Synthese (2024) 204:157
https://doi.org/10.1007/s11229-024-04803-9

ORIGINAL RESEARCH

Monothematic delusions are misfunctioning beliefs

Ema Sullivan-Bissett[1]

Received: 12 May 2022 / Accepted: 19 October 2024

Abstract

Monothematic delusions are bizarre beliefs which are often accompanied by highly anomalous experiences. For philosophers and psychologists attracted to the exploration of mental phenomena in an evolutionary framework, these beliefs represent—notwithstanding their rarity—a puzzle. A natural idea concerning the biology of belief is that our beliefs, in concert with relevant desires, help us to navigate our environments, and so, in broad terms, an evolutionary story of human belief formation will likely insist on a function of truth (true beliefs tend to lead to successful action). Monothematic delusions are systematically false and often harmful to the proper functioning of the agent and the navigation of their environment. So what are we to say? A compelling thought is that delusions are *malfunctioning* beliefs. Compelling though it may be, I argue against this view on the grounds that it does not pay due attention to the circumstances in which monothematic delusions are formed, and fails to establish *doxastic* malfunction. I argue instead that monothematic delusions are *misfunctioning* beliefs, that is, the result of mechanisms of belief formation operating in historically abnormal conditions. Monothematic delusions may take their place alongside a host of other strange beliefs formed in difficult epistemic conditions, but for which no underlying doxastic malfunction is in play.

Keywords Telological function · Malfunction · Misfunction · Normal conditions · Monothematic delusion · Belief

✉ Ema Sullivan-Bissett
e.l.sullivan-bissett@bham.ac.uk

[1] Department of Philosophy, University of Birmingham, Birmingham B15 2TT, UK

Published online: 22 November 2024

1 Preliminaries

Some of those attracted to naturalistic approaches in the philosophy of mind have sought to biologize belief, that is, to give an account of the way in which human beliefs are formed and maintained in an evolutionary framework, appealing to the biological functions of our mechanisms for belief formation. Biological functions here are of the historical (selected effects) or teleological[1] type (e.g. Millikan, 1984; Neander, 1991; Papineau, 1993). On such an approach, the function of a trait is whatever ancestral tokens of that trait type did which got them selected. It is taken to be one of the guiding virtues of historical accounts that trait tokens can have functions that they cannot perform, and historical accounts have been thought uniquely able to accommodate malfunction (cf. Davies, 2000; and replies in Franssen, 2009; Sullivan-Bissett 2017b).[2] By adopting a historical account of function then, we can ascribe functions to monothematic delusions even whilst recognizing that these functions are often not performed or even never performed (from here I often drop the 'monothematic' but should be read as referring to these delusions in particular, see Sect. 3). Because historical accounts divorce the possession of a function from any current day dispositions to perform or even capacity to perform said function, delusions can be really bad beliefs, but still have functions of belief.

What is meant by *mechanisms of belief formation*? Since my arguments turn on being able to at least roughly delineate these mechanisms and their functioning from (1) the broader functioning of cognition in general, and (2) the role of environmental factors, answering this up front is crucial. In a landmark paper, Ryan McKay and Daniel Dennett characterize the 'belief formation system' as 'an information processing system that takes certain inputs (e.g. perceptual inputs) and (via manipulations of these inputs) produces certain outputs (beliefs […])' (2009: 496). 'Certain inputs' will capture perceptual inputs, but also background beliefs, testimony, and other sources of information or evidence (depending on your epistemological persuasion) we take on board in forming beliefs. Talk of 'via manipulation of these inputs' picks out the various ways human beings form beliefs, through a process of mere endorsement of what is perceived, or with help from the influence of cognitive and motivational biases on belief, and also a range of ways a subject might interact with evidence (or their *epistemic style*, see Flores, 2021b for discussion). Imprecision here is permissible given my target conclusion. I am not committed to a detailed story of what

[1] See Garson (2019: 26–7) for concerns with using the term 'teleological' for describing the selected effects theory of biological function. I use 'teleological' here in the way he takes to be legitimate (even if not desirable), that is, as picking out a style of explanation.

[2] As opposed to non-historical accounts which focus on a token trait's present-day properties (e.g. Boorse 1976; Cummins 1975) or forward-looking dispositions (e.g. Bigelow & Pargetter, 1987). Such accounts tie function possession with the capacity to perform said function (or being disposed to contribute to inclusive fitness), thus leaving no gap between having a function and performing a function (and thus no possibility of malfunction). Another kind of view focuses on the modal properties of a trait token (Nanay, 2010). This view can accommodate malfunction, of a sort. Since Bence Nanay is not interested in functions at the level of trait *type* (but only *token*), he would not have anything to say about whether delusions are, in general, malfunctioning beliefs. At best he could say that a particular delusion was an instance of a malfunctioning belief, but if that judgement were grounded in any of the putative loci of malfunction to be discussed shortly, my arguments equally apply.

'manipulation of these inputs' looks like, but only to the claim that that manipulation, such as it is, is not properly described as malfunctional when it generates a delusion.[3]

Mechanisms of belief formation are prime candidates for a function-based treatment, given that beliefs produce changes to the way we navigate within and respond to our environment. Thus, adaptive pressures on their formation will be in play. A natural, well-worn, and perhaps obvious thought is that beliefs are adaptive when they are *true*, as Quine put it, 'creatures inveterately wrong in their inductions have a pathetic but praiseworthy tendency to die before reproducing their kind' (1985: 39). To accept this is not to accept that true beliefs are always adaptive (see notable examples in Stich, 1990; Stephens, 2001; Williams, 2021), but exceptions ought not deter us from accepting the claim that they usually or often enough are (Street, 2009: 235; Cowie, 2014: 4007; Hannon, 2019: 35).[4]

Once we recognize that true beliefs are adaptive, we are in a position to ascribe proper functions to the mechanisms which produce them. And again, we are in the territory of well-worn claims in suggesting that these mechanisms have the proper function of producing true beliefs (see for example Papineau, 1987; Millikan, 1995). Crucially, for historical theories of function, ascriptions of biological functions do not amount to reliable generalizations or even statistical likelihoods. It is only that in cases where functions are not performed (and such cases may well be ubiquitous), we are licensed to say that something has gone awry.

In what follows I help myself to the claim that our mechanisms of belief formation have the proper function of producing true beliefs.[5] This is a claim I have defended as part of a broader biological account of belief defended elsewhere (Sullivan-Bissett 2017a, 2018, 2020, *forthcoming*[b]). It is also one that anyone interested in the question of whether delusions are malfunctions or misfunctions of belief likely already has in the background. Going forward, let us call the thesis that delusions are malfunctions of belief the *malfunction thesis*, and the thesis that delusions are misfunctions of belief the *misfunction thesis.*

[3] This kind of picture of the mind is a modular one, which is a natural way of carving things up for those interested in the proper functioning of particular aspects of cognition. As Elselijn Kingma has noted, however, '[i]t may be the case that our minds operate in diffuse ways that do not lend themselves to easy carving into traits and functions, or that do not lend themselves to carving at all' (2013: 368). The point is a fair one. Nevertheless, here I follow the trend set by my predecessors in the discussion in ascribing functions specifically to mechanisms of belief formation.

[4] For a more comprehensive overview of evolutionary pressures on belief capacities see Sullivan-Bissett (*forthcoming*[a]).

[5] Not all philosophers understanding belief in terms of its proper function have ascribed a function in terms of *truth.* On Kate Nolfi's account, the constitutive proper function of belief 'is to inform our decisions to act by serving as a kind of map of the way things are so that we achieve whichever ends our actions aim to achieve' (Nolfi 2015: 197). My arguments do not require that mechanisms of belief formation have the function of producing true beliefs, although it is this function that is most obviously not performed in cases of delusion. Nevertheless, the claim that delusions are misfunctions rather than malfunctions of belief can tolerate some differences with respect to how the function of belief is understood.

2 Malfunction and misfunction

In characterizing *malfunction* we must take care not to do so in an overly permissive way, that is, as requiring only that a trait fails to perform its function. This is clearly inadequate– any account of function ought to allow a distinction between mere failure to perform and malfunction proper. In light of this, let us introduce some theoretical architecture which will frame the discussion, beginning with Ruth Millikan's sense of *normative historical* normality, signified with capitalization, and distinct from *statistical* normality. To demonstrate: sperm Normally fertilize ova, but it is not the case that sperm normally fertilize ova (Millikan, 1984: 34). This normative historical Normality gives us a way to contextualise the broader picture in which functionally characterized items operate (or not). That is, in cases of non-performance of a given biological function, we can say of a trait token that it *malfunctioned*, or that the circumstances for functional performance were abNormal, and that it merely failed to perform its function, or as I'll label it, *misfunctioned*.[6]

Millikan suggests that many putatively malfunctioning items should be captured under the heading of *mere failure to perform.* She emphasizes the contribution of the environment to proper functioning and suggests that the high rate of failure of biological designs is not down to breakage, but down to the absence of special conditions required for proper functioning (Millikan, 1994: 78, 2017: 84, 2024: 54). For her, malfunction proper takes place only in cases where 'there are abnormalities in the constitution of the device itself' (Millikan, 2013: 40; see also Garson 2017: 125; cf. Hundertmark & van den Bos, 2024: 10–13). For example, subjects with Holt-Oram syndrome have congenital heart defects caused by mutations in the TBX5 gene. This might be considered a case of *malfunction* (Sullivan-Bissett 2017b: 2512–13). In contrast, a heart which failed to pump blood during its time in a cold box ahead of transplantation would be failing to perform its function, due to being in abNormal conditions, but it wouldn't be malfunctioning (Davies, 2000: 32–3). This would be a case of *misfunction.*

When it comes to the proper performance of *cognitive* functions, Millikan claims that 'most cognitive failures are owing to outside conditions that are not Normal for the particular cognitive functions attempted' (Millikan, 2017: 84, see also 86). For Millikan, most cognitive failures are *misfunctions* rather than *malfunctions* of cognition. Let us now go forward with the malfunction/misfunction distinction in mind as we turn to biological function and belief.

[6] In addition to these two ways that functionally characterized traits might 'go wrong', John Matthewson and Paul Griffiths add two more. One is where 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). Delusions have been understood along these lines by Sarah Fineberg and Philip Corlett, who argue from within a predictive coding framework that the formation of a delusional belief is adaptive insofar as it allows for continued engagement with the world (2016: 3). The fourth way of going wrong is where an organism initiates an expensive developmental pathway in the presence of imperfect information (Matthewson & Griffiths, 2017: 455). I see little mileage in modelling the formation of a delusion in such terms. Thanks to Elselijn Kingma for drawing my attention to this work.

All of us carry with us a whole host of false beliefs. Some, plausibly, by design,[7] others not. For each of these cases, the function of true belief production has failed to be performed. Consider the following example: an authoritative source (say, an encyclopedia) asserts that *p.* After inspecting the encyclopedia's credentials, and checking for defeaters, I form the belief that *p.* However, the encyclopedia suffered an unfortunate typographical error in its most recent run, and printed *p* instead of the intended ~*p*. I have ended up with the false belief that *p.* This is a case of belief misfunction: my mechanisms for belief formation merely failed to perform their function of producing true beliefs—the environment did not cooperate, and the circumstances were abNormal for proper functional performance.[8]

What grounds the *abNormal* claim in this case? Something like this: the environment in which our mechanisms for belief formation were selected for producing true beliefs did not include misprinted encyclopedias, or more generally, did not include a sufficient number of accidentally unreliable sources mimicking reliable ones. Now, of course, if our environment had been abound with such cases, our mechanisms of belief formation *might* have been such that we were more discerning, or exercised more epistemic vigilance (if this additional cognitive work did not substantially slow the formation of attitudes which assisted our navigation of the environment). The encyclopedia case and others akin to it are not the kinds of condition to which we would need appeal in an explanation of the Normal operation of our mechanisms of belief formation. There is *no fault* to be found in their operation here, and so they are better considered cases of misfunction (mere failure to perform), rather than cases of malfunction.

Let us turn to malfunctioning belief. On one way of thinking about the matter, doxastic malfunction is a ubiquitous and everyday occurrence. Consider cases where we misuse rules of inference, miscount, are temporarily forgetful, fail to update background beliefs, and so on. Such errors might be explicable by appeal to fatigue, hunger, laziness, or stress, but they are errors nonetheless, on which, at least sometimes, perhaps, the blame cannot be lain at the door of an uncooperative environment.[9] Might delusional beliefs involve doxastic malfunction so understood? Surely! After

[7] I have argued elsewhere that we ought to ascribe a second function to our mechanisms of belief formation to accommodate cases where they seek to produce a belief because it is useful, but not insofar as it approximates to truth (Sullivan-Bissett 2017a, 2020, *forthcoming*[b]). I say no more about this since all sides agree that delusions are produced by mechanisms seeking to produce a true belief. The matter at issue is whether their failure to do so is a result of malfunction or misfunction.

[8] A different way of characterising the situation would be to index the possession of function to Normal conditions. Given this, the function of mechanisms of belief formation would be to produce true beliefs *in Normal conditions.* So when these mechanisms produce a false belief as a result of abNormal conditions, there is not in fact a function that they are failing to fulfil, and so no misfunction after all. For the project of this paper, characterizing things thus is merely terminological. My claim that delusions are not malfunctions of belief, but beliefs formed in abNormal conditions for proper functioning can equally be made in this alternative framework. I am grateful to Paul Noordhof for encouraging me to note this.

[9] A case could be made for the conditions facilitating mistakes of this kind to fall outside of mechanisms of belief formation and be counted as abNormal conditions (in which case, doxastic mistakes arising from fatigue, hunger, etc. would result in misfunctioning beliefs after all). Or they might be the result of limitations of the system, since the widely accepted principle that evolution does not generate optimally designed systems applies equally to our mechanisms of belief formation (McKay and Dennett: 2009: 497).

all, why suppose delusional beliefs in particular are immune from the everyday phenomena that cause mistakes in belief formation? If the malfunction thesis is to be understood in these terms– as suggesting that delusions are helped along by these everyday errors shared by non-delusional beliefs, that would be a fairly unremarkable thesis indeed. Characterizing delusions as malfunctioning beliefs in this sense would only place them alongside beliefs which are far more mundane in their epistemic imperfections. It would also not play the envisaged role in an account of the pathology of delusion (Miyazono, 2015, 2019: Ch. 3).[10]

The malfunction thesis then should be understood as something more substantial. We can capture the strength of the thesis by drawing a distinction within our category of malfunctioning belief between *everyday malfunction* and *abnormal malfunction.* There is no moving of the goal posts here since the malfunction posited is said to be grounded in a range of possible abnormalities. Let us be more precise about what is meant by *abnormality.*

There are broadly two ways of understanding *abnormality* in our discussion: functionally or statistically. We can take *functional normality* to pick out the property of being within the range of belief formation and evaluation styles which evolutionary selection has not distinguished, and functional abnormality the opposite. We can take *statistical normality* to pick out the property of occurring in non-delusional belief formation or evaluation, and statistical abnormality the opposite. Of course, often a belief that is abnormal in one of these ways will be abnormal in the other too, but they are separable, and we should be clear about which is in play when we're distinguishing everyday from abnormal malfunction. For example, misusing rules of inference and miscounting may well be statistically normal, but they may also be properly characterised as functionally abnormal and examples of everyday malfunction. But we have already said that the claim that delusions are malfunctioning beliefs is meant to be more substantial than this. In what follows then I will understand *abnormal malfunction* as identifying a functional abnormality against a statistical assumption (that is, functional abnormalities which are also statistical abnormalities are the ones constitutive of the category *abnormal malfunction*).

We can now separate those malfunctions that are everyday from those on which the claim that delusions are malfunctioning beliefs is based. In what follows instances of 'malfunctioning belief' should be read with the idea of *abnormal malfunction* in mind.[11]

[10] I'm grateful to Paul Noordhof for discussion on this material.

[11] The plausibility of possible candidates of malfunctioning belief so understood will come down to one's other commitments. Cases might be made for malfunctioning mechanisms of belief formation in cases of conspiratorial ideation or self-deception. Although alternative approaches abound. Jan-Willem van Prooijen and Mark van Vugt (2018) argue that conspiracy beliefs were adaptively advantageous in historical environments in which suspicion might be directed at powerful coalitions. Levy (2021) has argued that the mechanisms responsible for *bad beliefs* (a label under which most conspiracy beliefs fall) are the products of rational processes operating in unideal epistemic environments (cf. Williams, 2023). For self-deception, Trivers (2000, 2011, 2013) has argued that the capacity for self-deception is an evolutionary adaptation (to make us better interpersonal deceivers among other reasons), and Livingstone-Smith (2014) has it that it is the job of a sub-personal mechanism to selectively prevent the organism's representational apparatus from performing its proper function of accurate representation. On the other hand, Van Leeuwen (2007,

Now that we have our distinction between misfunction and (abnormal) malfunction in place, I say something about what hangs on it. After all, it might be asked what difference it makes if delusions are malfunctions or misfunctions of belief? Don't all sides agree that we find ourselves with a false[12] belief that is (often) bizarre, (often) resistant to counterevidence, which (often) impedes good functioning, and so on? Why would it matter whether that belief results from malfunction or misfunction?

There are, in fact, two implications of the present work. First, for those attracted to a view which sees delusions as continuous with other irrational beliefs (most robustly defended by Bortolotti, 2009), that view is much more difficult to maintain if delusions are malfunctioning beliefs. That's because a role for doxastic malfunction sets delusions apart from other irrational beliefs whose formation is not facilitated by anything so severe. Second, in some contexts it matters where we locate the fault. As Justin Garson points out, when biomedical researchers talk of a trait malfunctioning, 'they're often indicating, in a pragmatic kind of way, that the trait is an appropriate target for medical intervention' (Garson, 2019: 23). If there is a malfunction *of belief* in monothematic delusion, this might, depending on its nature, suggest particular kinds of intervention. On the other hand, if we think that the doxastic mechanisms are working well enough, but operating in an abNormal environment, we might focus our efforts instead on adjusting the environment or the subject's relationship to it.

3 Monothematic delusion

Let us start to situate our understanding of delusion in the foregoing discussion. Subjects with delusions are said to form these beliefs on evidence which does not properly support their content, to maintain them in the face of counterevidence, and delusions may also be incompatible or badly integrated with subjects' other beliefs (Bortolotti & Broome, 2008). These beliefs, as well as being epistemically faulty, can have serious pragmatic costs: they adversely affect wellbeing in various ways, by for example, interfering with one's relationships, and subjects may face social sanction arising from mistrust (Bortolotti, 2015: 493). These beliefs are, finally, almost always false (a feature which was taken to be definitional in DSM-IV (200: 765), and DSM-5 (2013: 819), with a move from 'false' to 'fixed' elsewhere in DSM-5 (2013: 87).

As noted earlier, my discussion will concern *monothematic* delusions, those involving a single theme and which can 'present in isolation in people whose beliefs are otherwise entirely unremarkable' (Coltheart et al., 2007: 642).[13] Also presumed

2008) has argued that this capacity is an evolutionary spandrel, a byproduct of other (adaptive) features of human minds.

[12] Of course, it is wise not to rule out the bare possibility of a true delusion. As Martin Davies and colleagues have argued, a belief which otherwise looks like a delusion ought not to be excluded from the category merely on the grounds that it is true (Davies et al. 2001: 133). My hunch is that delusions with true contents may only feature in imaginative philosophical thought experiments, but at the very least, actual world cases are sufficiently rare that we may put them aside for ease of exposition.

[13] I restrict my focus in this way because the debate between one- and two-factor theorists of delusion formation has taken place in the context of monothematic delusions and the identified factors posited by these views have been offered as providing potential loci for doxastic malfunction. (For explicit restriction

in what follows is doxasticism about these delusions (i.e. that they are *beliefs*), again following the convention set by the accounts discussed herein (for defences of doxasticism see Bayne & Pacherie, 2005; Bortolotti, 2009; Noordhof 2024[a]).

Examples of monothematic delusion include the Capgras delusion (the belief that *one's loved one has been replaced by a near-identical looking imposter*), the Cotard delusion (the belief that *one is dead* or *one has ceased existing*), mirrored self-misidentification (the belief that *the person in the mirror is not oneself*), and somatoparaphrenia (the belief that *a body part does not belong to one's body*). Delusions of this kind are often associated with some highly anomalous experiences. In Capgras delusion, subjects experience a lack of affective response to somebody with whom they are close. In Cotard delusion, this lack of affective response may be generalized to the environment, or there is an experience of depersonalization (see Gerrans, 2024 for discussion of the Cotard experience). In mirrored self-misidentification the subject doesn't experience the image in the mirror as herself, and in somatoparaphrenia, the subject does not experience a limb as part of her body.

According to empiricist accounts of delusion formation, delusions are grounded in anomalous experiences of this kind. Within empiricism, the explanatory reach of such experiences is contested, in particular, there is debate over how many *factors* we need to appeal to in order to explain delusion formation and maintenance. It is important for this debate in general, and indeed the arguments in this paper, to be really precise with respect to what is meant by *factor.* An appeal to factors is a way of picking out features of the context that are explanatorily relevant not to belief in general, but to delusional belief in particular. The project of identifying factors is not one which identifies all of the background features and cognitive contributions of belief formation. Rather, such a background is *read in*, and then one- and two-factor theorists seek to identify whatever *else* we need to explain the formation and maintenance of delusional belief in particular. Factors then are not merely causal contributions, and nor are they any of the various quirks of cognition that might go into an explanation of why folk have any number of strange beliefs (in e.g. the paranormal, conspiracy theories, and so on). These are simply part of the wide catalogue of propensities and tendencies in human psychology. If factor-theorists were committed to producing a causal inventory, they would make an appearance. But there is no such commitment.

A *factor* then is a contribution which is *abnormal.* This is recognized by one- and two-factor theorists alike. For example, elsewhere I defend a one-factor approach, and understand a factor as 'an abnormality that explains the formation of abnormal beliefs' (Noordhof & Sullivan-Bissett, 2021: 10279). Two-factor theorists signing up to this conception include Davies and colleagues who take the second factor to be 'a departure from what is normally the case' (2005: 228), Tony Stone and Andrew Young who talk of delusional reasoning being 'abnormal' and 'differences between people with and without delusions' (1997: 342), and Chenwie Nie who identifies a

to the monothematic case by one- and two-factor theorists see e.g. Coltheart et al., 2011: 282; Coltheart, 2013: 103; Coltheart & Davies, 2021: 225–6; Davies et al. 2001: 137; Davies and Coltheart, 2024: 430; Noordhof & Sullivan-Bissett, 2021: 10277; 2023: 87; Sullivan-Bissett, 2020: 679, 2024: 414).

factor as a 'departure from normality' (2023: 1, 13).[14] Finally, Philip Gerrans suggests that a defence of the one-factor approach could proceed by showing that the putative second factor describes 'a rationalization process which is within the normal range' (2002: 48). This of course would only be a defence of one factor over two if the second factor is proposed to constitute an abnormality.

We talked about how to understand *abnormality* earlier when we distinguished everyday from abnormal malfunction. It is the latter kind in play here, that is, I understand two-factor theorists as seeking to identify a functional abnormality against a statistical assumption (that is, functional abnormalities purported to be involved in delusion are also taken to be statistical abnormalities).[15]

With the above set out, we can now state the commitments of one- and two-factor approaches, which will be key to some of what follows. According to the one-factor approach, anomalous experience is the only abnormality to which we need appeal in giving an explanation of delusion formation and retention. The two-factor approach grants that anomalous experiences play a role in delusion, but also appeals to an additional anomaly in the shape of a bias, deficit, or performance error in mechanisms of belief formation or evaluation.

My discussion will take place in the empiricist context, but I briefly say something about two other approaches before proceeding. Rationalism denies anomalous experience a causal role in the formation of a delusion, and instead has it that 'there is a top-down disturbance in the subject's beliefs' (Campbell, 2001: 91) which can then affect experience. Rationalism might be a natural ally of the malfunction thesis, since the formation of a delusion is explained by appeal to *organic malfunction* (Campbell, 2001: 97). However, these are not the grounds on which the malfunction thesis has been explored or defended, and so I put rationalism aside.

Predictive coding accounts have it that perceptual processing involves generating predictions about sensory input based on hypotheses about the world. Such approaches might locate a malfunction in the processing of prediction errors (Miyazono, 2019: 91).[16] Some predictive coding accounts have it that aberrant prediction error signals cause problems in the allocation of attention in people with delusions (Corlett et al., 2010; Fletcher & Frith, 2009). This might result in paying undue atten-

[14] Although sometimes abnormality in the context of positing factors falls way, for example in Max Coltheart and Martin Davies's (2024) two-factor account of the Koro (shrinking penis) delusion.

[15] It might be thought that there is little clear water between the malfunction thesis and the two-factor account, after all, if I'm right, both require a statistically abnormal functional abnormality of belief to play a role in delusion. So might the project of this paper proceed more simply by assessing the merits of the two-factor view, rather than adjudicating between the malfunction and misfunction theses? In fact, there is sufficient water between the views to justify a project of this kind. That is because some of the potential sites of malfunction do not require the adoption of a two-factor model. Kengo Miyazono for example has suggested that we might identify doxastic malfunction in anomalous experience, and in doing so we need not presume a two-factor theory. In addition, a two-factor theorist might deny that for something to be a factor it must be an abnormality (implicit in Coltheart & Davies, 2024). I think there's very little mileage in either of these claims, but nonetheless they demonstrate that the malfunction thesis and the two-factor theory have different commitments.

[16] Although as noted earlier (fn. 6), some predictive coding theorists conceive of delusions as adaptive. For discussion of malfunction and adaptation versions of the predictive coding approach to delusions, see Lancellotta (2021).

tion to particular things or events (those which defy expectations), and a delusion is formed to explain those things or events. Predictive coding accounts are built upon a denial of any sharp distinction between perceptual and doxastic mechanisms. For example, Fineberg and Corlett distinguish themselves from one- and two-factor accounts when they note that on their account 'these two factors are strongly interrelated' and that 'top-down and bottom-up processes sculpt one another' (2016: 5).

None of what I've said in describing predictive coding approaches suggests that they cannot identify a mechanism responsible for belief and say that that mechanism malfunctions when it produces a delusion. However working with the specification of belief formation mechanisms given earlier, some of my arguments will turn on the distinction between perceptual mechanisms on the one hand, and belief mechanisms on the other. Predictive coding accounts deny a sharp distinction between perception and belief, and so although they can allow for malfunctioning belief in their framework, they do so by collecting together the perceptual and doxastic components of the relevant processes. My project recognises the virtues of separating out the operations of mechanisms of belief from the rest of our cognitive economy, and making judgements about how they are performing in particular.

A final methodological point before proceeding. I take the discussion which follows to be neutral on whether delusions constitute a natural kind. My view concerning the biology of delusions does not require that we impose (from the biological facts) any theory about the broader nature of these beliefs, and is perfectly consistent with a range of possible positions. If I'm right, delusions are best understood as beliefs formed in abNormal conditions. This need not be unique to the etiology of delusions, but that is fine, after all, there is no reason to insist that incorporating delusions into a biological picture must involve them in a unique origin story. Rather, if we take seriously (as I do) the idea that delusions sit on a continuum with other irrational beliefs, their biological underpinnings may well be shared by those other beliefs. Delusions may often be formed in abNormal conditions, but this may not be unique to delusions, and is not common to all delusions. My view is that delusions are produced by mechanisms of belief formation seeking to produce a true belief, and failing. To say this is *consistent* with delusions being a natural kind (as argued by Samuels, 2009), but it does not presume or require such a claim.[17]

Relatedly and furthermore, it pays to be open to heterogeneity in the class of delusion. It might be thought that the category is held together so loosely, precariously even, that it is hopeless to expect all tokens of the category to map onto a notion of misfunction or malfunction. This is fair, and my claim that delusions are misfunctioning beliefs should not be taken as ruling out, from the armchair, a case of delusion stemming from doxastic malfunction. There are no reasons, arising from the study of delusions, to think that they *could not* arise from doxastic malfunction. However, this possibility ought not be taken as providing insight into what we should say about delusions as they *in fact* manifest. Given what we know about the etiology of delusion, our default position ought to be that they can be accommodated in terms of *misfunction*. This is, I take it, especially friendly to the heterogeneity of delusion, since

[17] My impression is that proponents of the malfunction thesis need also not be committed to a natural kind claim for delusion, although the details of that non-commitment are for them to specify.

doxastic misfunction might arise for a number of reasons relating to the conditions in which the belief is formed. Indeed, the only commitment on the nature of delusion imposed by my view is that they are produced by mechanisms failing to perform the function of producing true beliefs. That's a very minor imposition on our more general theorizing.

4 Delusions as malfunctioning beliefs

Delusions might, at first glance, look like a clear case of something having gone profoundly awry with subjects' mechanisms of belief formation or evaluation. Delusions are characterized across various literatures as extreme cases of *belief gone wrong*. They are proposed as cases of severe irrationality, and it is often taken for granted that delusions are the paradigmatic case of *pathological belief* (see e.g. Bortolotti, 2018; Petrolini, 2017; 2024; Miyazono, 2015; Sakakibara, 2016; cf. Bortolotti, 2022). If one is attracted to integrating delusion into a broader programme of biologizing human cognition, before we even get to the details, the idea that they are malfunctioning beliefs may strike one as carrying serious prima facie plausibility. But let us get to those details and see if the initial plausibility can deliver.

With empiricist accounts of delusion formation as the backdrop, we can consider two possible sites for the presence of malfunction. The first is in the mechanisms responsible for anomalous experiences, and the second is in any one of a number of putative second factors (Miyazono, 2015: 566–7, 2019: 64–5).[18]

5 Delusions are not malfunctioning beliefs

In this section I will argue against the loci of doxastic malfunction noted above.

5.1 Anomalous experience

The first identified site of malfunction is in the anomalous experiences often associated with delusional beliefs. Not all delusions involve anomalous experiences (e.g. primary erotomania, for discussion see Coltheart, 2010: 24–5; Bell et al., 2008), but let us put such cases aside for the sake of argument.[19] Identifying the locus of malfunction here is problematic. Of course, some causes of anomalous experiences are properly characterized as malfunctions. For example, as noted earlier, in Capgras delusion subjects experience a lack of affective response when looking at someone with whom they are close. This has been traced to ventromedial prefrontal damage (Tranel et al., 1995; Coltheart, 2007). This damage may well be properly character-

[18] Miyazono in fact offers three possibilities in his account of delusions as doxastic malfunctions: anomalous experience, attention mechanisms in the predictive coding framework, and second factors. I do not discuss attention mechanisms since I put aside predictive coding approaches earlier.

[19] Coltheart suggests that erotomania might arise from an erroneous attribution of salience to events in experience (Coltheart, 2010: 24–5). If that's right, we retain continuity between erotomania and other monothematic delusions with respect to them involving something strange in experience.

ised as a malfunction. But we ought not overreach with our claim of malfunction to any cognitive consequences that might follow from it. There are not grounds for positing malfunction of *belief* from the fact that there is malfunction elsewhere in the cognitive architecture. It is consistent with normally functioning mechanisms of belief formation that the inputs result from malfunction in experience. These are not grounds for the claim that the resulting delusion is a malfunction *of belief.*

Indeed, Tim Bayne and Jordi Fernández say of one-factor accounts of delusion:

> Although experience-based accounts conceive of delusions as grounded in psychological malfunction, they see that malfunction as restricted to experiential mechanisms, broadly construed; on their view, delusion involves no damage to the mechanisms of belief formation as such. (Bayne & Fernández, 2015: 6)

This point isn't restricted to one-factor theories, but generalizes. A two-factor theorist might be attracted to the *malfunction thesis*, but wouldn't locate the relevant malfunction *in anomalous experience* for the reason I have just given. Malfunction in experiential mechanisms which can give rise to strange experiences, does not entail that beliefs formed downstream of such experiences are instances of malfunction. Consider an analogy: someone not familiar with a particular optical illusion ought not be said to have a malfunctioning belief when they form a belief based on the illusion's presumed veridicality (consider someone who doesn't know that sticks merely *look* bent when submerged in water). To take another example, a malfunctioning heart may well have downstream consequences for other physiological processes, involving, for example, the lungs. But it would be a mistake to locate the presence of malfunction in the lungs. By parity of reasoning, the *malfunction thesis* ought not base a claim of doxastic malfunction on malfunction in anomalous experience.[20]

Miyazono's version of the malfunction thesis might be thought to do better here, since his hypothesis is that 'delusions directly *or indirectly* involve some malfunctioning cognitive mechanisms' (2019: 4, my emphasis), where 'directly' maps belief-forming mechanisms and 'indirectly' maps mechanisms causally related to belief-forming mechanisms (2019: 4, fn. 6). So perhaps he would find nothing with which to disagree in the above. Rather, by locating the relevant malfunction in mechanisms responsible for anomalous experience, we would just have a case of *indirect* malfunction.

However, the idea that delusions involve malfunction *of some kind* is different from the idea that they are malfunctions specifically *of belief.* And this should be kept in mind as we turn to Miyazono's analogy: 'A delusion […] is analogous to a diseased or malformed heart. The category of belief, just like the category of heart, is defined in terms of distinctively belief-like functions' (Miyazono, 2019: 4, see also p. 105). Miyazono also identifies his central hypothesis as 'delusions are malfunctional

[20] As Garson points out, intuitions differ here. His own example is of a heart failing to perform its function due to a ruptured blood vessel. As noted earlier (Sect.2) malfunctions indicate the appropriateness of intervention (identifying the locus of a malfunction is to identify what needs to be fixed): 'if the heart cannot circulate blood because of a ruptured vessel, we don't want to fix the heart. We want to fix the artery!' (Garson 2017: 116). This is part of his larger argument for proper functions being proximal functions (Garson 2017: Ch. 7; Fagerberg and Garson *forthcoming*).

beliefs' (2019: 4), with the relevant functions constitutive of the biological category of belief including accurate representation and action guidance (2019: 12).

Now, of course, in almost all cases of delusion some of these functions are not performed, in particular, accurate representation. But a function not being performed by some mechanism is not the same as that mechanism *malfunctioning*, it could simply be *misfunctioning* (i.e. operating in conditions abNormal for proper functional performance). *That* is what we should say about cases Miyazono considers 'indirect malfunctions'. If there's a malfunction to be found in the realisers of anomalous experience which contributes to the formation of a delusion, that malfunction is not properly characterized as a malfunction *of belief.* Just as a lung may perform suboptimally due to a non-cooperative heart, so too might a belief fail to perform its function due to non-cooperative malfunctioning experiential mechanisms. In neither case should we ascribe malfunction to the biological items labouring under the consequences of distinct malfunctioning items. Talk of 'indirect malfunction' retains the language of malfunction at the expense of a more precise interpretation of what is going on.

5.2 Second factors

The second possible location for doxastic malfunction is in the putative second factor involved in delusion. There are broadly three routes to rebutting the idea that there is a second factor constitutive of doxastic malfunction. One can argue that (1) the proposed factor does not in fact characterize people with delusions, or one can argue that even if it does, it either (2) does not constitute an abnormality, or that it (3) does not constitute an abnormality *of the relevant kind.* Any of these routes would support the conclusion that the putative factor will not provide the loci for doxastic malfunction. For that the factor must be (1) present and (2) a statistically abnormal functional abnormality, and (3) a statistically abnormal functional abnormality *of belief*. We will see that all of the proposed factors I consider fail on at least one of these grounds.

5.2.1 Performance error

Some two-factor theories have it that the second factor is a *performance error*. That is, subjects with delusions have the capacity to, for example, form or evaluate beliefs appropriately, but they fail to put that capacity into practice (see e.g. Gerrans, 2001). Whether we have grounds for malfunction here will depend on the reason why the relevant capacity is inhibited. Suffice to say for now that performance error theorists have not appealed to malfunction in mechanisms of belief formation to explain the performance failure. Gerrans for example doesn't say much about what causes the performance failure (his interest in in establishing a failure of *performance* rather than *competence*), but suggests that the issue is 'possibly based in the cause of [the] anomalous experience' which is 'both extremely distressing and cognitively intractable' (Gerrans, 2001: 170). We have seen already that even if what lies behind the anomalous experience is properly characterized as malfunction, that does not justify a claim of malfunctioning belief.

Carolina Flores's recent work might also lend itself to an approach of this kind (indeed she recognizes similarities between her view and Gerrans's (2021a: 6303, fn. 8). Flores has argued that subjects with delusions have the capacity to respond to evidence against their delusion (2021a), but that the capacity is *masked* (2021a, for critical discussion see Noordhof 2024[b]: 312–14). More generally, she has appealed to masks to reconcile the idea that beliefs are constitutively evidence-responsive with the observation that many beliefs seem not to be so (Flores *forthcoming*). She takes motivational factors to be the 'central culprit' across cases, but also suggests that abnormal perceptual experiences may play this role (Flores *forthcoming*, fn. 20). Anomalous experiences might mask a subject's capacity to respond to counter-evidence to their delusion by functioning as reoccurring evidence *for* the delusion (Flores, 2021a: 6315).

It has, though, long been recognised that motivational influences play a role in delusion. Perhaps most obviously in cases where the content believed represents a desired state of affairs (e.g. erotomania or Reverse Othello delusion), where mechanisms of belief formation and maintenance familiar in explaining self-deception may be operative. But even in cases where someone believes something unwelcome, relief from the distress caused by anomalous experience as well as intellectual satisfaction might be had by a subject upon forming the delusion and *figuring things out* (Mishara, 2010: 10). These benefits may be held to tightly.

In addition, often the alternative hypothesis for explaining one's experience isn't motivationally neutral. Consider the case of Capgras, where accepting that the strange experience one has when looking at a loved one is neurobiological in origin might not be a 'particularly uplifting prospect' (Bortolotti, 2023: 59, see also 107). Furthermore, as Brendan Maher points out, 'the social costs and consequence of major decisions made under the influence of the delusion may create a situation in which it is very difficult for the patient to re-examine the belief and publicly reject it' (Maher 2006: 182).

Further details are of course needed, but a performance error approach could be built on considerations of this kind: anomalous experiences or motivational factors might *mask* the capacity to respond to evidence in subjects with delusions. The presence of masks, so understood, constitutes an abNormal condition for the proper performance of our mechanisms of belief formation. Here too then, we find no role for doxastic malfunction.

5.2.2 Reasoning biases

Other two-factor theories are put in terms of biases related to processing information. Broadly speaking, there are two ways of understanding talk of biases in this context. The first way is to understand these biases as ones which occur within the normal range. Now, although it may well be interesting to get clear on the kinds of influences involved in delusion formation, that people with delusions fall at a particular point in the normal range with respect to some form of reasoning is not going to give us a candidate locus for malfunction of the relevant kind. At best the influence of cognitive biases on belief formation would give us a case of everyday malfunction, but we said earlier that the malfunction thesis is committed to something more substantial than

this. Interest in biases so understood would mirror what goes on in the literatures on e.g. paranormal belief, or conspiracy belief, where researchers look for normal range biases and styles which contribute to the formation or maintenance of beliefs of this kind (see e.g. French & Wilson, 2007, Gagliardi *forthcoming*). But none of this work supports a claim of doxastic malfunction.

The second way to understand talk of biases is as ones which do not occur within the normal range or, weaker, they do but are exaggerated in subjects who have delusions. Philippa Garety and Daniel Freeman claim that 'there is growing evidence of reasoning and attributional biases in delusions which suggests they may display *systematic differences in cognitive processes* from those in the general population' (1999: 116, my emphasis). This is also how Miyazono understands his preferred second factor when he says 'when I refer to "the observational adequacy bias" […] I am talking about the relative strength of the biases rather than the presence of them' (2019: 91). Biases so understood may give us a malfunction of the kind we are after.

However, the case for various biases of the *systematic difference* kind is weak. For example, there is evidence that the 'jumping to conclusions' bias is present in schizophrenia, but little evidence that it is present in monothematic delusion, and differences in reasoning between subjects with delusions and those without are often found not to be statistically significant (e.g. McKay et al., 2007: 368–9; Brakoulias et al. 2008: 157, 161–2; Jacobsen et al., 2012: 12). Furthermore, Justin Sulik and colleagues have recently shown that when 'careless participants' are removed from the data, the relationship between holding delusion-like beliefs and jumping to conclusions is 'severely attenuated' or 'disappeared entirely'. That's because careless participants are coded as being high in delusion-like beliefs in comparison with diligent participants, and careless participants request to see fewer beads[21] (Sulik et al., 2023: 757, see also Ross et al., 2016).

Consider also various attributional biases hypothesized to characterise the reasoning of people with delusions. For example, there is some evidence that people with persecutory delusions are more likely to attribute the cause of negative events externally (i.e. to other people or circumstances) and attribute the cause of positive events internally (i.e. themselves), with depressive delusions involving the opposite pattern (Kaney & Bentall, 1989). Leaving aside the issue of whether attributional biases in fact characterise delusions of various kinds (see Langdon & Coltheart, 2000: 193–7 for discussion), do they represent a statistically abnormal functional abnormality? They do not. As Robyn Langdon and Coltheart argue, it is a common feature of our doxastic lives that one's mood can influence the beliefs we form: '[d]ifferences in attributional bias play a role in many non-clinical aspects of everyday life' (2000: 198, see also Flores, 2021b: 47 for discussion of mood and epistemic style).[22] In

[21] In the Beads Task (see e.g. Garety et al., 1991), subjects are presented with two opaque jars of beads containing two colours in opposing ratios (e.g. 80:20 and 20:80). They are asked to say when they're confident that they know which jar beads are being drawn from. Some studies have found that subjects with delusions request fewer beads before deciding on the jar, hence the charge of *jumping to conclusions.*

[22] Langdon and Coltheart take the presence of attributional biases to be commonplace, suggesting that 'it is part of the normal human condition […] to have an attributional bias which favours personal-level causal explanations over subpersonal-level causal explanations' (2000: 196–7).

addition, subjects with delusions 'are found with all styles of attributional bias and *without an abnormal degree of bias*' (2000: 198, my emphasis).

Finally, let us consider the bias towards observational adequacy, that is, 'the biased tendency to place more emphasis on incorporating new observations into [one's] belief system ('observational adequacy') than keeping [one's] existing beliefs as long as possible ('doxastic conservatism')' (Miyazono, 2019: 90). This can also not be pressed into service of the malfunction thesis. One issue is that some delusions involve observational data desperately in need of an explanation which people with delusions simply do not reach. Consider anosognosia, where, for example, a subject may deny that her left arm is paralyzed (when it is). Now consider the observational data available to her: she sees the arm inactive (whilst still detecting sensation within it), she fails to clap, etc. If there were a bias here, it ought not be described as privileging what is observed (Davies et al., 2005: 217–27). If delusions of this kind are to find a home in the malfunction thesis, it will not be via an abnormal bias towards observational adequacy. More broadly, if people with delusions had their beliefs biased in this way, we should expect them to be more often taken in by visual illusions. But there is no evidence that this is so (Davies & Coltheart, 2000: 25–7).[23]

Overall we have seen that the biases claimed as a second factor either do not characterize monothematic delusions (jumping to conclusions bias, bias towards observational adequacy), or do not characterize monothematic delusions to a more severe degree than they characterize non-delusional belief (attributional biases). These biases thus cannot provide a site of doxastic malfunction.

5.2.3 Belief evaluation deficit

Let us turn to the idea that the second factor is a belief evaluation deficit, a position which might seem a natural fit for the malfunction thesis. Langdon and Coltheart have suggested that there is a 'failure of normal belief evaluation' (2000: 184), and Coltheart has identified this as arising from right hemisphere damage in the frontal lobe, hypothesized to interfere with the belief evaluation mechanism (Coltheart, 2007: 1046; Coltheart et al., 2007: 644). If such a view were correct, we might have good grounds for the claim that there is a malfunction in the mechanisms of belief evaluation, indeed, it might be thought that the presence of such damage straightforwardly establishes malfunction of the relevant kind.

However, we ought not presume that the presence of neurological damage gives us the malfunction of belief we have been looking for. As we saw earlier in my discussion of anomalous experience, we shouldn't generalize from one site of malfunction to a malfunction *of belief*. Even if the damage was confirmed to be in the regions of the brain responsible for belief evaluation, further questions would need to be answered before we could take ourselves to have identified a *malfunction* in the

[23] Miyazono understands this bias though as operating within a predictive coding framework, that is, there is a prioritizing of prediction errors over prior beliefs (2019: 96). However, the arguments against this bias being involved in delusion are not solved by placing it in the context of prediction coding accounts and understanding it as privileging prediction errors over background beliefs. That's because the worry can be equally put in terms of the availability of opposing observational data with corresponding prediction errors.

mechanisms of belief evaluation. Putting aside issues arising from capturing certain cognitive processes by appeal to particular brain regions, is damage to a trait sufficient for malfunction? Accepting this would deliver strange results in those cases where an item is damaged but nevertheless functions entirely appropriately. For example, I might sustain a particularly grievous papercut to my thumb, but nevertheless retain a reasonable range of movement in it. Do we have grounds to say that my thumb exhibits a malfunction in this case, based on damage caused by the papercut? The intuitive notion of malfunction we began with required, at the very least, that a function wasn't performed. Physical damage to a trait may often proceed functional failure, but not always. Thus, the presence of physical damage to mechanisms of belief evaluation does not yet establish doxastic malfunction. For that, the damage would need to be responsible for a failure of belief evaluation.

And so we come back to the starting motivation for the deficit approach– that it can explain why people with delusions are especially bad at evaluating their delusional belief and responding to counterevidence to it. If true, this might be explained by a belief evaluation deficit, which might in turn be said to arise from the neurological damage sometimes observed.

However, there are no grounds for taking the resistance to counterevidence displayed by subjects with delusions to be deficient beyond what we find in the normal range, and hence indicative of doxastic malfunction. Even if poor belief evaluation is true of subjects with delusions, so too is it true for a range of beliefs. Indeed, Maher took this to be analogous to what occurs in scientific theory change, where better theories are resisted because they conflict with a scientist's commitment to her own theory (Maher 1974: 107). A range of other, non-delusional beliefs also display evidence-irresponsiveness, including conspiracy beliefs (Bortolotti, 2023: 64; Gold and Gold 2024), religious beliefs (Van Leeuwen, 2017: 55; Ichino 2024: 84–5),[24] and self-deceptive beliefs (Van Leeuwen, 2007: 422), but the psychological work seeking to understand the formation and maintenance of such attitudes does not take there to be something profound at work needed to account for resistance to counterevidence.

In addition, there is some evidence that people with delusions *do* respond to evidence (or at least, what they take to be evidence). In terms of supporting evidence, the anomalous experiences which are the basis of the delusional belief are, first, taken to be reliable by the subject, and second, continue to support the delusional belief. In terms of responding to counterevidence, the ways subjects interact with such evidence suggest that it is being *incorporated* rather than dismissed or ignored. They might reason that they are an exception to a generalization, bite the bullet on an implausible consequence of their delusion, or contrive stories to accommodate the counter evidence (Flores, 2021a: 6307–9). This behaviour might be epistemically unideal, but it hardly merits the claim of a *deficit* in belief evaluation.

There is also evidence for people with delusions exhibiting normal belief evaluation. Subjects with delusions are often able to evaluate their beliefs for plausibility

[24] Van Leeuwen and Ichino are non-doxasticists about religious attitudes, and the relationship between these attitudes and evidence is part of their case for non-doxasticism. Nevertheless, one doesn't have to be a non-doxasticist in this domain to accept that there is an epistemically problematic relationship between religious attitudes and evidence, just as there seems to be in some other kinds of belief.

(even if they cannot bring themselves to abandon them), which suggests that they are perfectly well able to process information regarding a belief's plausibility. For example, M. P. Alexander and colleagues report on a Capgras subject who, when asked what he would think were someone to tell his story, replied 'I would find it extremely hard to believe' (Alexander et al., 1979: 335). Some months later, the subject 'recalled discussing the implausibility of his story', and 'could rationally discuss the preposterous nature of his reduplication' (Alexander et al., 1979: 336). Davies and colleagues take this as one example which shows that at least some people with delusions display 'considerable *appreciation of the implausibility* of their delusional beliefs' (Davies et al. 2001: 149). Gerrans suggests of the subject in this case that '[h] is grasp of the distinction between what is rationally required to believe in his context and what he actually believes is intact' (Gerrans, 2001: 171).[25] The subject's having these abilities might not be totally irreconcilable with the idea that they also exhibit a deficit in belief evaluation, but it is at the very least unfriendly to such a position. We would need to make plausible the idea that the deficit in belief evaluation is robust enough to affect the evaluation of one's beliefs when it comes to deliberation over retaining them, but not so robust as to remove one's capacity for evaluating them for plausibility third-personally.

In addition, sometimes people with delusions *do* abandon their beliefs, and first-personal accounts reflecting on the abandonment often discuss the role of counter-evidence (which, presumably, fed into a process of normal belief evaluation) (Flores, 2021a: 6312–13). This should not happen if there is a belief evaluation *deficit.* Indeed, Flores suggests that it is unlikely that the capacity to rationally respond to evidence is lost during the period that the delusion is held, and then re-acquired (Flores 2021: 6312). Again, this might not be logically incompatible with the idea that there's a malfunction in the mechanisms of belief evaluation, but it's certainly an inconvenient truth for such a view. Perhaps it is consistent with the idea that there's a malfunction in one's mechanisms of belief evaluation at time t, that there is no such malfunction at time $t+1$ (and *that* is why delusional belief abandonment is possible), but if the malfunction in belief evaluation is constituted by a neural deficit, this story becomes a much harder sell.

For those attracted to a claim couched in terms of difference in degree, this may even be granted in the case of delusion, but there are significant mitigations not present in the case of other evidence-resistant beliefs. As already discussed, Flores argues that people with delusions retain the capacity to rationally respond to evidence but 'are rarely in the right (internal) conditions' to do so (Flores, 2021a: 6306). Such capacities can be *masked* by anomalous experience and motivational influences on

[25] More recently, Debbie M. Warman and Joel M. Martin (2006) investigated the relationship between delusion proneness and impaired cognitive insight. They found that delusion prone participants were 'more certain in their own judgments than those who are less delusion prone', but that those participants who were highly delusion prone 'demonstrated more willingness to acknowledge fallibility than those who were low in delusion proneness' (2006: 302). They also found that those highly delusion prone displayed higher self-reflectiveness, understood as the extent to which subjects are likely to recognise that they have jumped too quickly to a conclusion, as well as the possibility of making mistakes in situations for which there is more than one potential explanation. The authors note that this result was 'unexpected' and 'demonstrates the complex nature of both delusion proneness and introspection' (2006: 303).

belief formation and evaluation. We can even forego talk of masks and, taking into proper consideration the context in which subjects with delusions find themselves, simply ask the this question (following Noordhof & Sullivan-Bissett, 2021: 10301): is the belief evaluation in subjects with delusions *significantly worse* than we would expect *given the presence of anomalous experiences which can be profound*, *distressing*, *and so on*? The discussion in this section suggests that it is not.

5.3 No malfunction

Taking the prospects for locating a malfunction of belief in anomalous experience or the putative second factor as a whole: things do not look promising. With respect to the first, assuming the legitimacy of carving out perceptual mechanisms from those involved in belief formation and retention, identifying a malfunction in the former is to look in the wrong place if one's project is one of identifying *doxastic* malfunction. When we turn to putative second factors, we are closer, but still not close enough. Performance error accounts are simply not in the business of identifying a malfunction in belief– the explanatory resources they bring to bear are ones which retain doxastic capacities of various kinds. Bias theories will not give us malfunction if they characterize the biases as present in non-delusional people, and a systematic difference claim, which would support malfunction, has not been established. Finally, deficit theories require more than neurological damage to establish a malfunction in belief evaluation, they also need to show that delusions involve abnormally faulty belief evaluation. However, there are no grounds for claiming that the belief evaluation displayed by subjects with delusions is beyond the ordinary, and in addition, there is at least some evidence from the study of people with delusions which suggests that there is no fault in belief evaluation, abnormal or otherwise.

No doubt there are things the malfunction theorist may say in reply to the foregoing, and so I do not take myself to have refuted the position (and as noted earlier, I should not be read as claiming that delusions *couldn't* arise from doxastic malfunction). Rather, I intended only to raise some concerns about it to motivate the search for an alternative. Let us turn now to making good on the claim that delusions are, instead, misfunctioning beliefs.

6 Delusions are misfunctioning beliefs

Let's begin with a quick reminder of our starting point. I identified *mechanisms of belief formation* fairly imprecisely– as those mechanisms which produce beliefs following inputs such as those arising from perception, background beliefs, and so on. I then helped myself to the oft-defended claim that such mechanisms have the biological proper function of producing true beliefs, as well as the utterly uncontroversial claim that delusions are *false* beliefs (notwithstanding fn. 12). With that as the machinery, we had two options when it came to characterizing what was going on with mechanisms of belief formation when they produced a delusion: they were (1) malfunctioning, or (2) misfunctioning. The arguments so far do double duty. First, they reveal the inadequacies of option (1), and so, via disjunctive syllogism, support

the plausibility of option (2). But the attractiveness of option (2) goes beyond its being merely what is left when we reject (1). Rather, the considerations brought to bear on the inadequacies of (1) have helped us to already paint a fuller picture of (2). In this section then, drawing on the discussions so far, I fill in some remaining details.

Recall our background of empiricism, according to which delusions are based on highly anomalous experiences. As noted earlier, debate within empiricism has focused on how many factors we need to appeal to explain delusion formation and maintenance. On a one-factor approach, the only clinical anomaly to which we need appeal is anomalous experience, on a two-factor approach, there is an additional factor in the shape of a performance error, reasoning bias, or deficit in belief evaluation.

Even granting a role for performance errors, reasoning biases, or poor belief evaluation in delusion, if I'm right that the putative second factors do not amount to doxastic malfunctions, by my lights it follows that they do not amount to *factors* either. That's because the status of something's being a malfunction and its being a factor hinged on the same thing, namely, being a statistically abnormal functional abnormality. If performance errors, reasoning biases, or kinds of belief evaluation are not doxastic malfunctions, they're not factors either. The misfunction thesis is therefore not compatible with a two-factor approach. Nevertheless, later in this section I speak to how some of the research motivating two-factor accounts may find a home in the misfunction thesis.

Let us return to anomalous experiences, which are plausibly attributable to perceptual malfunction. We have seen that although they do not constitute a doxastic malfunction, these experiences play a key role in the misfunction thesis, specifically, they support the claim that delusions are formed in abNormal conditions. That is, the presence of anomalous experiences place mechanisms of belief formation in conditions abNormal for them to take in relevant inputs and produce a true belief.

Recall our earlier characterization of our *mechanisms of belief formation* as mechanisms which encompass the range of ways one might move from experience to belief, or maintain a belief once it is formed (again, talk of *epistemic styles* à la Flores, 2021b might be instructive here). A naturalistic telling of the origin of this range of routes to belief will not include their being applied to highly anomalous experiences. Rather, when a person with a delusion forms a belief based on an anomalous experience, they are applying perfectly ordinary and normal range ways of forming their beliefs to an extraordinary situation. Everyday doxastic malfunctions may also play a role. Delusions are a case of Millikan's 'cognitive failures' that are owed to 'outside conditions that are not Normal for the particular cognitive functions attempted' (Millikan, 2017: 84). The way in which delusional beliefs are formed are perfectly everyday, with the subject seeking to form a true belief. However, the uncooperative environment (understood as perceptual mechanisms producing anomalous experiences, or the presence of motivational influences) 'does not cooperate in the ways necessary to support the particular type of cognitive functioning attempted' (Millikan, 2017: 86).

It is key at this point to note that it is no part of the misfunction thesis that the formation and maintenance of delusions is a *rational* response to anomalous experience. I am not committed to the claim that delusions ought to be construed as *rational* beliefs. Rather, it is consistent with the misfunction thesis that subjects with delu-

sions exhibit great irrationality, but the irrationality they exhibit is not qualitatively different from the irrationality exhibited by many folk whose ways of forming and maintaining belief fall squarely within the normal range. Such irrationality though, manifests in abNormal conditions.

Let me say a little more about what talk of *normal range* is doing here. Consider a line along which we might plot various kinds of reasoners: in one position along that line we might locate scientists or philosophers (the reader should fill in with her epistemic role models). At another position we might locate conspiracy theorists. At another we find occultists, with a wealth of strange beliefs from alien abduction to telekinesis. We could also do things in terms of falling into line with logical principles. At one position we find those who do not commit the conjunction fallacy (they guess that Linda is a mere bank teller), and perhaps nearby those who understand the logic of the material conditional (and so pass the Wason selection task). At a different (more populated) position in that range we find folk who take it to be more likely that a conjunction is true rather than just one of its conjuncts (Linda is a *feminist* bank teller), and they don't understand (or at least can't apply) the truth table for a conditional. The normal range for human belief formation and maintenance will tolerate an enormous variety of belief formation practices. This should be unsurprising given a wealth of examples of polarization; it is hardly remarkable when we find that people have different beliefs. Indeed, that we do find this has been the basis for a wealth of research in social psychology and social epistemology, much of which seeks to map normal range contributions to belief formation and evaluation.

So what does this have to do with the present issue? Well, in developing the misfunction thesis, we need not take a view on where in the normal range folk with delusions fall. We should only say that people with delusions fall somewhere within that range. Such a range will tolerate a variety of responses to anomalous experiences, which might involve the influence of motivation, and crucially, all such responses need not be taken to arise from doxastic malfunction. Delusions arise not because subjects who have them cannot be plotted along a normal range of ways of forming and evaluating belief, not because their mechanisms of belief formation are *malfunctioning.* Rather, that (albeit broad) range of ways of forming and evaluating belief, is being applied to the abNormal conditions constituted by the anomalous experiences to which they are responding.

I said earlier that some of what motivates the two-factor approach can find a home in the misfunction thesis. Even if the various two-factor theories do not give us malfunction (or a second factor), they may nevertheless be picking out interesting patterns in the ways people with delusions form or evaluate their beliefs, and if so, that will naturally be relevant to a whole bunch of explanatory projects. One such project is how delusional belief formation and maintenance fits into an overall biological theory of belief. Let us return then to some of the candidate second factors and see how they can be seen to fit within the misfunction thesis.

Recall that some two-factor theorists have it that the second factor is an error in performance, not competence. On such a view, subjects with delusions retain the competence to, for example, evaluate their beliefs appropriately, but simply fail to put that competence into practice. As we saw earlier, on one such view, the performance failure is identified as being based in the cause of the anomalous experience.

For our purposes then, what we have is the tracing of a cognitive consequence from the abNormal conditions constituted by anomalous experience. Some doxastic capacity fails to be performed because conditions for performance are abNormal.

Other two-factor theories are put in terms of biases relating to processing information. Again, bias two-factor theorists might identify abNormal conditions for belief formation in the anomalous experiences to which the subject (biasedly) responds. As we saw earlier, *if* such biases do characterize delusional belief formation, they are not properly characterized as malfunctions of belief. Nevertheless, they are explanatorily relevant to a full picture of delusional belief formation, but it is only in concert with the abNormality brought about by anomalous experience that they are generative of delusional belief.

One example might be attributional biases, which might be thought a normal part of the way in which we form beliefs. Even granting that such biases play a role outside of delusional belief formation, they may nevertheless be of interest to a broader project of understanding how delusions come about. For example, Langdon and Coltheart suggest that such biases might 'help us to explain some of the individual variation in delusional content (e.g. Capgras delusion versus Cotard delusion) and may help us to explain why deluded individuals are not swayed by medical evidence of subpersonal causality' (Langdon & Coltheart, 2000: 198).

Research on reasoning and belief formation styles in delusion should not be understood as supporting the case for a second factor or doxastic malfunction, but can nevertheless further our understanding of the formation and maintenance of monothematic delusions, understood as beliefs formed in abNormal conditions.

7 Conclusions

I have argued that the case for monothematic delusions involving doxastic malfunction has not been made. Rather, monothematic delusions are best understood as misfunctions of belief. No belief is formed in a vacuum, and it is a familiar fact that we all fall short in our doxastic lives. When that happens, we standardly appeal to features of the context which might excuse epistemic bad practice, and thus retain a picture of ourselves and each other as broadly rational. People with delusions are not afforded the same grace, a particular injustice when the relevant context is an especially challenging one for ideal epistemic performance. When we fail to pay due consideration to the wider circumstances in which beliefs of this kind are formed, we risk talk of substantial cognitive failure where there is none. Better to understand monothematic delusions as cases of believing in difficult conditions, ones for which our cognitive resources were not designed. I conclude then that for those in the business of biologizing delusions, they should take them to be the outputs of mechanisms of belief formation operating in abNormal conditions. Monothematic delusions are misfunctioning beliefs.

Acknowledgements I acknowledge the support of the Arts and Humanities Research Council (*Deluded by Experience*, grant no. AH/T013486/10) for funding the work of which this is a part. I am grateful to audiences at the Oxford Epistemology Group at the University of Oxford, the Illusion, Delusion, and Hal-

lucination Summer School at Bonn's International Center for Philosophy, the Cognitive Diversity Seminar at the National Autonomous University of Mexico, the Visiting Speaker Seminar at Southampton, the Empirical Epistemology Network Workshop at the University of Stirling, and the Philosophy of Psychiatry Colloquium at Bielefeld University. Thank you to undergraduate students of my *Fantastic Beasts* module (Ashmethaa Ashokumar, Ethan Cobb, Patrick Edis, Andreea-Karla Manea, Charlotte Skye, and Milo Wakefield) for generously discussing a very early draft of this material with me. For helpful comments on previous versions of the paper I am grateful to Lisa Bortolotti, Kengo Miyazono, and three reviewers for this journal. Finally, thank you to Paul Noordhof with whom I have previously developed some of the ideas in this paper, and whose substantial objections on some of the material therein helped me improve it.

Declarations

Conflict of interest I declare that there is no conflict of interest.

References

Alexander, M. P., Stuss, D. T., & Benson, D. F. (1979). Capgras' syndrome: A reduplicative phenomenon. *Neurology*, *29*, 334–339.

American Psychiatric Association. (2010). *Diagnostic and statistical manual of mental disorders: DSM-IV-TR.*

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders.* 5th Ed.

Bayne, T., & Fernández, J. (2015). Delusion and self-deception: Mapping the terrain. In T. Bayne and J. Fernández (Eds.) *Delusion and self-deception* (pp. 1–21). Psychology Press.

Bayne, T., & Pacherie, E. (2005). In defence of the doxastic conception of delusions. *Mind & Language*, *20*(2), 163–188.

Bell, V., Halligan, P. W., & Hadyn, E. (2008). Are anomalous perceptual experiences necessary for delusions? *The Journal of Nervous and Mental Disease*, *196*(1), 3–8.

Bigelow, J., & Pargetter, R. (1987). Functions. *The Journal of Philosophy*, *84*(4), 181–196.

Boorse, C. (1976). Wright on functions. *The Philosophical Review*, *85*(1), 70–86.

Bortolotti, L. (2009). *Delusions and other irrational beliefss*. Oxford University Press.

Bortolotti, L. (2015). The Epistemic innocence of motivated delusions. *Consciousness and Cognition*, *33*, 490–499.

Bortolotti, L. (2018). Delusion. In E. N. Zalta (Ed.) *The Stanford encyclopedia of philosophy.* Summer 2022 edition.

Bortolotti, L. (2022). Are delusions pathological beliefs? Symposium on Kengo Miyazono's delusions and beliefs. *Asian Journal of Philosophy*.

Bortolotti, L. (2023). *Why delusions matter.* Bloomsbury.

Bortolotti, L., & Broome, M. (2008). Delusional beliefs and reason giving. *Philosophical Psychology*, *21*, 821–841.

Brakoulias, V., Langdon, R., Sloss, G., Coltheart, M., Meares, R., & Anthony, H. (2008). Delusions and reasoning: A study involving cognitive behavioural therapy. *Cognitive Neuropsychiatry*, *13*(2), 148–165.

Campbell, J. (2001). Rationality, meaning, and the analysis of delusion'. *Philosophy Psychiatry & Psychology*, *8*(2/3), 89–100.

Coltheart, M. (2007). Cognitive neuropsychology and delusional belief. *The Quarterly Journal of Experimental Psychology*, *60*(8), 1041–1062.
Coltheart, M. (2010). The neuropsychology of delusions. *Annals of the New York Academy of Sciences*, *1191*, 16–26.
Coltheart, M. (2013). On the distinction between monothematic and polythematic delusions. *Mind & Languag*, *28*(1), 103–112.
Coltheart, M., & Davies, M. (2021). Failure of hypothesis evaluation as a factor in delusional belief. *Cognitive Neuropsychiatry*, *26*(4), 213–260.
Coltheart, M., & Davies, M. (2024). Koro: A socially-transmitted delusional belief. *Cognitive Neuropsychiatry*, https://doi.org/10.1080/13546805.2024.2313474.
Coltheart, M., Langdon, R., & McKay, R. (2007). Schizophrenia and monothematic delusions. *Schizophrenia Bulletin*, *33*(3), 642–647.
Coltheart, M., Langdon, R., & McKay, R. (2011). Delusional belief. *Annual Review of Psychology*, *62*(1), 271–298.
Corlett, P. R., Taylor, J. R., Wang, X. J., Fletcher, P. C., & Krystal, J. H. (2010). Towards a neurobiology of delusions. *Progress in Neurobiology*, *92*, 345–369.
Cowie, C. (2014). In defence of instrumentalism about epistemic normativity. *Synthese*, *191*, 4003–4017.
Cummins, R. (1975). Functional analysis. *The Journal of Philosophy*, *72*(20), 741–765.
Davies, P. (2000). Malfunctions. *Biology and Philosophy*, *15*, 19–38.
Davies, M., Coltheart, M., Langdon, R., & Breen, N. (2001). Monothematic delusions: Towards a two-factor account. *Philosophy, Psychiatry, & Psychology*, *8*(2/3), 133–158.
Davies, M., & Coltheart, M. (2000). Introduction: Pathologies of belief. *Mind & Language*, *15*(1), 1–46.
Davies, M., & Coltheart, M. (2024). The two-factor theory. In E. Sullivan-Bissett (Ed.), *The Routledge handbook of the philosophy of delusion* (pp. 430–449). Routledge.
Davies, M., Davies, A., Anna, & Coltheart, M. (2005). Anosognosia and the two-factor theory of delusions. *Mind and Language*, *20*(2), 209–236.
Fagerberg, H. and Garson, J. (forthcoming). Proper functions are proximal functions. *British Journal for the Philosophy of Science*.
Fineberg, S. K., & Corlett, P. R. (2016). The doxastic shear pin: Delusions as errors of learning and memory. *Cognitive Neuropsychiatry*, *21*(1), pp. 73–89.
Fletcher, P. C., & Frith, C. D. (2009). Perceiving is believing: A bayesian approach to explaining the positive symptoms of Schizophrenia. *Nature Reviews Neuroscience*, *10*(1), 48.
Flores, C. (2021a). Delusional evidence-responsiveness. *Synthese*, *199*, 3–4.
Flores, C. (2021b). Epistemic styles. *Philosophical Topics*, *49*(2), 35–55.
Flores, C. (forthcoming). Resistant beliefs, responsive believers. *Journal of Philosophy*.
Franssen, M. (2009). The Inherent normativity of functions in biology and technology. In U. Krohs and P. Kroes (Eds.) *Functions in biological and artificial worlds: Comparative philosophical perspectives.* (pp. 127–46). Cambridge: MIT Press.
French, C., & Wilson, K. (2007). Cognitive factors underlying Paranormal beliefs and experiences. In S. Della Sala (Ed.), *Tell tales about the mind and brain: Separating fact from fiction* (pp. 3–22). Oxford University Press.
Gagliardi, L. (forthcoming). The Role of cognitive biases in conspiracy beliefs: A Literature review. *Journal of Economic Surveys*.https://doi.org/10.1111/joes.12604.
Garety, P. A., & Freeman, D. (1999). Cognitive approaches to delusions: A critical review of theories and evidence. *British Journal of Clinical Psychology*, *38*, 113–154.
Garety, P. A., Hemsley, D. R., & Wessely, S. (1991). Reasoning in deluded schizophrenic and paranoid patients: Biases in performance on a probabilistic inference task. *The Journal of Nervous and Mental Disease*, *179*(4), 194–201.
Garson, J. (2019). *What biological functions are and why they Matter*. Cambridge University Press.
Gerrans, P. (2001). Delusions as performance failures. *Cognitive Neuropsychiatry*, *6*(3), 161–173.
Gerrans, P. (2024). Cotard Syndrome: The experience of Inexistence. In E. Sullivan-Bissett (Ed.), *Belief, imagination, and delusion* (pp. 181–204). Oxford University Press.
Gold, I, and Gold, J, (forthcoming). Delusion and culture. In E. Sullivan-Bissett (Ed.), *The Routledge Handbook of the philosophy of delusion* (pp. 533–543) Oxon: Routledge.
Hannon, M. (2019). *What's the point of knowledge? A function-first epistemology*. Oxford University Press.
Hundertmark, F., & van den Bos, M. (2024). Biological functions and dysfunctions: A selected dispositions approach *39*(8). https://doi.org/10.1007/s10539-024-09944-2.

Ichino, A. (2024). Religious imaginings. In E. Sullivan-Bissett (Ed.), *Belief, imagination, and delusion* (pp. 81–106). Oxford University Press.
Jacobsen, P., Freeman, D., & Salkovskis, P. (2012). Reasoning bias and belief conviction in obsessive-compulsive disorder and delusions: Jumping to conclusions across disorders? *British Journal of Clinical Psychology*, *51*(1), 84–99.
Kaney, S., & Bentall, R. P. (1989). Persecutory delusions and attributional style. *British Journal of Medical Psychology*, *62*, 191–198.
Kingma, E. (2013). Naturalist accounts of mental disorder. In K. W. M. Fulford (Ed.), *The Oxford Handbook of Philosophy and Psychiatry* (pp. 363–384). Oxford University Press.
Lancellotta, E. (2021). Is the biological adaptiveness of delusions doomed? *Review of Philosophy and Psychology*, *13*, 47–63.
Langdon, R., & Coltheart, M. (2000). The cognitive neuropsychology of delusions. *Mind & Language*, *15*(1), 184–218.
Levy, N. (2021). *Bad beliefs: Why they happen to good people*. Oxford University Press.
Livingstone-Smith, D. (2014). Self-deception: A teleofunctional approach. *Philosophia*, *42*, 181–199.
Maher, B. (1974). Delusional Thinking and Perceptual Disorder. *Journal of Individual Psychology, 30*(1), 98–113.
Maher, B. (2006). The Relationship Between Delusions and Hallucinations. *Current Psychiatry Reports, 8*, 179–183.
Matthewson, J., & Griffiths, P. E. (2017). Biological Criteria for disease: Four ways of going wrong. *Journal of Medicine and Philosophy*, *42*, 447–466.
McKay, R., & Dennett, D. (2009). The evolution of misbelief. *Behavioral and Brain Sciences*, *32*, 493–561.
McKay, R., Langdon, R., & Coltheart, M. (2007). Jumping to delusions? Paranoia, probabilistic reasoning and the need for Closure. *Cognitive Neuropsychiatry*, *12*(4), 362–376.
Millikan, R. (1984). *Language, Thought and other Biological categories*. MIT Press.
Millikan, R. (1994). On unclear and indistinct ideas. *Philosophical Perspectives*, *8*, 75–100.
Millikan, R. (1995). Explanation in biopsychology. In R. Millikan (Ed.) *White Queen Psychology and Other Essays for Alice* (pp. 171–192). Cambridge, MA: MIT Press.
Millikan, R. (2013). 'Reply to Neander'. In D. Ryder, J. Kingsbury, and K. Williford (Eds.) *Millikan and her critics* (pp. 37–40) West Sussex: Wiley-Blackwell.
Millikan, R. (2017). *Beyond concepts. Unicepts, language, and natural information*. Oxford University Press.
Millikan, R. (2024). Teleosemantics and the frogs. *Mind & Language*, *1*, 52–60.
Mishara, A. L. (2010). Klaus Conrad (1905–1961): Delusional mood, psychosis, and beginning Schizophrenia. *Schizophrenia Bulletin*, *36*(1), 9–13.
Miyazono, K. (2015). Delusions as harmful malfunctioning beliefs. *Consciousness and Cognition*, *33*, 561–573.
Miyazono, K. (2019). *Delusions and beliefs.* Routledge.
Nanay, B. (2010). A modal theory of function. *The Journal of Philosophy*, *107*(8), 412–431.
Neander, K. (1991). The teleological notion of function. *Australasian Journal of Philosophy*, *69*(4), 454–468.
Nie, C. (2023). Revising Maher's one-factor theory of delusion. *Neuroethics*, *16*, 15.
Noordhof, P., & Sullivan-Bissett, E. (2021). The clinical significance of anomalous experience in the explanation of monothematic delusion. *Synthese*, *199*, 10277–10309.
Noordhof, P., & Sullivan-Bissett, E. (2023). The everyday irrationality of monothematic delusion. In P. Henne and S. Murray (Eds.) *Advances in Experimental Philosophy of Action*. (pp. 87–111) Bloomsbury.
Noordhof, P. (2024a) Delusion and doxasticism. E. Sullivan-Bissett (Ed.), *The Routledge handbook of the philosophy of delusion*. (pp. 292–307) Oxon: Routledge.
Noordhof, P. (2024b). Delusion and doxasticism. E. Sullivan-Bissett (Ed.), *The Routledge handbook of the philosophy of delusion* (pp. 308–323). Oxon: Routledge.
Papineau, D. (1987). *Reality and representation*. Basil Blackwell Limited.
Papineau, D. (1993). *Philosophical naturalism*. Blackwell.
Petrolini, V. (2017). What makes delusions pathological? *Philosophical Psychology*, *30*(4), 502–523.
Petrolini, V. (2024). Delusion and Pathology. In E. Sullivan-Bissett (Ed.), *The Routledge handbook of the philosophy of delusion* (pp. 33–45). Routledge.
Quine, W. V. O. (1985). Natural kinds. In H. Kornblith (Ed.) *Naturalizing epistemology* (pp. 57–76). Cambridge, MA: MIT Press.

Ross, R. M., Pennycook, G., McKay, R., Gervais, W. M., Langdon, R., & Coltheart, M. (2016). Analytic Cognitive style, not delusional ideation, predicts data gathering in a large beads task study. *Cognitive Neuropsychiatry*, *21*(4), 300–314.
Sakakibara, E. (2016). Irrationality and pathology of beliefs. *Neuroethics*, *9*, 147–157.
Samuels, R. (2009). Delusion as a Natural Kind. In M. Broome, & L. Bortolotti (Eds.), *Psychiatry as Cognitive Neuroscience: Philosophical perspectives* (pp. 49–80). Oxford University Press.
Stephens, C. L. (2001). When is it selectively advantageous to have true beliefs? *Philosophical Studies*, *105*(2), 161–189.
Stich, S. (1990). *The fragmentation of reason* London: MIT.
Street, S. (2009). Evolution and the normativity of epistemic reasons. *Canadian Journal of Philosophy*, *39*, 213–248.
Sulik, J., Ross, R. M., Balzan, R., & McKay, R. (2023). Delusion-like beliefs and data quality: Are classic cognitive biases artifacts of carelessness? *Journal of Psychopathology and Clinical Science*, *132*(6), 749–760.
Sullivan-Bissett, E. (2018). Explaining doxastic transparency: Aim, norm, or function? *Synthese*, *195*(8), 3453–3476.
Sullivan-Bissett, E. (2020). We are like American robins. In S. Stapleford and K. McCain (Eds.) *Epistemic Duties: New Arguments, New Angles.* (pp. 94–110) New York: Routledge.
Sullivan-Bissett, E. (2024). The one-factor theory. In E. Sullivan-Bissett (Ed.), *The Routledge handbook of the philosophy of delusion* (pp. 414–429). Routledge.
Sullivan-Bissett, E. (2017a). Biological function and epistemic normativity'. *Philosophical Explorations* 20, 1, pp. 94–110.
Sullivan-Bissett, E. (2017b). Malfunction defended. *Synthese*. *194*(7), 2501–2522.
Sullivan-Bissett, E. (forthcominga). Evolutionary pressures on belief capacities. V. N. Leeuwen, & T. Lombrozo (Eds.), *The Oxford handbook of the cognitive science of belief*. Oxford University Press.
Sullivan-Bissett, E. (forthcomingb) In Defence of Ontic austerity for belief. In J. Jong and E. Schwitzgebel (Eds.) *The Nature of Belief*. Oxford University Press.
Tranel, D., Damasio, H., & Damasio, A. R. (1995). Double dissociation between overt and covert face recognition. *Journal of Cognitive Neuroscience*, *7*(4), 425–432.
Trivers, R. (2000). The elements of a scientific theory of self-deception. *Annals of the New York Academy of Sciences*, *907*, 114–131.
Trivers, R. (2011). *Deceit and self-deception: Fooling yourself the better to fool others*. The Penguin Group.
Trivers, R. (2013). *The folly of fools. The logic of deceit and self-deception in human life*. Basic Books.
Van Leeuwen, N. (2007). The product of self-deception. *Erkenntnis*, *67*, 419–437.
Van Leeuwen, N. (2008). Finite rational self-deceivers. *Philosophical Studies*, *139*(2), 191–208.
Van Leeuwen, N. (2017). Do religious beliefs respond to evidence? *Philosophical Explorations*, *20*, 52–72.
van Prooijen, J. W., & van Vugt, M. (2018). Conspiracy theories: Evolved functions and psychological mechanisms. *Association for Psychological Science*, *13*(6), 770–788.
Warman, D. M., & Martin, J. M. (2004). Cognitive insight and delusion proneness: An Investigation using the beck cognitive insight scale. *Schizophrenia Research*, *84*, 297–304.
Williams, D. (2021). Socially adaptive belief. *Mind & Language*, *36*, 333–354.
Williams, D. (2023). Bad beliefs: Why they happen to highly intelligent, vigilant, devious, self-deceiving, Coalitional apes. *Philosophical Psychology*, *36*(4), 819–833.