claim for delusions as strategy to be found here, let alone delusions as adaptive, indeed, Lancellotta refers to this view as the "maladaptive view of delusions" (2021, 49). However, some predictive coding theorists have conceived of delusions as generated by a *shear pin,* understood as a mechanism which is designed to break in certain conditions, to prevent further damage. A *doxastic* shear pin, that is, one at work in the machinations of belief, might be designed to break in particular circumstances (e.g. psychological distress) which would then allow for the

formation of strange beliefs that would not have been formed in normal circumstances (McKay and Dennett 2009, 501-2).

Sarah Fineberg and Philip Corlett have developed an account in these terms. In the face of the anomalous prediction errors, delusions are produced which explain the anomalous experiences and allow for the resumption of learning and engagement with reality. The delusion might be thought of as a strategy resulting from shear pin breakage which enables continued learning (as against cognitive resources being used up in making sense of the experience, at the expense of continued learning and engagement with the world).

We can also find support for the idea of delusions as strategies in some empiricist approaches, which have it that delusions arise, in part, from anomalous experiences. On endorsement versions of empiricism, the delusion is an *endorsement of* what is presented in experience, and the delusional content *is identical to* the content of the anomalous experience. On explanationist versions of empiricism, the content of the anomalous experience falls short of the content taken up in delusional belief, but the delusional belief is nevertheless thought to *explain* the anomalous experience (Bayne and Pacherie 2004, 3). We'll talk in explanationist terms, although we think that both endorsement and explanationist versions of empiricism could support a delusion-as-strategy approach insofar as delusions function as explanations of experience.[2]

Let us consider some examples. A subject may have visual and auditory hallucinations of a second head on her shoulder (a case of this kind is discussed in Ames 1984). One way of making sense of this experience is to form the belief that *there is a second head on my shoulder*, after all, that's exactly the kind of experience one might expect to have if the belief were true. Or consider the anomalous experience hypothesized to precede the Capgras delusion (the belief that *someone familiar has been replaced by an imposter*). It is thought that in such cases the subject has reduced affective response to familiar faces traceable to ventromedial prefrontal cortex damage (Tranel, Damasio, and Damasio 1995). One explanation of the lack of an expected affective response to the appearance of a loved one is that *that is not in fact one's loved one, but rather someone who looks a lot like them*.

We have seen some natural theoretical allies of the idea that delusions are *strategies.* Broadly speaking, they are strategies deployed in the context of

[2] Chenwie Nie has suggested that endorsement approaches to delusion can't cast delusions as *explanations* of anomalous experiences (see Sullivan-Bissett and Noordhof 2024, 2 in reply).

anomalous experiences which explain those experiences and keep the learning system moving. We have not yet said anything though about adaptation, something written into Garson's characterisation of madness-as-strategy. Garson is interested in the idea that delusions (among other madnesses) "could stem from mechanisms that are *performing their evolved functions perfectly well*" (Garson 2022, 251-2). Let us turn then to the idea of delusions as *adaptive* strategies.

2. Delusions as *adaptive* strategies

We are interested in the route from strategy to adaptation, between which we think there is significant water. In this section we'll explore this in the context of predictive coding and empiricism.

First, though, let us introduce Ruth Millikan's *Normal* (noting the capitalization) which will be important for some of what follows. *Normal* picks out a historical sense of normalcy rather than a statistical sense. Whilst it might be Normal for sperm to fertilize ova, it is not normal for them to do so (Millikan 1984, 34). There are Normal conditions for the proper performance of a functional item, and when a function fails to be performed, we might say that that's because the item was *dysfunctional*, or that the conditions for proper functional performance were abNormal. Now let us help ourselves to the well-worn claim that (in standard cases) mechanisms of belief production have the function of producing true beliefs (Papineau 1987; Millikan 1995). Delusions are, set against this background, clear candidates for dysfunctioning beliefs (Miyazono 2019). However, if mechanisms of belief production are operating in abNormal conditions for proper functioning, that is, in the presence of highly anomalous experiences, we have two more options for characterising them with respect to evolution. We might say that they are merely failing to perform their function due to abNormal conditions (Sullivan-Bissett 2024b), or we might say that those abNormal conditions are a trigger for a new function to be performed, and that delusions are adaptations.

One thing to note before exploring these options is that when we're thinking about delusions as responses to anomalous experience, at least in some cases, the idea that there is dysfunction *somewhere* is unavoidable. As we've noted, some experiences are hypothesized to arise from prediction errors or neurological damage, and no one wanting to argue that delusions are adaptative strategies to cope with anomalous experiences must thus deny that there's dysfunction *anywhere.* When Garson conceives of delusions as strategies to cope with anomalous experiences, he is not denying that that to which delusions are a *response* might arise from

dysfunction (Garson 2022, 260). This is why we will characterise the approach in opposition to delusion-as-(adaptive)-strategy as delusion-as-(*mere)*-dysfunction (Figure 2).

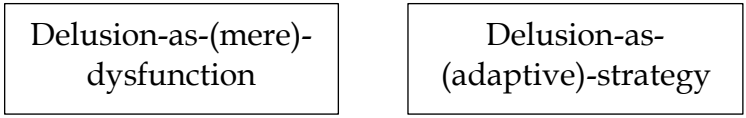

Fig. 2 (Mere) dysfunction versus (adaptive) strategy

Let us turn now to whether there's a route from delusions-as-*strategy* to delusions-as-*adaptive* in predictive coding and empiricist approaches.

2.1 Predictive coding and adaptation

John Matthewson and Paul Griffiths (2017) identify four ways in which a functional trait might *go wrong*, one of which is particularly instructive here: a trait does what it is supposed to do, but its Normal conditions for doing so are ones where something else has gone wrong for the organism (2017, 454-5). This is likely the kind of picture Garson has in mind when he speculates that delusions might have "an adaptive or functional role in compensating for perceptual abnormalities, or for yielding an appearance of meaning in a seemingly absurd or cruel world" (Garson 2022, 12). Perhaps forming a belief, as against not forming a belief, is adaptive, lest one live in the paralysis of uncertainty. This idea seems to be suggested in Garson's discussion of Snyder when he says "[d]elusions are the only thing that enable me to continue to navigate my environment; the delusions rescue me from madness" (Garson 2022, 226).

Some predictive coding approaches are a natural home for thinking in these terms. Indeed, the idea that delusions are *biologically*, rather than merely *psychologically* adaptive has only been defended in this context.[3] As we have seen, some versions of the predictive coding approach focus on the importance of resuming learning and engagement with the environment. On these approaches, delusions are adaptive insofar as they maintain behavioural interactions in the face of abnormal prediction-error signaling (Fineberg and Corlett 2016; Mishara and Corlett 2009). Or, as Garson puts it, delusions could be seen as "an attempt to buffer the mind from events or perceptual experiences that would otherwise totally disrupt our ability to get around the world" (Garson 2022, 252). We see then a route from strategy to adaptation.

[3] A similar approach for delusions in schizophrenia has been defended by Pablo López-Silva (2023), although from a phenomenological framework.

We make two points here. First, as noted earlier, it is more common for predictive coding accounts of delusion to cast them as by-products of a single dysfunction (Lancellotta 2021, 56). Accounts which cast them as adaptive responses which resume learning represent deviation from these "standard" predictive coding accounts. That's no objection of course, but we might ask which of the dysfunction or adaptation versions of the predictive coding approach to delusions is to be preferred. Lancellotta (2021) has argued that the dysfunction approach is superior to the adaptation account for several reasons (including that the former is simpler and more compatible with the available empirical evidence).

In addition, some philosophers who think delusions arise from *doxastic* dysfunction have identified prediction errors as the site of such dysfunction (Miyazono 2019, 65). Recall that someone who understands delusions as adaptive strategies needn't deny that they are responses to dysfunction elsewhere (e.g. in perceptual mechanisms). However, predictive coding accounts are built upon a denial of any sharp distinction between perceptual and doxastic mechanisms. Fineberg and Corlett for example propose a single impairment in prediction error, occurring in three stages: (1) *delusional mood,* in which "attention is drawn to irrelevant stimuli", (2) *delusion formation,* in which "explanatory insight occurs and flexible processing is disabled", and (3) *explaining things with the delusion,* in which the delusion becomes habitual and "enables patients to stay engaged with the environment and exploit its regularities" (Fineberg and Corlett 2016, 4). It looks, then, like the error can characterize either perception or belief. Indeed, Fineberg and Corlett themselves note that on their model "top-down and bottom-up processes sculpt one another" (2016, 5), and Corlett and Paul Fletcher (2014) suggest that prediction error dysfunction could result in deficits in both experience and belief. So although delusions could be an adaptive strategy in the face of dysfunction elsewhere (i.e. in perceptual mechanisms), given that predictive coding approaches deny a sharp distinction between perception and belief, it is difficult to isolate the delusional response from the dysfunction which characterises the single impairment in prediction error.

All told, predictive coding accounts which take delusions to be adaptive responses to anomalous experiences break from more standard predictive coding approaches, and face challenges arising from that break regarding theoretical complexity and consistency with the empirical evidence. It might also be difficult to characterise the delusion as formed by mechanisms of belief "*performing their evolved functions perfectly well*" (Garson 2022, 251-2), given that the approach does not sharply distinguish perceptual and doxastic mechanisms.

2.2 Explanationism and adaptation

Let us return now to explanationism, which could conceive of delusions as strategies to explain anomalous experiences. From here, we won't find an easy route to the idea that these strategies are *adaptive* ones. The orthodox position in the empiricist framework is the two-factor theory, which posits (1) an experiential abnormality and (2) an abnormality in belief. A delusion may well be formed as a strategy to make sense of anomalous experience, but it is a bad strategy, and the choice of a delusional content arises from abnormalities in mechanisms of belief formation or evaluation.

Part of what motivates two-factor approaches is that delusional explanations of anomalous experiences are *bad* explanations. Indeed, Max Coltheart and colleagues suggest that any theory of delusion formation should answer two questions:

> The first question is, what brought the delusional idea to mind in the first place? The second question is, why is this idea accepted as true and adopted as a belief when the belief is typically bizarre and when so much evidence against its truth is available to the patient? (Coltheart et al. 2011, 271)

Whilst all explanationists agree that the first question is answered by appeal to anomalous experience, the second question is taken by two-factor theorists to be answerable only by appeal to a second factor (an abnormality of belief) (Davies 2009, 72). Some go further, suggesting that delusional explanations for anomalous experiences are not only *poor,* they're "unintelligible" (Nie 2023; cf. Sullivan-Bissett and Noordhof 2024, 3-5), or "nonstarters", and "the explanations of the delusional patients are nothing like explanations as we understand them" (Fine et al. 2005, 160). Delusions may well be strategies, but they are bad ones.

We are not two-factor theorists (one of us has defended the one-factor approach elsewhere, Sullivan-Bissett 2020, 2024c; Noordhof and Sullivan-Bissett 2021, 2023). In our view, the selection of the delusional hypothesis can be explained by appeal to a range of normal (which is not to say good, or rational) ways of forming beliefs that we find across human psychology, and without appeal to doxastic abnormalities. However, something being within a range of statistically normal responses does not make it adaptive, and here we can usefully distinguish between *everyday dysfunction* and *abnormal dysfunction.* This distinction allows for a defence of delusions as strategies, without abnormalities in belief, but stops short of the idea that these strategies are *adaptive*.

Doxastic dysfunction of a minor kind is likely ubiquitous. It is not exactly rare for us to fail to update background beliefs in light of new evidence, miscount, be temporarily forgetful, misuse rules of inference, and so on. We think of these as cases of *everyday dysfunctions* of belief formation. Might delusions involve dysfunctions of this kind? We think so (it would be strange if delusions were immune to the kind of everyday dysfunction that causes error in belief). Delusions involving dysfunctions of this kind would place them alongside other irrational beliefs. It is also a familiar fact that there can be motivational influences on belief; such influences are most obviously at work in cases of self-deception, but they might also be of use in explaining the formation and maintenance of delusions with welcome contents (something Garson considers on page 169).

Delusions might be conceived of as strategies, but, we say, they are strategies mounted on a host of everyday doxastic dysfunctions and, in some cases, the epistemically inappropriate influence of motivation. A better strategy when faced with the uncertainties of an anomalous experience would be to form a non-delusional belief about the origins of the experience (e.g. *I am ill,* or *I am hallucinating*). The anomalous experiences with which people with delusions must deal might represent abNormal conditions for the proper performance of belief formation. However, even on this view, things aren't going well doxastically for the subject with a delusion, and delusion formation is helped along by various errors in reasoning that might also go into explanations of other kinds of irrational beliefs. It is thus difficult to see how delusions might be conceived of as stemming "from mechanisms that are *performing their evolved functions perfectly well*" (Garson 2022, 251-2).[4]

None of this is to deny that the delusional hypothesis may hold gifts not held by the non-delusional hypothesis. And the benefits that may accrue as part of delusion's sense-making have been recently discussed in a few contexts. For example, Rosa Ritunnano and colleagues have used their case study of Harry to show that some delusions can enhance a person's sense of *meaningfulness*, understood as "the extent to which one's life is subjectively experienced as making sense, and as being motivated and directed by valued goals" (Ritunnano et al. 2022, 110). This is no doubt important, both for a more comprehensive understanding of the nature of delusion and its impact on people's lives, as well as in the clinical setting. But there's no route here to a claim of adaptation.[5]

[4] A view of this kind need not be committed to the claim that anomalous experience is *sufficient* for the formation of a delusion. Human belief formation is broad and varied, and a whole range of individual differences can contribute to the formation of a delusion in response to an anomalous experience, *or not.*

[5] This point is made by Garson in an essay for Aeon (2022).

There is also the term *epistemic innocence*, which Garson suggests is used by Lisa Bortolotti and Ema Sullivan-Bissett "to describe the apparent *reasonableness* of some delusions, given the experiences that fostered them" (Garson 2022, 260). These authors do argue that the framework of *epistemic innocence* allows for a more nuanced assessment of the epistemic and pragmatic benefits of delusions, but they also deny that the delusion is overall epistemically *good* (Bortolotti 2015, 490; Sullivan-Bissett 2018, 924), and have it that delusions "compromise good functioning to a considerable extent" (Bortolotti 2015, 496).

Here then, is the theoretical terrain as we now see it (Figure 3).

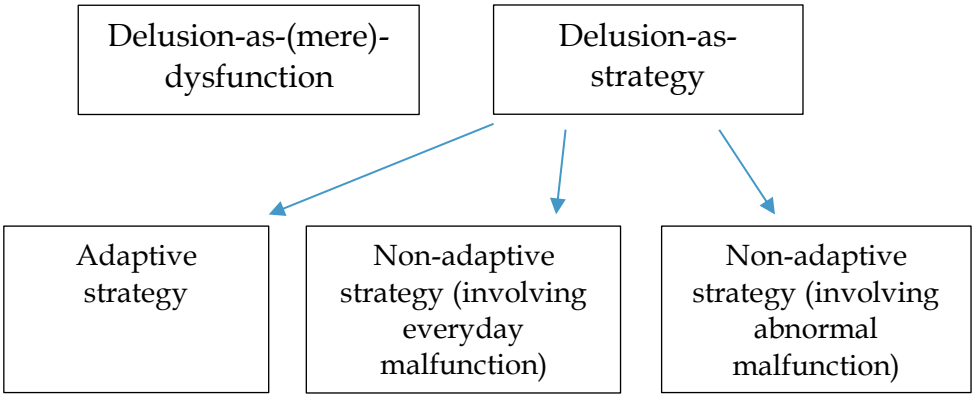

Fig. 3 Delusion as dysfunction versus three strategy approaches

Overall, when it comes to understanding the formation and maintenance of delusions, our view is that the distinction between *(mere) dysfunction* and *strategy* (the top row of Figure 3) is not where the theoretical goods lie. Rather, we should explore the kind of strategy in play, and, in particular, whether the doxastic dysfunctions prompting or facilitating its execution are *adaptive,* or involve *everyday* or *abnormal dysfunction* (the bottom row of Figure 3). We think this question is more illuminating with respect to the nature and role of delusion, and it is one on which major accounts of delusion formation might be distinguished.

Of course, even if delusions arise from doxastic dysfunction of some kind, they can nevertheless be considered worthwhile epistemic contributions to interpersonal exchange, and we suspect that this is something on which we and Garson agree, but also something which is not kept in firm view in psychiatry. We turn now then to epistemic injustice and delusion. Here we think there is theoretical fruit in the distinctions of both rows of Figure 3. That is, sometimes the distinction between *(mere) dysfunction* and *strategy* might make a difference to ameliorating epistemic injustice in psychiatric

encounters, whilst other times it is the distinction between *kinds of strategy* that makes the difference.

3. Delusion-as-strategy and epistemic injustice

Epistemic injustices, broadly construed, involve being wronged as an epistemic agent, in one's capacity as a *knower* (Fricker 2007, 20). Among other phenomena, this can involve having one's testimony unfairly discredited, distrusted, or discounted, as in *testimonial injustice*, having one's participation in epistemic practices unfairly undermined, as in *participatory injustice* (Hookway 2010), or having one's ability to make sense of and articulate one's own experiences unfairly constrained, as in *hermeneutical injustice*. In the literature exploring epistemic injustice and psychiatry, it is generally agreed that people with psychiatric conditions can and do experience various kinds of epistemic injustice in psychiatric contexts (e.g. Crichton, Carel and Kidd 2017; Scrutton 2017; although see Kious, Lewis and Kim 2023; Kidd, Spencer, and Harris 2023 in reply).

Epistemic injustices in psychiatry can be *interpersonal*, that is, perpetrated by healthcare practitioners or others in clinical contexts. They can also be *structural*: caused or maintained by unequal social or political dynamics. As well as these interpersonal and structural causes, Ian James Kidd and Havi Carel (2018) argue that epistemic injustices in healthcare (including psychiatry) can also be generated and exacerbated by the underlying *theoretical conception of health* which informs clinical practice and policy. Here we consider how the theoretical conception of delusion in play in psychiatric contexts might relate to the notions of *credibility*, *relevance*, and *intelligibility*, and how these affect the vulnerability of communicators with delusions to testimonial, participatory, and hermeneutical injustices.

We focus on the distinctions between delusion-as-(mere)-dysfunction and delusion-as-strategy, and between strategy accounts which posit everyday or abnormal doxastic dysfunction. The most commonly discussed cases of epistemic injustice relate to *credibility*, and here, we think the key distinction is the latter, that is, *between strategy accounts.* Nonetheless, when considering epistemic injustice related to notions of relevance and intelligibility, the broader distinction between delusion-as-(mere)-dysfunction and delusion-as-strategy can make a difference. All told, what is important for epistemic injustice in clinical encounters is not only whether we take delusions to be (mere) dysfunctions or strategies, but the details of the strategy at work.

3.1 Credibility in clinical interactions

In its most general formulation, testimonial injustice occurs when a communicator sustains an "*identity-prejudicial credibility deficit*" (Fricker 2007, 28), whereby they are perceived as having deflated credibility due to the operation of some identity prejudice. Consequently, their testimony is dismissed or distrusted. A key way in which identity prejudice deflates credibility is via stereotypes (Fricker 2007, 17).[6] Abdi Sanati and Michalis Kyratsous argue that people with delusions are frequently stereotyped as being "*bizarre*, *incomprehensible*, and *irrational*" (Sanati and Kyratsous 2015, 484), and are consequently vulnerable to testimonial injustices in clinical contexts. Here, we focus on *irrationality*, since this dimension of the stereotype relates most directly to credibility, although we touch on *bizarreness* and *incomprehensibility* in our discussion of intelligibility later (sect. 3.3).

According to Sanati and Kyratsous, the irrationality ascribed to people with delusions is not restricted to the irrational delusional belief(s), but is *generalised* to other beliefs the person holds (2015, 483). Thus, communicators with delusions are stereotyped as *globally epistemically irrational*. The irrationality that is stereotypically ascribed is not the unremarkable, everyday irrationality that even communicators without delusions exhibit. Instead, the stereotype should be understood as attributing an important difference in the *degree* or *kind* of irrationality characteristic of communicators with delusions (Palafox-Harris 2024, 261).

This stereotype deflates perceptions of testimonial credibility, as someone whose belief-system—indeed "their *general psychic life*" (Sanati and Kyratsous 2015, 484, our emphasis)—is characterised by global irrationality may not be perceived as an epistemically competent or credible communicator. If that's right, the testimony of people with delusions so stereotyped will not be given uptake in clinical interactions, and delusional communicators sustain testimonial injustices.

Let's have this basic picture in the background, as we turn to accounts of delusion and how they might impact on testimonial injustices related to assessments of *credibility.* If we think that what matters is the accuracy of the stereotype deflating credibility, then it is not (mere) dysfunction versus strategy which might make a difference here (because both kinds of

[6] For a recent discussion on the epistemic costs and benefits of stereotyping, see Puddifoot (2021) and the book symposium organized by Trakas (2025) in EuJAP.

approach, taken broadly, are consistent with the stereotype). What matters is whether the strategy involves everyday or abnormal dysfunction.

Let's start with strategy accounts which appeal to abnormal dysfunction, e.g. an account which explains the belief *that a loved one has been replaced by an imposter* by appeal to a cognitive deficit in belief evaluation. A conception of delusion that posits abnormality in belief lends theoretical justification to the credibility deficits communicators with delusions sustain in clinical interactions. Firstly, it lends theoretical support to the idea that people with delusions are characterised by irrationality that is different in kind or degree from the everyday irrationality of others. This is because positing an abnormal dysfunction in doxastic mechanisms provides a way to demarcate delusional irrational beliefs (those that *are* the product of abnormal doxastic dysfunction) from other unremarkable irrational beliefs (those that are *not* the result of abnormal doxastic dysfunction).[7] Thus, a conception of delusions as resulting from abnormality in belief reinforces the idea that people with delusions are irrational in a different way or to a greater extent than those without delusions, precisely because the irrationality associated with delusion can be traced back to an abnormal dysfunction in belief which people without delusions do not have.

Secondly, positing an abnormal doxastic dysfunction in those with delusions might also legitimise the *generalisation* of irrationality from one site (a particular delusional belief) to the presumed broader epistemic irrationality communicators with delusions are stereotypically taken to possess. It might be thought that an abnormal dysfunction in belief would lead to many of a person's beliefs being affected.[8] Conceptualising delusions as the products of abnormal doxastic dysfunction thereby sanctions the credibility deficits communicators with delusions can sustain in psychiatric contexts, because, on such a conception, people with delusions can be characterised as irrational (and thus, not credible) in a way which is global and remarkable (that is, different in an important respect from everyday irrationality). Therefore, we suggest that accounts

[7] This is not to say that a strategy account which posits abnormal dysfunction needs to be committed to delusions being the only kinds of belief that could arise from doxastic dysfunction. However, research on the etiology of other irrational beliefs (e.g. conspiratorial, self-deceptive, paranormal) has not proceeded by seeking to identify doxastic dysfunction. Rather these beliefs are theorised as arising from adaptive mechanisms, or as arising from normal range irrationalities or dysfunctions. For comparisons along these lines see Noordhof and Sullivan-Bissett (2023, 88-96) for monothematic delusion and paranormal beliefs, and Ichino and Sullivan-Bissett (2024) for monothematic delusions and beliefs in conspiracy theories.

[8] Accounts positing abnormal dysfunction can resist the stereotype of global epistemic irrationality, but this depends on the details of the second factor and delusion circumscription (Palafox-Harris 2024, 268-270).

of delusion which appeal to an abnormality in belief can contribute to testimonial injustices in psychiatry.

Let us turn to strategy accounts which deny abnormal dysfunction, e.g. an account on which the belief *that a loved one has been replaced by an imposter* can be explained as an unremarkable (which does not mean *rational*) response to weird perceptual experience, without appeal to cognitive deficit (or other abnormalities in belief). Such accounts do not support the stereotype that communicators with delusions are globally epistemically irrational, and therefore do not provide theoretical justification for downgrading the testimonial credibility of people with delusions. On an account that takes delusions as strategies for explaining anomalous experience without positing abnormal doxastic dysfunction, the irrationality of delusional beliefs is not different in kind nor in degree from other irrational beliefs. Instead, delusional beliefs can be explained by appeal to a range of *normal responses* to anomalous experience, utilizing cognitive processes which are "in no important respect different from those by which normal beliefs are formed" (Maher 1992, 262). A strategy approach of this kind does not reinforce the stereotype that people with delusions are globally epistemically irrational. If people with delusions are not irrational in a remarkable way, we avoid pre-emptively discrediting the testimony of delusional communicators on considerations of credibility.

Of course, testimonial injustices involving people with delusions might still occur in clinical interactions even if we take delusions to be strategies facilitated by everyday doxastic dysfunction. Delusional beliefs are, after all, irrational (even if they are irrational in an unremarkable way), and can contain content that is bizarre and difficult to comprehend, such as the belief *that I am dead* (a variation of Cotard delusion). Delusional testimony is therefore vulnerable to being distrusted or discounted even if we do not take delusions to result from abnormal doxastic dysfunction. Nevertheless, we suggest that if we're in the business of reducing vulnerability to testimonial injustice, it is not *(mere) dysfunction* versus *strategy* but *everyday* versus *abnormal* dysfunction within strategy approaches that could make the difference. That's because divergence on theoretical justification for the credibility deficits sustained by communicators with delusions is found *between strategy account*s, not between *(mere) dysfunction* versus *strategy*.

However, in clinical interactions, a communicator with a delusion may have her claims dismissed not only on the basis of irrationality (and therefore, of lacking credibility), but because her claims are deemed to be *irrelevant* or *unintelligible*. With respect to injustices arising from this, we

think that the distinction between *(mere) dysfunction* and *strategy* is key. Let us turn to that now.

3.2 Relevance in clinical encounters

Kidd and Carel argue that:

> A conception of disease is hermeneutically influential in two related ways: it affects which experiences can be candidates for discussion and interpretation and, secondly, shapes the forms of intelligibility applicable to them. (2018, 230)

Clinicians, scaffolded by medical institutions and the theoretical framework that informs clinical practice and policy, are the arbiters of what is *relevant* in clinical encounters. It is the clinician's job to glean which information is relevant to a patient's condition from sources such as the patient's testimony, testimony from others (e.g. family members), and their medical history. Naturally, what exactly is considered relevant to clinical practice will turn on the theoretical framework for health and ill-health that is at work. If we are the papists from the Middle Ages, then the patient's moral and spiritual character—their vices, their sins—are relevant to diagnosing and treating their malady (Garson 2022, 28-29). If, however, we are Kantians, then we need only investigate which of the mind's faculties are erring (Garson 2022, 79-94).

If delusions are mere dysfunctions, the clinician's role would be to diagnose and treat the dysfunction from which the delusion arises. On such an approach, there is little room for the patient's personal narrative or interpretation. If a delusion is simply the result of a broken doxastic mechanism inside the person with the delusion, then the clinician need not look further than the dysfunction to determine how best to treat them; the patient's subjective experiences and personal interpretations sit outside the bounds of clinical relevance, and need not be given clinical uptake.

This narrow conception of relevance affects the epistemic agency of a communicator with a delusion in at least two related ways. Firstly, it constrains the *scope* or *degree* of her participation in a given clinical encounter, as the clinician alone has "special authority to delineate and treat" mental disorders (Garson 2022, 240). If the rich subjective life of the person with delusions is not clinically relevant in any substantive sense, whilst the clinical interpretation of the subject's experiences is considered authoritative, the patient's perspective need not be solicited, appealed to, nor given uptake. Indeed, the extent of the patient's participation might be exhausted by "confirming biographical details or reporting symptoms"

(Kidd and Carel 2017, 181). Thus, people with delusions might be very restricted in what areas of clinical investigation they can participate in—they can *report* symptoms, but not *interpret* them—and in the degree to which their participation is sought or taken seriously by clinicians.

Secondly, this narrow sense of clinical relevance delimits the *kind* of participation a patient can perform. According to Christopher Hookway (2010), participating in collaborative epistemic projects "is not simply a matter of exchanging information" (2010, 156). Instead, participation involves asking questions, putting forward ideas, evaluating explanations, considering possible alternative explanations, and so on. By constraining clinical relevance in a way that excludes patient contributions beyond providing information about symptoms and history, people with delusions are precluded from performing the richer participative role Hookway describes. In so doing, delusion-as-(mere)-dysfunction takes delusional communicators as sources of information for clinical inquiry, instead of as *collaborators* in the epistemic pursuits of diagnosis and treatment.

Kidd and Carel (2017) argue that "ill persons (…) are typically regarded as the objects of the epistemic practices of medicine rather than as participants in them" (2017, 181). Therefore, a (mere) dysfunction-centric view of clinical relevance treats the person with delusions as an *epistemic object* (Fricker 2007; McGlynn 2021), or perhaps as a *truncated subject* (Pohlhaus 2014), but not properly as a *participant* in clinical encounters. In restricting patient participation in these two related ways, delusional communicators sustain *participatory injustices* (Hookway 2010). Both of these effects undermine the epistemic agency of the person with delusions, whilst bolstering the epistemic authority of the clinician, thereby consolidating the *epistemic privilege* of healthcare professionals (Carel and Kidd 2014) and unequal power dynamics in clinical interaction. Thus, (mere) dysfunction accounts of delusion undermine epistemic agency by tightly circumscribing what is *relevant* to the clinical encounter, to the exclusion of the patient's rich subjective experiences.

Alternatively, if delusions are conceived of as strategies, the clinician's role is to "discern the secret purpose that madness is trying to fulfil" (Garson 2022, 1) and "target the situation or event or arrangement" (Garson 2022, 12) to which delusions are a strategic response. This conception of madness carries a much broader notion of relevance to clinical practice, and we think this broadening comes with thinking in terms of mere strategy, regardless of what kind of dysfunction the strategy involves. The personal interpretations and experiences of the communicator with a delusion are not automatically discredited as irrelevant. Rather than being relegated to the periphery of clinical investigation (if not excluded

altogether), on a delusion-as-strategy framework the patient's rich subjective experiences are the very *substance* of clinical interest: in order to diagnose what the delusion is a response to, the clinician needs to look at what is happening in the patient's life, how they are experiencing their world, and how they make sense of their experiences.

This broader notion of relevance can scaffold the patient's epistemic status and agency in clinical encounters. Firstly, it means that *more* of the subject's testimony is clinically relevant compared to a (mere) dysfunction account; their personal experiences, interpretations, and meaning-making matters to a clinical inquiry that takes delusional beliefs to be *doing something*, that is, serving some purpose for the person who holds them, even if that purpose is helped along by everyday or abnormal dysfunction. In this way, the scope and degree of the person with delusion's participation is not so tightly circumscribed as on a conception of delusion as (mere) dysfunction. Secondly, conceptualising delusions as strategies allows for the patient to perform the richer participative role Hookway (2010) describes. If a patient's perspective is clinically relevant in a substantive way, then the patient is empowered to put forward their own interpretations, to participate in evaluating alternative interpretations, and to ask relevant questions. In this sense, delusional communicators are collaborators or co-investigators in the shared epistemic project of understanding their delusional beliefs, rather than the objects (or truncated subjects) of that epistemic project. Therefore, taking delusions as strategies for explaining anomalous experience can mitigate the participatory injustices people with delusions are vulnerable to in clinical encounters.

Moreover, we take the broader conception of clinical relevance contained in delusion-as-strategy to be more closely aligned than the narrow sense of relevance on delusion-as-dysfunction with frameworks like *phenomenology*, *co-production*, and *expertise-by-experience*, all of which take seriously the lived experiences of people with psychiatric conditions and promote epistemic agency. A strategy approach to delusion appears a natural ally to these more collaborative approaches to healthcare, which are often actively engaged in the project of ameliorating epistemic injustice in psychiatry (e.g. Ritunnano 2022; Carel and Kidd 2014).

3.3 Intelligibility in clinical contexts

As well as defining standards of clinical relevance, Kidd and Carel (2018, 230) emphasise that a conception of health and ill-health also defines standards of clinical intelligibility, that is, it sets the authoritative language for discussing medical conditions by producing the concepts, terms, and diagnostic categories used to interpret and talk about illness. In so doing, a

conception of health and ill-health can *hermeneutically marginalise* (Fricker 2007, 153) patients. Hermeneutical marginalisation occurs when a group is not afforded equal participation in the creation of shared hermeneutical resources. Patients are hermeneutically marginalised when the interpretative resources available to make sense of and talk about an experience of ill-health are produced by healthcare professionals, rather than by people with similar experiences.

Kidd and Carel argue that aspects of patient experience "cannot gain purchase within an exclusively naturalistic conception of health", which is equipped to interpret "physiological dimensions of the process of illness" rather than subjective experiential dimensions (Kidd and Carel 2018, 228-229). A similar thing might be said for (mere) dysfunction accounts of delusion. For example, claims of *meaning* in delusion are not easily understood against a theoretical backdrop according to which delusions simply are, or arise from, dysfunction. When the theoretical conception of delusion is ill-equipped to interpret subjective experience, such experiences might either be dismissed as unintelligible, or shoehorned into the language and style the theoretical conception is equipped to interpret. Kristen Steslow argues that adopting the medicalised language used by clinicians in order to render oneself intelligible involves "forsaking the uniqueness of [one's] own perspective, understanding, and expression" (2010, 30). In this event, the subject's interpretation is given clinical uptake, but something is lost in translation. In either case, whether the patient's interpretations are dismissed as unintelligible, or made intelligible only by translation into ill-fitting medicalised concepts, the person with delusions sustains a *hermeneutical injustice* (Fricker 2007, 155): they are unfairly hindered in their own interpretative efforts by a theoretical framework that cannot accommodate such efforts.

Indeed, the problem of unintelligibility might be particularly pronounced in cases of delusion, as the lifeworld of those with delusions might be radically altered from that of the clinician, and because delusional beliefs often have bizarre content which is difficult to comprehend. As we have seen, communicators with delusions are stereotyped as not only irrational, but also bizarre and incomprehensible (Sanati and Kyratsous 2015, 484). In this way, people with delusions have to contend not only with the "communicative roadblocks" generated by an unjust hermeneutical environment which privileges the interpretative resources of clinicians (Palafox-Harris 2024, 260), but also the incommunicability of delusional experience itself.

However, conceptualising delusions as strategies to explain anomalous experience might alleviate some of the apparent incomprehensibility of delusional content. Philip Gerrans argues that "delusions often seem to be

situated in chains of reasoning which, while incorrect, are *intelligible*" (2014, 113, our emphasis). A clinical approach that aims to "discern the secret purpose" (Garson 2022, 1) of a delusional belief and uncover the strategy in a delusion makes intelligible the *reasoning behind* the formation or maintenance of a delusion—reasoning which might be left obscure on an approach focused solely on diagnosing and treating dysfunctions.

Consider Capgras delusion, where someone believes *a familiar person has been replaced by an imposter*. Thinking of the Capgras delusion as a strategy for explaining anomalous perceptual experience makes the delusion more comprehensible. That is, even if the reasoning is epistemically *faulty* in some way (owing to everyday or abnormal doxastic dysfunction), we can nevertheless start to see why the particular delusional hypothesis was adopted or maintained. Understanding the strategy behind the delusion might help clinicians in understanding aspects of delusional experience and delusional content which a dysfunction-centric approach cannot make sense of. In this way, delusion-as-strategy can help to reduce the apparent incomprehensibility of delusions themselves.

Nevertheless, delusional communicators might still be hermeneutically marginalised in psychiatric contexts even on delusion-as-strategy, as we can imagine that patients would be prevented from participating *equally* in the creation of interpretative resources. However, if thinking of delusions as strategies reduces participatory injustices for the reasons we have discussed, then hermeneutical marginalisation may also be lessened. That is, a strategy approach may afford delusional communicators *more* participation in the production of shared interpretative resources than a dysfunction-centric approach. Moreover, an interpretative framework that takes delusions to be strategies for explaining or coping with anomalous experiences will be better equipped to comprehend patient interpretations of those experiences because of the standards of intelligibility such a framework carries. For example, claims that a delusional belief is *personally meaningful* are intelligible against a theoretical backdrop that assumes that delusions can serve a purpose for the person with delusions. If a particular patient interpretation is clinically intelligible, then it would not be dismissed or forced into other pre-existing concepts on the grounds of unintelligibility.

We suggest, then, that reconceptualising delusions as strategies reduces the susceptibility of delusional communicators to hermeneutical injustice in psychiatry for three reasons: (i) the standards of clinical intelligibility are equipped to comprehend patient attempts at interpretation and meaning-making, (ii) there is greater patient participation in creating hermeneutical

resources, and (iii) seeking to uncover and understand the strategy behind a delusional belief reduces the apparent incomprehensibility of delusions themselves.

Of course, adopting a strategy approach to delusion (and to psychiatry generally) does not make anyone immune from acting in epistemically unjust ways. Moreover, contingent features of healthcare contexts, such as time and economic pressures, unequal power dynamics, and broader socio-political factors might still obtain even if madness-as-strategy was adopted as the underlying theoretical conception of mental ill-health. Therefore, we do not suggest that reconceptualising delusions as strategies to explain anomalous experience would automatically and wholly mitigate the various epistemic injustices delusional communicators are vulnerable to. However, strategy approaches appear better equipped to promote the epistemic agency of people with delusions than (mere) dysfunction approaches.

4. Conclusions

We have explored conceiving of delusion as (mere) dysfunctions versus strategies. We found little mileage in the idea that delusions are *adaptive* strategies, but suggested that thinking of them as strategies finds a home in some contemporary approaches to delusion formation. Garson suggests that thinking in terms of madness-as-strategy rather than madness-as-dysfunction has interesting implications for "research and treatment itself, that is, the manner in which people are healed" (Garson 2022, 11). We have suggested, instead, that the key distinction with respect to the nature of delusion is in fact *between* strategy approaches, in particular, between those that appeal to abnormal malfunction versus those that appeal to everyday malfunction. Settling this is, for example, at the heart of the debate within explanationism about the number of *factors* involved in delusion.

We have also explored implications for epistemic injustice in psychiatry, and have suggested that the distinction between (mere) dysfunction and strategy approaches to delusion matters for epistemic injustices related to notions of clinical relevance and intelligibility (participatory and hermeneutical injustices). However, what matters for testimonial injustice is whether a strategy approach appeals to abnormal or everyday dysfunction. Therefore, reconceptualising delusions as strategies (broadly conceived) rather than arising from mere dysfunction is not enough to mitigate the credibility deficits communicators with delusions sustain. To focus solely on the distinction between delusion-as-(mere)-dysfunction

and delusion-as-strategy obscures important differences between strategy approaches which bear on considerations of credibility. Therefore, the alleviation or amelioration of epistemic injustices in psychiatry is best served not by a move from mere dysfunction to strategy, but rather to strategy *without abnormal doxastic dysfunction*.

In sum, when it comes to delusion, thinking in terms of *strategy* over *(mere) dysfunction* might make a difference to the amelioration of some epistemic injustices. However, overall, it is consideration of the *kind of strategy* involved in delusion which bears more theoretical fruit when considering questions of the nature of delusion and epistemic injustice in psychiatry.

Acknowledgments

We acknowledge the support of the Arts and Humanities Research Council for funding the work of which this is a part (Midlands4Cities Doctoral Training Partnership for Eleanor Palafox-Harris, and *Deluded by Experience,* grant no. AH/T013486/10 for Ema Sullivan-Bissett). Thanks also to Lisa Bortolotti and two reviewers for this journal for their comments on an earlier version.

REFERENCES

American Psychiatric Association. 2000. *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV-TR.* Washington, DC: American Psychiatric Association.

American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders. DSM-5*. Washington, DC: American Psychiatric Association.

Ames, David. 1984. "Self-shooting of a Phantom Head." *The British Journal of Psychiatry* 145 (2): 193-4.

Bayne, Tim, and Elisabeth Pacherie. 2004. "Bottom-Up or Top-Down: Campbell's Rationalist Account of Monothematic Delusions." *Philosophy, Psychiatry, and Psychology* 11 (1): 1-11.

Bortolotti, Lisa. 2015. "The Epistemic Innocence of Motivated Delusions." *Consciousness and Cognition* 33: 490-99.

Bortolotti, Lisa. 2018. "Delusion." In *The Stanford Encyclopedia of Philosophy*, edited by Edward Zalta, Summer 2022 edition.

Bortolotti, Lisa. 2022. "Are Delusions Pathological Beliefs?" *Asian Journal of Philosophy* 1 (31): 1-10.

Carel, Havi, and Ian J. Kidd. 2014. "Epistemic Injustice in Healthcare: A Philosophical Analysis." *Medicine, Health Care and Philosophy* 17: 529-540.

Coltheart, Max, Robyn Langdon, and Ryan McKay. 2011. "Delusional Belief." *Annual Review of Psychology* 62: 271-98.

Coltheart, Max, Robyn Landon, and Ryan McKay. 2007. "Schizophrenia and Monothematic Delusions." *Schizophrenia Bulletin* 33: 642-7.

Corlett, Philip, and Paul Fletcher. 2014. "Computational Psychiatry: A Rosetta Stone Linking the Brain to Mental Illness." *Lancet Psychiatry* 1 (5): 399-402.

Crichton, Paul, Havi Carel, and Ian Kidd. 2017. "Epistemic Injustice in Psychiatry." *BJPsych Bulletin* 41 (2): 65-70.

Davies, Martin 2009: "Delusion and Motivationally Biased Belief. Self-Deception in the Two-Factor Framework." In *Delusion and Self-Deception*, edited by Tim Bayne and Jordi Fernández, Jordi, 71–86. Psychology Press.

Fine, Cordelia, Jillian Craigie, and Ian Gold. 2005. "The Explanation Approach to Delusion." *Philosophy, Psychiatry, and Psychology.* 12 (2): 159-63.

Fineberg, Sarah K., and Philip R. Corlett. 2016. "The Doxastic Shear Pin: Delusions as Errors of Learning and Memory." *Cognitive Neuropsychiatry* 21 (1): 73-89.

Fricker, Miranda. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.

Garson, Justin. 2022. *Madness: A Philosophical Exploration.* New York: Oxford University Press.

Garson, Justin. 2022. "The Helpful Delusion." *Aeon. https://aeon.co/essays/evidence-grows-that-mental-illness-is-more-than-dysfunction*

Gerrans, Philip. 2014. *The Measure of Madness: Philosophy of Mind, Cognitive Neuroscience, and Delusional Thought*. Cambridge, MA: MIT Press.

Hookway, Christopher. 2010. "Some Varieties of Epistemic Injustice: Reflections on Fricker." *Episteme* 7 (2): 151-163.

Ichino, Anna, and Ema Sullivan-Bissett. 2024. "Conspiracy Beliefs and Monothematic Delusions: A Case for De-pathologizing." *Erkenntnis.* https://doi.org/10.1007/s10670-024-00881-w

Kidd, Ian J., and Havi Carel. 2017. "Epistemic Injustice and Illness." *Journal of Applied Philosophy* 34 (2): 172-190.

Kidd, Ian J., and Havi Carel. 2018. "Healthcare Practice, Epistemic Injustice, and Naturalism." In *Harms and Wrongs in Epistemic Practice*, edited by Simon Barker, Charlie Crerar, and Trystan S. Goetz, *Royal Institute of Philosophy Supplements:* 84, 211-233. Cambridge: Cambridge University Press.

Kidd, Ian J., Lucienne Spencer, and Eleanor Harris. 2023. "Epistemic Injustice Should Matter to Psychiatrists." *Philosophy of Medicine* 4 (1): 1-4.

Kious, Brent M., Benjamin R. Lewis, and Scott Y. H. Kim. 2023. "Epistemic Injustice and the Psychiatrist." *Psychological Medicine* 53 (1): 1-5.

Lancellotta, Eugenia. 2021. "Is the Biological Adaptiveness of Delusions Doomed?" *Review of Philosophy and Psychology* 13: 47–63.

Lancellotta, Eugenia, and Lisa Bortolotti. 2020. "Delusions in the Two-factor Theory: Pathological or Adaptive?" *European Journal of Analytic Philosophy* 16 (2): 37-57.

López-Silva, Pablo. 2023. "Minimal Biological Adaptiveness and the Phenomenology of Delusions in Schizophrenia." In *The Philosophy and Psychology of Delusions: Historical and Contemporary Perspectives* edited by Ana Falcato and Jorge Gonçalves, 126-140. New York: Routledge.

Maher, Brendan. 1992. "Delusions: Contemporary Etiological Hypotheses." *Psychiatric Annals* 22 (5): 260-8.

Matthewson, John, and Paul Griffiths. 2017. "Biological Criteria for Disease: Four Ways of Going Wrong." *Journal of Medicine and Philosophy* 42: 447-66.

McGlynn, Aidan. 2021. "Epistemic Objectification as the Primary Harm of Testimonial Injustice." *Episteme* 18 (2): 160-176.

McKay, Ryan, and Daniel Dennett. 2009. "The Evolution of Misbelief." *Behavioral and Brain Sciences* 32: 493-561.

Millikan, Ruth. 1984. *Language, Thought and Other Biological Categories*. Cambridge, MA: MIT Press.

Millikan, Ruth. 1995. "Explanation in Biopsychology." In *White Queen Psychology and Other Essays for Alice*, authored by Ruth Millikan, 171-192. Cambridge, MA: MIT Press.

Mishara, Aaron L., and Philip Corlett. 2009. "Are Delusions Biologically Adaptive? Salvaging the Doxastic Shear Pin." *Behavioral and Brain Sciences* 32 (6): 530-1.

Miyazono, Kengo. 2015. "Delusions as Harmful Malfunctioning Beliefs." *Consciousness and Cognition* 33: 561-73.

Miyazono, Kengo. 2019. *Delusions and Beliefs.* London and New York: Routledge.

Nie, Chenwei. 2023. "Revising Maher's One-Factor Theory of Delusion." *Neuroethics* 16 (15): 1-16.

Noordhof, Paul, and Ema Sullivan-Bissett. 2021. "The Clinical Significance of Anomalous Experience in the Explanation of Monothematic Delusions." *Synthese* 199: 10277-10309.

Noordhof, Paul, and Ema Sullivan-Bissett. 2023. "The Everyday Irrationality of Monothematic Delusion." In *Advances in*

Experimental Philosophy of Action, edited by Paul Henne and Sam Murray, 87-111. Bloomsbury.

Palafox-Harris, Eleanor. 2024. "Delusion and Epistemic Injustice." In *The Routledge Handbook of Philosophy of Delusion*, edited by Ema Sullivan-Bissett, 258-273. Oxon: Routledge.

Papineau, David. 1987. *Reality and Representation*. Oxford: Basil Blackwell Limited.

Petrolini, Valentina. 2017. "What Makes Delusions Pathological?" *Philosophical Psychology* 30 (4): 502-23.

Petrolini, Valentina. 2024. "Delusion and Pathology." In *The Routledge Handbook of Philosophy of Delusion*, edited by Ema Sullivan-Bissett, 33–45. Oxon: Routledge.

Pohlhaus, Gaire, Jr. 2014. "Discerning the Primary Harm in Cases of Testimonial Injustice." *Social Epistemology* 28 (2): 99-114.

Puddifoot, Katherine. 2021. *How Stereotypes Deceive Us*. Oxford: Oxford University Press.

Ritunnano, Rosa. 2022. "Overcoming Hermeneutical Injustice in Mental Health: A Role for Critical Phenomenology." *Journal of the British Society for Phenomenology* 53 (3): 243-260.

Ritunnano, Rosa, Clara Humpston, and Matthew R. Broome. 2022. "Finding Order Within the Disorder: A Case Study Exploring the Meaningfulness of Delusions." *BJPsych Bulletin* 46: 109-15.

Sakakibara, Eisuke. 2016. "Irrationality and Pathology of Beliefs." *Neuroethics* 9: 147-57.

Sanati, Abdi, and Michalis Kyratsous. 2015. "Epistemic Injustice in Assessment of Delusions." *Journal of Evaluation of Clinical Practice* 21 (3): 479-485.

Scrutton, Anastasia. 2017. "Epistemic Injustice and Mental Illness." In *The Routledge Handbook of Epistemic Injustice*, edited by Ian J. Kidd, José Medina, and Gaile Pohlhaus Jr., 347–355. Abingdon: Routledge.

Steslow, Kristen. 2010. "Metaphors in Our Mouths: The Silencing of the Psychiatric Patient." *Hastings Centre Report* 40 (4): 30-33.

Sullivan-Bissett, Ema. 2018. Monothematic Delusion: A Case of Innocence from Experience." *Philosophical Psychology* 31 (6): 920-47.

Sullivan-Bissett, Ema. 2020. "Unimpaired Abduction to Alien Abduction: Lessons on Delusion Formation." *Philosophical Psychology* 33 (5): 679-704.

Sullivan-Bissett, Ema 2024a. "Introduction." In *The Routledge Handbook of Philosophy of Delusion*, edited by Ema Sullivan-Bissett, 1-29. Oxon: Routledge.

Sullivan-Bissett, Ema. 2024b. "Monothematic Delusions are Misfunctioning Beliefs." *Synthese* 204, 157. https://doi.org/10.1007/s11229-024-04803-9.

Sullivan-Bissett, Ema. 2024c. "The One-factor Theory." In *The Routledge Handbook of Philosophy of Delusion*, edited by Ema Sullivan-Bissett, 414–29. Oxon: Routledge.

Sullivan-Bissett, Ema and Paul Noordhof. 2024. "Revisiting Maher's One-Factor Theory of Delusion, Again." *Neuroethics* 17 (1): 1-8.

Trakas, Marina. 2025. "Stereotypes Deceive Us, but Not in the Way We Commonly Think: Introduction to the Book Symposium." *European Journal of Analytic Philosophy* 21 (1): 1-8. https://hrcak.srce.hr/328465

Tranel, Daniel, Hanna Damasio, and Antonio R. Damasio. 1995. "Double Dissociation between Overt and Covert Face Recognition." *Journal of Cognitive Neuroscience* 7 (4): 425-32.

Wakefield, Jerome C. 1992. "The Concept of Mental Disorder: On the Boundary between biological Facts and Social Values." *American Psychologist* 47: 373-88.