Penultimate draft. Please cite final draft forthcoming in *Neuroethics*.

# DELUSION, PROPER FUNCTION, AND JUSTIFICATION

## INTRODUCTION

Among psychiatric conditions, delusions have received significant attention in the philosophical literature. This is partly due to the fact that many delusions are bizarre, and their contents interesting in and of themselves. But the disproportionate attention is also due to the notion that by studying what happens when perception, cognition, and belief go wrong, we can better understand what happens when these go right. In this paper, I attend to delusions for the second reason—by evaluating the epistemology of delusions, we can better understand the epistemology of ordinary belief. Given recent advancements in our understanding of how delusions are formed, the epistemology of delusions motivates a proper functionalist account of the justification of belief. Proper functionalist accounts of the justification of belief hold that whether a belief is justified is partly determined by whether the system that produces the belief is functioning properly.

There are several candidate accounts of how delusions are formed. I argue that they all motivate proper functionalism. Further, any account of delusion formation according to which delusions are epistemically inappropriate beliefs that result from dysfunctional neurobiology or cognition motivates proper functionalism. The structure of this argument is roughly that proper function figures into the epistemic properties of delusional beliefs; since what we say about the epistemology of delusional beliefs also applies to the epistemology of ordinary, non-pathological beliefs, proper function figures into the epistemology of ordinary, non-pathological beliefs. Most of the paper is spent establishing the conclusion that proper function figures into the epistemology of delusional beliefs. Considering the epistemology of delusions not only

motivates proper functionalism, it also shapes the theory. Contemporary proper functionalism includes a condition on the aim of belief, namely that it aims at truth. This is a common take on the aim of belief. Accounting for the pathology of delusion suggests that belief aims at something else. I argue that the below accounts of delusion formation suggest that belief aims not at truth, as proper functionalists and most everyone else thinks, but instead at explanatory adequacy or stability.

This is not to say, however, that when beliefs miss their aim they are necessarily pathological. People often hold false or explanatorily inadequate or unstable beliefs and do so non-pathologically. Further, beliefs are penetrated by other propositional attitudes such as hopes and desires (resulting in beliefs formed by wishful thinking, for instance), and this penetration sometimes causes beliefs to miss their aim. But even when this penetration causes beliefs to miss their aim, the beliefs are not necessarily pathological, or even statistically or functionally abnormal. Indeed, if beliefs were delusions just in case they failed to hit their aim, whatever that aim happens to be, delusion would perhaps be the most widespread pathological condition known to humans. Though it is possible to make inferences from the accounts of delusion formation about what the aim of belief is, these accounts in no way depend on what it happens to be.

In the next section I briefly describe the different candidate accounts of delusion formation. In the section that follows I show that according to these accounts epistemically inappropriate beliefs result from dysfunctional neurobiology or cognition. I then argue that repairing this dysfunction resolves the inappropriate epistemology, which means that at a minimum a thing doing what it's supposed to do—a thing performing its proper function—influences the

epistemic properties of a belief. Once established, I expand on the nature of this influence and finish with a tentative account of proper functionalism.

## DELUSION FORMATION

Along with the candidate accounts of delusion formation, I assume delusions to be doxastic, that someone with Capgras delusion *believes* that their spouse has been replaced by an imposter [1–3]. There is some debate about whether this is true, with some authors suggesting that delusions are not beliefs because they don't work in ways that beliefs work, such as by being consistent with other beliefs or sensitive to evidence [2]. Although I write as though delusions are beliefs, nothing that follows hinges on whether delusions are in fact beliefs, so long as whatever they are if not beliefs has a mind-to-world direction of fit, that the content of the state changes according to observations of world. Beliefs have a mind-to-world direction of fit, but so do other attitudes and experiences.

Considered as such, delusions are pathological beliefs. What distinguishes the different candidate accounts from each other is what they say about this pathology.

One-factor accounts of delusion formation hold that only one pathology is needed to account for the formation of delusion. Maher claims that delusions are reasonable responses to abnormal experiences [4]. The malfunction is in the production of the abnormal experience, but given the abnormal experience the delusion is a reasonable response. For example, Capgras delusion—the delusion that one's loved one has been replaced by an imposter—is an abnormal experience that results from an experience of a familiar face without the concomitant affective state, an

experience that may be due to brain damage. But, given the experience that one has, the delusion itself is a reasonable response.

Two-factor accounts locate an additional malfunction. This is motivated in part by the fact that some people with the indicated brain damage and the abnormal experience fail to adopt delusion. Thus, another factor must be contributing to the delusion's adoption or maintenance or both. Coltheart, Menzies, and Sutton adopt a Bayesian approach to delusion formation and argue that while a neurobiological malfunction results in an abnormal experience, which is sufficient to result in adopting the delusion, a second malfunction of hypothesis evaluation is responsible for its maintenance [5]. Delusions are stubborn and resist countervailing evidence. Why doesn't a person with delusion, once adopted, abandon the delusional belief in favor of one that incorporates this other evidence? Coltheart et al. claim that it's a bias toward the conservation of adopted beliefs. Given the weight of adopted beliefs, other hypotheses can't compete in the explanation of the abnormal experience. Consider again Capgras delusion. Coltheart et al., like others, claim that people with Capgras have an abnormal experience. When looking at a familiar person, a properly functioning process of facial recognition is one in which the capacity for visual recognition of faces connects to the autonomic nervous system. Normally, looking at a familiar person triggers the recognitional capacity, which then predicts an autonomic response such as affection. But in people with Capgras, the facial recognitional capacity is cleaved from the autonomic nervous system, perhaps from Alzheimer's or stroke. Thus, when one looks at a familiar face, the recognitional capacity is triggered, which then predicts a response from the autonomic response. But because they are disconnected from each other, there is no such response. Thus, the person's response to the abnormal experience is the delusional belief that

one's loved one has been replaced by an imposter (factor one). Once adopted, a second malfunction in the belief evaluation system allows the delusional belief to persist (factor two).

Ryan McKay also argues for a two-factor account of delusion formation [6]. And like Maher and Coltheart et al., he agrees that the first factor in delusion formation is an abnormal experience that results from neurobiological malfunction (e.g., the disconnected facial recognitional capacity and autonomic nervous system). But he locates the second factor not in a dysfunctional belief evaluation system (which permits persistence). Rather, he holds that the second factor is in the adoption of the delusional belief. Where Coltheart et al. hold that the second factor is the bias toward doxastic conservatism, McKay claims that the second factor is a bias toward explanatory adequacy. For the person with Capgras delusion, the delusional imposter belief is *adopted* only because the believer's Bayesian updating is biased toward the adequacy of the explanation of the abnormal experience. The explanation that the person is an imposter is weighted too heavily, and this swamps the adoption of alternative explanations.

The one- and two-factor accounts above are bottom-up accounts of delusion formation: sensory input results in data that prompt a cognitive and then doxastic response. But perceptual belief formation generally and delusion formation specifically may also be top-down. Since everyone agrees that perception influences belief, this amounts to the view that beliefs (or other cognitive states) can influence perception. This is a weak claim that is certainly true in a trivial sense. My belief that my coffee mug is behind me influences the movement of my head and eyes, thereby influencing my perception of my coffee mug. The influence at issue is rather something stronger: the content of a perceptual experience is partially influenced by the person's beliefs, where if one were to attend to the same object of perception but have different beliefs, the experience would have been different [7]. If we view the perception-belief process this way,

then there is an alternative account of delusion formation available: The dysfunction responsible for delusion starts with cognitive dysfunction, resulting in an inappropriate belief. This belief then exerts top-down influence upon the experience, which then reinforces the original belief.

Recently, an account of delusion formation that permits this sort of top-down influence has become prominent. The account views perception and belief as different levels in a hierarchy of inferences that operates according to Bayes' Theorem. For any given experience and belief, the mind brings to bear a set of priors. These priors are expectations of what is likely to be perceived, a prediction that is passed down through the hierarchy. These predictions exert top-down influence. At the same time, perception, under the influence of these priors, detects features of the environment. When this feature detection matches the prediction, the priors are updated accordingly, increasing the probability of those predictions. But if the bottom-up information from the environment doesn't match the top-down predictions, an error signal is produced. The prediction was wrong. This error signal then gets passed up the hierarchy until it is reconciled with higher level predictions. These predictions are then updated accordingly, so that when they are brought to bear on a future experience, they will have been altered slightly by the error signal. Thus, the new prior will be a different prediction exerting top-down influence on the experience, which will then match or not and move up the hierarchy, updating priors as it does so. It's a powerful, economical—only error signals are transmitted up the hierarchy—and coherent account of the perception and belief.

This view is supposed to account for the formation of delusions. Andy Clark explains [8]:

> The key idea…is that understanding the positive symptoms of schizophrenia [such as delusion] requires understanding disturbances in the generation and weighting of prediction error. The suggestion is that malfunctions within that complex economy

> (perhaps fundamentally rooted in abnormal dopaminergic functioning) yield wave upon wave of persistent and highly weighted "false errors" that then propagate all the way up the hierarchy forcing, in severe cases (via the ensuing waves of neural plasticity) extremely deep revisions in our model of the world. The improbable (telepathy, conspiracy, persecution, etc.) then becomes the least surprising, and—because perception itself is conditioned by the top-down flow of prior expectations—the cascade of misinformation reaches back down, allowing false perceptions and bizarre beliefs to solidify into a coherent and mutually supportive cycle…Such a process is self-entrenching. As new generative models take hold, their influence flows back down so that incoming data is sculpted by the new (but badly misinformed) priors so as to "conform to expectancies." (p. 196)

Delusions are formed when, due to perhaps a malfunction in the production and transmission of dopamine, there are more frequent or more heavily weighted errors, mismatches between the prediction and incoming perception, that get passed up the hierarchy, updating priors, which then come back down and influence perception, which reinforces the erroneous prediction, which strengthens the prior so that the erroneous prediction exerts even more top-down influence. Thus, what starts as a mismatch between incoming perceptual information and top-down prediction results in a bizarre and false belief, which only gets reinforced as that belief influences subsequent perceptions. Consider, for example, how this accounts for Capgras delusion, the delusional belief that a loved one has been replaced by imposters. Corlett et al. explain [9]:

> Capgras results when patients experience an anomalous lack of affective responding when confronted with their relatives the delusion constitutes a new prior driven by the experience, a means for explaining it away. It is possible that the initial affective

> disturbance results from a failure to guide affect perception by prior experience, that is, just like sensory perception, emotions are predicted; we have emotional priors, indeed, it is the prior expectancy of a familiar face combined with an emotional response (learned through experience) which breaks down in Capgras patients; fostering the misidentification of someone (or something) familiar as unfamiliar. (p. 358)

Corlett et al. explain the formation of Capgras delusion via predictive coding as a mismatch between the prediction of emotion accompanying the recognition of a loved one and the incoming perceptual information, which lacks an affective component. This surprising predictive error gets passed up the hierarchy until the experience (no affective component) can be reconciled with the predictive error (prediction of affection): the person must be an imposter. This explanation becomes the prior that bears upon subsequent experiences of the person, which only goes to reinforce the delusion.

## DYSFUNCTION AND EPISTEMIC PROPERTIES

For one- and two-factor accounts and the predictive coding account, dysfunction is responsible for the delusion. For all three accounts, the process leading to the experience is dysfunctional. For two-factor accounts, there is additional dysfunction after the experience, leading to either the adoption or maintenance of the delusion or both. Mechanisms don't do what they are supposed to do, and because of that a person develops a delusional belief. If the mechanisms are dysfunctional—if they are not performing their function—then there must be something that they are supposed to do. The mechanisms responsible for delusion formation must have a proper function.

I say more about proper functions below, but for now it is sufficient to recognize that delusion results from mechanisms failing to do what they are supposed to do. Identifying these mechanisms is the point of the accounts of delusion formation. If delusion formation is not due to the failure of mechanisms to satisfy their proper function—if there's nothing that these mechanisms are supposed to do—then it can't be that pathology is a matter of dysfunction. To put it another way, if the pathology of delusion is *not* at least partly a matter of a mechanism not doing what it's supposed to do, then the pathology of delusion is independent of dysfunction, which implies that pathology more generally is independent of dysfunction.

That pathology is not at all a matter of dysfunction is certainly a view that some people hold [10]. But it is much more common to hold that pathology is at least partly a matter of proper function (even if what functions count as proper is determined by human interests rather than some mind-independent feature of the mechanism). And in any case, the accounts of delusion formation are clearly accommodating of such objectivist accounts of pathology. For my argument that follows, it is unnecessary to further specify an account of pathology; so long as the true account of pathology is one in which proper function plays a role, it doesn't matter what the specific relation is between proper function and pathology.

Delusions like Capgras result from some biological or cognitive mechanism failing to do what it is supposed to do. But delusions are also epistemically inappropriate. This is to say that delusional beliefs fail to meet the standards for the epistemic permissibility of belief. Often delusional beliefs are described as being irrational. And whether a belief is rational is at least partly a matter of whether it coheres and is consistent with other propositional attitudes. But something more specific can be said about the epistemic status of delusions: they are unjustified beliefs. Rationality, at least in discussions about the epistemic inappropriateness of delusions, is

a property of subject: subjects are irrational for adopting or maintaining a delusional belief. The subject ought not hold that belief. But the delusional belief itself is also epistemically inappropriate. The belief ought not be held. Why the belief ought not be held will differ according to the different accounts of doxastic justification. But delusional beliefs are not justified.[1]

The rejection of this claim is that delusional beliefs can be justified. This amounts to the view that a delusional belief can be justified despite it being irrational for the person to believe it. If a delusion is justified, then the person holding it has justification to believe the content of the delusion (e.g., that one's wife has been replaced by an imposter).[2] Thus, if delusions can be justified, then relative to the proposition believed, the person has justification to believe the proposition but is also irrational for doing so. In the absence of a deflated notion of rationality, it seems absurd to assert such a conjunction. And if it is absurd, then either the claim that delusional beliefs can be justified must be abandoned or the claim that the subject of the belief holds it irrationally must be.

Consider an example in which holding a particular belief is irrational for a person, because the belief is probabilistically or logically inconsistent with other beliefs or is otherwise incoherent. Can someone at the same time have justification to hold that belief? It seems absurd to say they can: inconsistency or incoherence defeats justification. Thus, if a person irrationally holds a particular belief in spite of its incoherence, the incoherence itself will also defeat the

[1] This is not to say that delusional beliefs are lacking entirely epistemic value. Recently, some authors have argued that monothematic delusions are epistemically innocent, which is an epistemic property of states that are epistemically faulty but nevertheless confer some epistemic benefit [3, 22, 26].

[2] John Turri argues that doxastic justification explains propositional justification, which is contrary to the orthodox view that the order of explanation goes the other way [27]. Even so, his view wouldn't permit an irrationally held justified belief.

justification the person may have for the belief. Or, in other words, whatever defeats rationality of a belief also defeats the justification a person has to hold it. And if a person lacks justification to hold the belief, the belief itself cannot be justified.

The above is true regardless of one's account of justification: even externalist accounts of justification, such as reliabilism, have a "No Defeater" condition [11]. Suppose someone has just finished a meeting that began at 1:30 pm and looks at their watch. Unbeknownst to the person, the watch stopped working at 1:00 pm, so that when they look at their watch, they form the belief that it is 1:00 pm.[3] So long as the only input into the belief is the sensory evidence, the belief that it is 1:00 pm is justified (given the reliability of the watch) and the person is rational to hold it. But once other inputs go into the belief, such as the defeating evidence that one just finished a meeting that started at 1:30 pm, the justification that one has to hold the belief that it is 1:00 pm is defeated and along with it the rationality to hold it. So, even according to reliabilism, once all the inputs are considered, it's not the case that the person is justified in holding the belief that it is 1:00 pm but irrational in doing so.

Maher rejects the claim that the person holding the delusional belief is irrational, as the delusion on his one-factor account is a reasonable response to an abnormal experience. But his one-factor account of delusions is compatible with the notion that beliefs that result from the abnormal experience are nevertheless unjustified. Susanna Siegel has recently argued that experiences themselves are epistemically evaluable [12]. The orthodox view is that experiences are epistemically neutral. But Siegel holds that they can be "epistemically charged" prior to any cognitive response to them. If she is right that experiences can be epistemically charged, and this charge can be transmitted to the subsequent belief, then even one-factor accounts like Maher's

[3] Thanks to an anonymous referee for this example.

can hold that delusional beliefs are unjustified. If an abnormal experience is negatively charged and this results in a delusional belief, then the belief may be unjustified.

By committing to the view that delusional beliefs are held irrationally, two-factor and predictive coding accounts of delusion seem committed to the view that delusional beliefs are unjustified. One-factor accounts are merely compatible with that view. From here on, I treat delusional beliefs as unjustified beliefs. As such, delusional beliefs are, among other things, unjustified beliefs that result from a biological or cognitive mechanisms failing to do what they are supposed to do.

## RESTORING FUNCTION

Suppose a person suffers from Capgras delusion. For one- and two-factor accounts, the delusion is prompted by an abnormal experience that results from the disconnection of the facial recognitional capacity from the autonomic nervous system. Seeing his wife triggers the recognitional capacity and predicts a concomitant affective response. But because the two systems are disconnected, there is none, producing the abnormal experience. For two-factor accounts an additional cognitive pathology results in the adoption (e.g., McKay's bias toward explanatory adequacy) of the (unjustified) delusional belief or the maintenance (Coltheart et al. bias toward doxastic conservatism) of it. The end product is the unjustified delusional belief that an imposter has replaced his wife.

On predictive coding accounts, the delusion is generated and maintained by neurobiological dysfunction that results in his perceptual system being more prone to predictive error, requiring more top-down influence on the reconciling of these errors. The top-down

predictions expect that low-level perceptual recognition of his wife is accompanied by an affective component. Due to the neurobiological dysfunction, there is no such affective component. This unexpected prediction error is surprising, and gets passed up the Bayesian hierarchy until it is reconciled with other priors. The explanation that reconciles the mismatch between the priors and the posteriors is that his wife has been replaced with an imposter. This results in an unjustified delusional belief that an imposter has replaced his wife.

For all of the accounts, the pathology is partly neurobiological, and the delusion is a symptom of this pathology. But a secondary symptom of this pathology is the epistemic condition—the man's delusional belief that his wife has been replaced by an imposter is not justified.

This case illuminates the nature of justification, because resolving the epistemic condition will indicate which properties or processes factor into a belief being justified. Returning the man's neurobiological dysfunction to its proper function has the downstream effect of resolving this epistemic condition. On one-factor accounts resolving the secondary symptom (lack of justification) is straightforward. Restore the proper function of the system connecting the facial recognitional capacity to the autonomic nervous system. Then when the man sees his wife the recognitional capacity is triggered, which predicts the concomitant affective response from the autonomic nervous system. Because that system is doing what it is supposed to do, there is such a response. It's a perfectly ordinary experience the cognitive response to which is to adopt the perfectly ordinary—and justified—belief that his wife just walked into the room. Restoring the system to its proper function resolves the primary symptom, the delusion, and in so doing resolves the secondary symptom, the lack of justification.

Even if Maher is right that people rationally hold delusional beliefs, proper function can still influence justification. So long as restoring the underlying pathology to its proper function increases the degree of justification, then proper function has a role to play in the story of how beliefs come to be justified. And it seems intuitive that the pathological belief that one's wife has been replaced by an impostor, even if rationally held, is *less* justified than the non-pathological belief that the man's wife has just walked into the room. If so, then restoring function influences justification, even if the delusional belief is a rationally held justified belief, and repairing function merely increases the belief's degree of justification.

The same is true of two-factor accounts: restore the neurobiological pathology to its proper function, and no abnormal experience is generated. With no abnormal experience, the second factor doesn't come into play and the person goes on to have an ordinary experience of his wife, which then results in an ordinary, justified belief about his wife. But for two-factor accounts, restoring whatever the cognitive mechanism is responsible to its proper function will also resolve the secondary symptom. Suppose a person has the first factor of Capgras delusion. Restoring the cognitive pathology to its proper function involves, if McKay is right, appropriately weighting the disposition to explanatory adequacy and, if Coltheart et al. are right, appropriately weighting the disposition to conserve already adopted beliefs. But either way the person's response to the abnormal experience will not result in an unjustified belief that an imposter has replaced his wife. On McKay's view, if the second factor is functioning properly, the delusional belief won't be adopted. On Coltheart et al's view, the delusional belief won't be maintained. In either case, the epistemic condition will be resolved; he won't adopt and maintain the delusion. In such a person (dysfunctional first factor, properly functioning second factor) it is not clear what belief the person will adopt. It's possible that the person will withhold belief

altogether. But at least the secondary symptom will be resolved. So, restoring either factor to its proper function will resolve the epistemic condition.

It's a similar story for predictive coding accounts. Suppose the particular neurobiological dysfunction is dopaminergic. Also suppose that we are able to restore the transmission of dopamine to its proper function. If it is correct that the formation and persistence of the man's delusion is due to unstable low-level perception and prediction error, which is in turn due to the dysfunctional transmission of dopamine, then returning the transmission to its proper function should result in different low-level perceptual information. That is, we should expect that returning the man's dopamine transmission to its proper function results in more stable low-level perception, perceptual information that is more likely to match the predictions from higher levels in the hierarchy and less likely to generate predictive errors. With no surprising prediction error passing up the hierarchy, there is no need for the man's higher-level priors to be updated—there is no surprising error in need of explanation. Since the delusional belief that the man's wife has been replaced by an imposter results from this need for explanation, the delusion would not be generated. What would be generated is presumably stable low-level perceptual information that matches the predictions coming from higher levels and the subsequent ordinary belief that his wife just walked into the room. There is no apparent epistemic failure in this perfectly ordinary process. Returning the neurobiological dysfunction to its proper function resolves one symptom, the delusional belief, and in so doing resolves the secondary symptom, the lack of justification.

The fact that restoring the proper function of the mechanisms responsible for delusion formation resolves the secondary symptom, the unjustified belief, indicates, that at a minimum, proper function has a role to play in determining the epistemic properties of delusional beliefs, and, presuming that delusional beliefs are the same sort of thing as ordinary beliefs, the

epistemic properties of ordinary beliefs. More specifically, proper function influences justification. Some might think that a better example to demonstrate the influence of proper function on justification would be to hold fixed the belief that one's wife has been replaced by an imposter and change the functional status of the process leading to the belief. If upon inspection we find a corresponding change in justification, then that's better evidence that proper function influences justification. But the challenge for the proper functionalist is that by changing the proper function, the content of belief changes as well—the content is tied to the functional status.

Consider a person with schizophrenia who has the sensory experience of parasites crawling under their skin and the delusional belief that parasites are crawling under their skin. The belief is unjustified, and it results from an underlying pathology, possibly related to the transportation of dopamine [13]. Now consider someone who has the belief that parasites are crawling under their skin (so the belief content is the same as the person with delusional parasitosis) which is based on the sensory experience of parasites crawling under their skin (so the phenomenal character is the same as the person with delusional parasitosis). But suppose that this person is infected with *Dirofilaria repens*, a parasite infected mosquitos transmit to humans, which results in parasites that migrate under the human's skin, commonly the face [14]. Although the contents of the experience and the belief are the same between the two people, the belief of the person with delusional parasitosis seems unjustified, while the belief of the person infected with *D. repens* seems justified. My claim is that what accounts for this difference in justification is that the belief-forming faculties of the person with delusional parasitosis are not functioning properly, but the faculties of the person infected with *D. repens* are functioning as they should.

It is important to view the epistemic condition of those with delusional beliefs as a symptom of an underlying pathology. By analogy, a punctured lung can be a factor in the lung collapsing, which causes shortness of breath. To demonstrate that a punctured lung is a factor in the shortness of breath, we wouldn't find people with a collapsed lung but no puncture (e.g., from COPD), look to see whether they have shortness of breath, and, finding that they do, conclude that the puncture is therefore not a factor in the shortness of breath. Rather, we would expect that to resolve the shortness of breath, the underlying pathology (i.e., the puncture) should be resolved.

Similarly, to resolve the epistemic condition of those with delusions, the underlying pathological condition should be restored to its proper function. In doing so, we should expect that the belief that one's wife has been replaced with an imposter vanishes, improving the person's epistemic condition in the way that a repaired lung is better able to breathe. The same could be said for other delusions: restore whatever neurobiological or cognitive faculty is responsible to its proper function and the person's epistemic condition improves along with it.

In cases of delusion, returning a process to its proper function results in a change in the justification of the resulting belief. The influence of a properly functioning neurobiological or cognitive faculties upon justification doesn't show that proper function is either necessary or sufficient for justification. But it does show that it has a role to play in the story of how beliefs come to be justified.

## CONTEMPORARY PROPER FUNCTIONALISM

Drawing on Alvin Plantinga's account of warrant [15], Michael Bergmann has proposed the most sophisticated proper functionalist account of justification [16]. His account is that a subject's belief B, is justified if, and only if, (i) the subject does not take B to be defeated and (ii) the cognitive faculties producing B are (a) functioning properly, (b) truth-aimed, and (c) reliable in the environments for which they were "designed." The conditions are severally necessary and jointly sufficient for a belief to be justified.

To say that a subject takes a belief to be defeated is to say that she has a further belief that either the belief is false or that the grounds upon which the belief is based are unreliable; she has further evidence to doubt the belief in question. It is a mark of delusional beliefs that the subjects do not take them to be defeated. They are rather resistant to countervailing evidence. So delusional beliefs satisfy this condition.

The second condition imports proper function. There are several accounts of proper function that a proper functionalist might adopt. The most prominent type of account holds that a mechanism's (or a system or a trait type) proper function is the thing it does that has contributed to organism's ancestors' survival.[4] It is a historical notion according to which proper function is tied to natural selection. For example, Karen Neander's account [17]is that:

> It is the/a proper function of an item (X) of an organism (O) to do that which items of X's type did to contribute to the inclusive fitness of O's ancestors, and which caused the genotype, of which X is the phenotypic expression, to be selected by natural selection.

If Neander is correct, then to determine the proper function of a particular neurobiological mechanism, we should look to see what that mechanism did to contribute to a person's ancestors'

[4] I adopt 'proper function' rather than perhaps a more common terminology, 'etiological function,' out of convenience, as the epistemological theory (i.e., proper functionalism) adopts this language.

fitness. In the case of Capgras, the proper function of the facial recognitional capacity arguably includes passing information to the autonomic nervous system, perhaps as a way to prompt behaviors appropriate to the identity of the person recognized (or not recognized), such as behaviors involved in fight or flight.

Not all accounts are historical, in the way that Neander's and other prominent accounts are [18, 19]. Cummins' account ties proper function to a thing's causal role within a given system rather than to any goal or aim or history [20]. More recently, Nanay, like Neander and Millikan, has argued that proper function is tied to inclusive fitness, but his is an a-historical account. Instead, he proposes a modal account of proper function [21]:

> Performing *F* is a function of organism *O*'s trait *x* at time *t* if and only if some 'relatively close' possible worlds where *x* is doing *F* at *t* and this contributes to *O*'s inclusive fitness are closer to the actual world than any of those possible worlds where *x* is doing *F* at *t* but this does not contribute to *O*'s inclusive fitness.

For Nanay, what counts as a "relatively close" possible world is determined by the explanatory project of interest. Once specified, if in those relatively close possible worlds *x* is doing *F* and doing so is contributing to the organism's inclusive fitness, and there are no closer worlds in which *x* is doing *F* but it doesn't contribute to the organism's inclusive fitness, then the proper function of *x* is to *F*. Further, unlike Neander (or Millikan's) account of proper function, Nanay's account attributes proper function to tokens, not types. Thus, two tokens of a trait type might have different proper functions.

In the case of Capgras, Nanay's account would likely arrive at the same proper function of the neurobiological deficit responsible for delusion, whatever that happens to be: it is the proper function of the person's facial recognitional capacity to pass information to the autonomic

nervous system, because the possible worlds in which it does this and contributes to inclusive fitness are closer than all of the possible worlds in which it does this but doesn't contribute to inclusive fitness. However, because it is not a historical account and it attributes proper function to tokens rather than types, Nanay's account may have interesting implications for how we think of delusion. For example, it's plausible that some tokens of motivated delusions may contribute to inclusive fitness [3, 22]. If proper function is a-historical and attributed only to trait tokens, then on Nanay's account it may turn out that such delusions result from a properly functioning trait token (depending on whether it contributes to inclusive fitness in other, relatively close possible worlds). This would imply either that delusions are not pathological or that pathology is not a matter of whether a thing is malfunctioning.

Regardless of what the best account of proper function is, to establish my claim that proper function figures into the justification of delusional beliefs, and for this reason the justification of ordinary beliefs, it is unnecessary for me to commit to a particular notion of proper function. This is Bergmann's strategy. So, when I claim that proper function figures into whether a belief is justified, it doesn't matter whether the proper function of a mechanism is determined by what's happening in possible but not actual worlds or whether it has contributed to the survival of an organism's ancestor or whether it has been thoughtfully designed by an omnipotent creator. Of course, if a mechanism having a proper function implied intelligent design, then that would be a decisive reason to reject that notion of proper function. But my point is simply that it is open to me to adopt whatever it is that ends up being the best account of proper function.

The motivation for the condition that the faculties producing the belief be truth-aimed stems from several considerations. One is simple intuition. It seems like our beliefs are trying to

get it right—they're trying to represent the world accurately. When one believes that the meeting will go late, the belief that it will and the hope that it won't feel different, and that difference is that one attitude is trying to get it right and other isn't. A second consideration is that belief aiming at truth has significant explanatory power. In particular, belief aiming at truth can explain why beliefs seem to be stubbornly resistant to regulation from other types of states. You can't believe that Hillary Clinton won the 2016 United States' presidential election even though you may want to very much. Similarly, it doesn't seem as though our beliefs can be influenced by practical reasons, such as how much we are being paid to believe that *p*. In other words, it seems like nothing other truth-directed states, not even our will, can regulate belief. The explanation for this phenomenon is that belief aims at truth. However, considering delusions and their etiology suggests that the aim of belief is not truth, which, if correct, would imply that Bergmann's account of proper functionalism would need to be amended. In the section after next I discuss possible amendments.

Bergmann's final condition is that the cognitive faculty be reliable in the environment for which it is "designed." All this means is that the faculty tend to produce true beliefs in the environment in which the function of the faculty evolved or was designed. Suppose that in our environment our belief-forming faculties evolved to aim at truth. If so, then as long as the cognitive faculties responsible for belief-formation tend to hit that aim in our environment, then the condition is satisfied. But put us in a different environment, one in which the conditions cause us to constantly suffer from perceptual illusions, and our cognitive faculties wouldn't be reliable. But that wouldn't fail the condition, because that's not the environment for which they were designed (or in which they evolved). In our environment, our cognitive faculties are reliable, as they tend to produce true beliefs.

With this account of proper functionalism in hand, notice that according to proper functionalism, delusional beliefs are not justified. For subjects with delusional beliefs, the first condition is satisfied—they usually don't take them to be defeated. But the second condition is not satisfied. Although delusional beliefs are produced by cognitive faculties that are reliable in the environments in which they evolved (e.g., perception and reasoning in such people are still generally reliable for people who have delusional beliefs), and beliefs may be aimed at truth, the faculties producing the delusional belief are not functioning properly (either at the neurobiological level or the cognitive level). Thus, on this proper functionalist account, delusional beliefs are not justified.

## DELUSIONAL BELIEF TO ORDINARY BELIEF

I have argued that the proper function of the mechanisms responsible for delusion formation figure into the justification of delusional belief. Proper function figures into justification regardless of whether one adopts a one-factor, two-factor, or predictive coding account of delusion formation. It is a further step to claim that because proper function figures into justification in the case of delusion it also figures into justification in the case of ordinary belief. But it is not a big step, especially if one already accepts the candidate accounts of delusion formation.

The process leading to delusion is either a deviant version of the process that leads to ordinary belief, or it is a different process altogether. If the process leading to delusion is a different sort of process from the one that results in ordinary belief, then the accounts of delusion formation begin to look a lot less impressive. In such a case, they would be accounts of a process

that leads to a psychiatric condition, but one that tells us nothing about how beliefs are formed in the ordinary way. Furthermore, if the process leading to delusion were a different type of process, it's not clear that delusions could be considered abnormal. A process can't be abnormal, if there is no normal. If delusions aren't abnormal relative to normal belief-forming processes, then they aren't abnormal at all.

If it makes sense to think of delusions as the ordinary belief-forming process gone wrong, then we have good reason to think that what goes wrong epistemically reveals the epistemology of ordinary beliefs. That is, if the psychological properties of delusion reveal the psychology of ordinary belief, then we can say the same thing about the epistemic properties of delusion and ordinary beliefs. This is especially true since most theories of how beliefs become justified explain that justification in terms of the process leading to the adoption of the belief.[5] Since in the case of delusions proper function figures into the justification of the delusional belief, it is likely that proper function also figures into the justification of ordinary belief.

## REVISED PROPER FUNCTIONALISM

Bergmann's version of proper functionalism is that a subject's belief B, is justified if, and only if, (i) the subject does not take B to be defeated and (ii) the cognitive faculties producing B are (a) functioning properly, (b) truth-aimed, and (c) reliable in the environments for which they were "designed." This version of proper functionalism accounts for the epistemic failure of

[5] This is obviously true for accounts such as reliabilism or proper functionalism. But even internalist accounts propose some grounding condition on the belief.

delusional belief, if a one-factor, two-factor, or predictive coding account of delusion formation is correct. But if these accounts are correct, we need a revised proper functionalism.

First, the conditions (ii)(a)-(ii)(c) of Bergmann's proper functionalism only apply to the subject's cognitive faculties. But justification is not simply a matter of persons' cognitive faculties, as restoring proper function to presumably non-cognitive, neurobiological mechanisms also influences justification. So, instead of the conditions being for cognitive faculties, they should be for all belief-forming faculties.

Second, and more importantly, considering the accounts of delusion formation may warrant reconsidering whether the aim of belief is truth. If a predictive coding account is right and beliefs are the result of a Bayesian hierarchy of predictions exerting top-down influence of incoming posteriors, which then update those priors, then it isn't as obvious as the beliefs aim at truth [23]. And if they don't aim at truth, then the reliability of the mechanism that produces them is less of a concern.

Instead, if beliefs are the result of predictive coding, then it seems more accurate to say that beliefs aim at stability of the hierarchy, if they aim at anything at all. Or, to put it in terms common to predictive coding accounts of the mind, the aim of belief is to minimize "free energy" or prediction error [24, 25]. The immediate aim of belief is not to represent the world accurately, but to represent whatever the hierarchy serves up. And the Bayesian hierarchy aims at stability. This is not to say that the aim of belief cannot also be truth. It is plausible that in creatures like us a stable hierarchy is one that produces truths—a properly functioning stable predictive coding mechanism tends to result in accurate representations, because doing so contributes to fitness, for example. Achieving the primary aim (stability) results in achieving the secondary aim (truth).

Considering two-factor accounts of delusion formation also warrant reconsidering the aim of belief. Coltheart et al. claim that the second factor is a bias toward doxastic conservatism. McKay's account attributes the second factor to the bias toward explanatory adequacy (noting, however, that his account is compatible with predictive coding). In either case, the fit between the abnormal experience and the priors is off, due to the bias toward either doxastic conservatism or the bias toward explanatory adequacy. It may then be plausible to think of beliefs not aiming at truth but more immediately at fitness of the belief to both the priors and incoming abnormal sensory information (which, again, may tend to produce true beliefs). Fitness could then be fleshed out in Bayesian terms. Such an account may end up being the same as, or very similar to, the idea that belief aims at stability.

How conditions (ii)(b) and (ii)(c) are specified will depend on what the proper function of belief-formation is. If the proper function of belief-formation is to represent the world accurately, then conditions tying justification to truth make sense. But the proper function of belief-formation may not be to represent the world accurately. Maintaining explanatory adequacy between priors and incoming information or minimizing prediction error may contribute to our inclusive fitness more so than representing the world accurately. And if the proper function of belief-formation is maintaining explanatory adequacy or minimizing prediction error, then a proper functionalist account of justification must heed these attributions of function.

A proper functionalist account of justification could therefore be amended. Instead of conditioning justification on the cognitive faculty being truth-aimed and reliable, it should be conditioned on the belief-forming faculty being stability-aimed and stable in the environment for which it was designed or conditioned on being fitness-aimed and fit in the environment for which it was designed. Thus, a proposal for a proper functionalist theory of justification is that a

subject's belief B, is justified if, and only if, (i) the subject does not take B to be defeated and (ii) the belief-forming faculties producing B are (a) functioning properly, (b) stability-aimed (or fitness-aimed), and (c) stable (or fit) in the environments for which they were "designed." Like Bergmann's version, this version implies that delusional beliefs are not justified, but it more neatly reflects the functioning of a Bayesian mind, while allowing for our faculties to be aimed mediately at truth. It is a starting point for an updated proper functionalism.

## CONCLUSION

Proper function has a role to play in the justification of beliefs. I have not argued that its role is to the exclusion of other possible factors such as reliability or whether a belief is grounded in one's evidence. A delusional belief's lack of justification is a secondary symptom of the pathology and restoring the belief-forming faculty's proper function resolves this secondary symptom in the same way that repairing a punctured lung results in better breathing. And just as it is right to say that repairing the puncture influences breathing, it is right to say that repairing the pathological belief-forming faculty influences justification. The epistemology of delusions illuminates the epistemology of ordinary belief in the way that understanding disease helps to illuminate health. Proper function may not be the only factor in determining a delusion's epistemic properties, but how the mechanisms responsible for forming beliefs function figure into whether that belief is justified.

## REFERENCES

1. Bayne, Tim, and Elisabeth Pacherie. 2005. In Defence of the Doxastic Conception of Delusions. *Mind & Language* 20. John Wiley & Sons, Ltd (10.1111): 163–188. doi:10.1111/j.0268-1064.2005.00281.x.
2. Bortolotti, L. 2010. *Delusions and Other Irrational Beliefs*. International Perspectives in Philosophy & Psychiatry. OUP Oxford.
3. Bortolotti, Lisa. 2016. Epistemic Benefits of Elaborated and Systematized Delusions in Schizophrenia 67: 879–900. doi:10.1093/bjps/axv024.
4. Maher, B A. 1974. Delusional thinking and perceptual disorder. *Journal of individual psychology* 30. United States: 98–113.
5. Coltheart, Max, Peter Menzies, and John Sutton. 2010. Abductive inference and delusional belief. *Cognitive Neuropsychiatry* 15: 261–287. doi:10.1080/13546800903439120.
6. Mckay, Ryan. 2012. Delusional Inference 27: 330–355.
7. Macpherson, Fiona. 2017. The relationship between cognitive penetration and predictive coding. *Consciousness and Cognition* 47. The Author: 6–16. doi:10.1016/j.concog.2016.04.001.
8. Clark, Andy. 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences* 36: 181–204. doi:10.1017/S0140525X12000477.
9. Corlett, P. R., J. R. Taylor, X. J. Wang, P. C. Fletcher, and J. H. Krystal. 2010. Toward a neurobiology of delusions. *Progress in Neurobiology* 92: 345–369. doi:10.1016/j.pneurobio.2010.06.007.
10. Reznek, Lawrie. 1987. *The Nature of Disease*. Routledge & Kegan Paul.

11. Beddor, Bob. 2015. Process reliabilism's troubles with defeat. *Philosophical Quarterly* 65: 145–159. doi:10.1093/pq/pqu075.

12. Siegel, Susanna. 2017. *The Rationality of Perception*. Oxford University Press.

13. Huber, M., E. Kirchler, M. Karner, and R. Pycha. 2007. Delusional parasitosis and the dopamine transporter. A new insight of etiology? *Medical Hypotheses* 68. Churchill Livingstone: 1351–1358. doi:10.1016/J.MEHY.2006.07.061.

14. Ermakova, Larisa Alexandrovna, Sergey Andreevich Nagorny, Elena Yurievna Krivorotova, Natalia Yurievna Pshenichnaya, and Olga Nikolaevna Matina. 2014. Dirofilaria repens in the Russian Federation: current epidemiology, diagnosis, and treatment from a federal reference center perspective. *International Journal of Infectious Diseases* 23. Elsevier: 47–52. doi:10.1016/J.IJID.2014.02.008.

15. Plantinga, Alvin. 1993. *Warrant and Proper Function*. New York: Oxford University Press. doi:10.1093/0195078640.001.0001.

16. Bergmann, M. 2006. *Justification Without Awareness: A Defense of Epistemic Externalism*. Oxford Scholarship Online. Philosophy Module. Clarendon Press.

17. Neander, Karen. 1991. The teleological notion of 'function.' *Australasian Journal of Philosophy* 69. Routledge: 454–468. doi:10.1080/00048409112344881.

18. Neander, Karen. 1991. Functions as Selected Effects: The Conceptual Analyst's Defense. *Philosophy of Science* 58. [University of Chicago Press, Philosophy of Science Association]: 168–184.

19. Millikan, Ruth Garrett. 1989. In Defense of Proper Functions. *Philosophy of Science* 56. The University of Chicago Press: 288–302. doi:10.1086/289488.

20. Cummins, Robert. 1975. Functional Analysis. *The Journal of Philosophy* 72: 741.

doi:10.2307/2024640.

21. Nanay, Bence. 2010. A Modal Theory of Function 107: 412–431.
22. Bortolotti, Lisa. 2015. The epistemic innocence of motivated delusions q. *Consciousness and Cognition* 33. Elsevier Inc.: 490–499. doi:10.1016/j.concog.2014.10.005.
23. McKay, Ryan T, and Daniel C Dennett. 2009. The evolution of misbelief. *The Behavioral and brain sciences* 32. England: 461–493. doi:10.1017/S0140525X09990975.
24. Friston, Karl, James Kilner, and Lee Harrison. 2006. A free energy principle for the brain 100: 70–87. doi:10.1016/j.jphysparis.2006.10.001.
25. Friston, Karl. 2010. The free-energy principle : a unified brain theory ? 11. Nature Publishing Group: 127–138. doi:10.1038/nrn2787.
26. Sullivan-Bissett, Ema. 2018. Monothematic delusion : A case of innocence from experience Monothematic delusion : *Philosophical Psychology* 31. Routledge: 920–947. doi:10.1080/09515089.2018.1468024.
27. Turri, John. 2010. On the Relationship between Propositional and Doxastic Justification. *Philosophy and Phenomenological Research* 80: 312–326.