Transparent Delusion

Just because I'm paranoid doesn't mean they're not out to get me

— Joseph Heller, Catch-22 —

Abstract: In this paper, I examine a kind of delusion in which the patients judge that their occurrent thoughts are false and try to abandon them precisely because they are false, but fail to do so. I call this delusion *transparent*, since it is transparent to the sufferer that their thought is false. In explaining this phenomenon, I defend a particular two-factor theory of delusion that takes the proper integration of relevant reasoning processes as vital for thought-evaluation. On this proposal, which is a refinement of Gerrans's (2014) account of delusion as unsupervised by decontextualized processing, I can have all my reasoning processes working reliably and thus judge that my delusion is false but, if I cannot use their outputs when revising the thought itself, the delusion will persist. I also sketch how this framework explains some interesting cases of failed belief-revision in the general population in which people judge that ~*p* but nonetheless continue to believe that *p*.

Keywords: delusion; two-factor framework; thought-evaluation; internal incoherence; decontextualized processing.

1. Introduction

In this paper, I analyse a phenomenon that I call *transparent delusion*. I use the term 'transparent' in a strictly linguistic sense: 'readily understood' (Merriam-Webster). On my view, a delusion is transparent in the sense in which someone's business practice can be transparent. Transparently delusional subjects are those who simply see their occurrent thoughts for what they are, i.e., *false delusional thoughts*, but who cannot revise them despite their attempts to do so. A delusional thought that *p*, therefore, is transparent iff the sufferer

1. (Thinks that he) believes that *p*;
2. Judges that his belief that *p* is false;
3. Because of (2), attempts to abandon the thought that *p*; but
4. Cannot (permanently) revise the thought that *p*.

The cases I have in mind are not those typically reported in the literature as involving 'awareness of one's illness' (e.g., Amador et al. 1994; Cuffel et al. 1996; Langdon and Ward 2009). In the study conducted by Amador et al. (1994, 829), for instance, a response that

counts as being 'Somewhat Aware/Unaware [of one's delusion],' is 'I hear voices because of the implant the researchers put in my brain.' These delusions are not transparent to the subject as false beliefs/thoughts but rather as troubling experiences; they do not satisfy condition 2 of transparent delusion. Interestingly, recognizing the implausibility of one's thought need not make the delusion transparent. Consider this case documented by Alexander, Stuss, and Benson (1979, 335).

E: Isn't that [two families] unusual?
S: It was unbelievable!
E: How do you account for it?
S. I don't know. I try to understand it myself, and it was virtually impossible.
E: What if I told you I don't believe it?
S: That's perfectly understandable. In fact, when I tell the story, I feel that I'm concocting a story. … It's not quite right. Something is wrong.
E: If someone told you the story, what would you think?
S: I would find it extremely hard to believe. I should be defending myself.

Mr S. was able to express amazement and disbelief at his contradictory statements. He could not use his awareness to revise his delusion but, nonetheless, he was not transparently delusional. The reason is that S thought neither that his thought is false nor that it should be revised; rather, he 'admitted distress about the implausibility of this story but insisted that it was correct, even though he realized it sounded ridiculous' (Alexander, Stuss, and Benson 1979, 335).

Even those patients who report that they are delusional need not think that their thought is false. Indicative of this trend and in line with Catch-22, a patient who willingly participated in a group therapy defined delusion as (italics added) 'a thing that you believe is true *and that may or may not be true*, but most people believe is not true' (Landa et al. 2006, 13). This patient recognizes that his environment has a different view on the relevant issue and that he is probably mistaken but he does not exclude the possibility that his delusion is true. He seems to be one of those sufferers who seek therapy because they were being told that they are delusional for years or because their condition prevents them from being successfully integrated into their environment, not because they do not want to believe a falsehood. These people do not satisfy condition 3 of transparent delusion. The kind of cases I am interested in are those in which patients desperately want and try to abandon their delusional thought *principally because they identify it as false* and only then (if at all) because it causes them distress or impedes their social interactions.

Transparently delusional subjects are unresponsive to counterevidence or counterarguments in a way atypical for delusion: unlike other sufferers, they accept that their thought is false (condition 2) and, *because of that*, they want to revise it (condition 3) but are unable to do that successfully (condition 4). Transparent delusion is philosophically very important because it suggests that I may continue to believe that *p*, and consciously so, while judging that ~*p*. Therefore, this condition suggests that the idea that I can answer the question of *whether I believe that p* by simply considering *whether p* (e.g., Evans 1982; Peacocke 2000, 71) is mistaken. Also, insights obtained by analysing this condition can help us better understand not only the nature of delusion but also the behaviour of the general population. Even non-delusional people may think that their occurrent thoughts must be false without being able to revise them. I will call this phenomenon closely related to transparent delusion *transparently irrational belief*. On my view, your belief is transparently irrational if you think that the available evidence suggests that it is false and that you should not believe it but you are unable to revise it notwithstanding.

Transparent delusion and transparently irrational beliefs are cases of what Davidson (1982, 1985) famously calls *internal incoherence*, a kind of irrationality in which a person has a set of attitudes which are inconsistent by her own standards. In akrasia, for example, I φ intentionally in the face of my all-things-considered judgement that I ought *not* to φ and my principle that I should only do what I judge all-things-considered as best. According to Davidson, this is a conceptually problematic case of irrationality: 'the reasons an agent has for acting must, if they are to explain the action, be the reasons on which he acted' (Davidson 1982, 173) and, in this situation, I acted on a reason that is not mine; I had better reasons *not* to φ. Davidson's solution is that a semi-autonomous structure in my mind supplies the reason that causes my incontinent action. Naturally, one might think that, if transparent delusion is real, the same solution should be applied.

In what follows (Section 2), I will first discuss a case that, I argue, involves a transparent delusion and then compare it to some cases of transparently irrational beliefs found in the general population. In Section 3, I will inspect how the received theories deal with transparent delusion. I will, then, suggest an account of the relevant thought-evaluation impairment that does not require partitioning the mind (Subsections 4.1 and 4.2) and apply it to the relevant cases (Subsection 4.3). In Section 5, I will offer some brief concluding remarks.

2. The Peculiar Cases of Mr F. and Martin

Below is an excerpt from the interview with Mr F., a 39-year-old male ultra-orthodox follower of Judaism, diagnosed with delusions of grandeur; he thought that he was King David, Moses, Hitler, and Satan (Zislin, Kuperman, and Durst 2011). The interview was conducted in Hebrew. Mr F. was first diagnosed with paranoid schizophrenia at the age of 21. Since then, he experienced multiple psychotic episodes and had a history of numerous serious suicidal attempts. In a complete psychotic state, Mr F. took his own life two years after the interview (italics and numbering added).

F_1: They want to kill me, I fear, because *I am* the Satan.
J_1: *They* think you are the Satan, or *you* think that you are the Satan?
F_2: Sometimes *I think* I am the Satan …
J_2: What is Satan?
F_3: I can be like Moses, or like Satan. You understand, doctor?… I am a very sick man. The best thing you can do to me is to kill me. I would like to have a normal life, like everyone. […] But I cannot manage with a simple life because *I always have thoughts of grandeur in my head*. […]
  J_3: What are those thoughts?
  F_4: [Thoughts] that I can *be* a very big person in this life.
  J_4: Like who?
  F_5: Like King David, and people like that…
J_5: Is it not good to be *like* King David?
F_6: It's frightening… I can be like Hitler.
J_6: King David and Hitler, who else?
F_7: That's it. These are the two persons. *I am afraid to think that I can be like Hitler, since I am a Jew*.
J_7: How is it to feel like Satan?
F_8: I want to ask [for] forgiveness from the entire mankind. Yes, *I am* Satan.
J_8: What can you do as Satan?
F_9: To ask [for] forgiveness. Can you kill me?

Understanding the claim 'being *like x*' requires additional effort. Mr F. seems to have used this phrase in, at least, two different meanings during the interview and sometimes even by automatism, namely, with no specific meaning attached to it. For the first interpretation, I draw your attention to the indented part of the interview, J_3–F_5. Mr F. says that he can 'be' (rather than 'be like') a very big person. '*Like* who?' – Zislin asks. '*Like* King David, and people *like* that' – F. replies. Considering Zislin's question, 'like' here means 'for example' in the first instance and, '[people] similar to [David]' in the second.

However, this analysis need not generalize to the whole interview. True, in F_7, Mr F. could be saying that he has different reasons against each thought, and that being Hitler, *for*

example, is impossible because he is a Jew, but this is not the only way to look at this particular claim. Another interpretation is that Mr F.'s thought that he can be *like* Hitler explains a compelling illusion, where everything is *like x* but it is not *x*. While this hypothesis may explain F_3, I do not think it explains F_7, and I will argue against it in Section 3.2 in detail. Here, I draw your attention to the fact that J uses 'like' in two different meanings in J_4 ('for example') and J_5 ('as if,' 'comparable,' 'resembling') and the discussion from J_5 and F_6. In J_5, Zislin asks 'Is it not good to be *like* King David?'. 'It's [being *like* David] frightening,' Mr F. replies in F_6, and continues with 'I can be *like* Hitler.' In F_6 and F_7, Mr F. is merely following the way Zislin describes Mr F.'s condition and uses 'like' by automatism. That is, Mr F. is not giving his evaluation of his own condition ('I suffer from an illusion where everything is *like*'); rather, he is just continuing the conversation in the way Zislin frames it.

None of this is to say that Mr F. did not understand his condition. Although Mr F. is obviously confused at some points, at some other points (e.g., in F_5), he was essentially saying that some of his thoughts are false but that he cannot fight them off. In fact, the recognition that his thoughts are false – *even those he is currently having*, such as 'I *am* the Satan' (F_{8b}) – was preoccupying his attention: his explanation from F_3, which should answer the question from J_2, is actually an elaboration of F_2, not an answer to J_2 (F. was talking to himself). The case, thus, fits the suggested description of transparent delusions: Mr F. is afraid of thinking that he is Moses or David not because being Moses or David is unpleasant but rather *because he knows that he is neither*.

Having said that, one may think that his delusional thought about Hitler is terrifying for a different reason: it could be that Hitler is a dangerous person for Jews, F. is a Jew, and he does not want to be (like) Hitler who kills fellow Jews. In short, it could be that F.'s desire to stop believing this proposition is motivated by fear or guilt rather than by his judgment that the belief is false, which would make the thought an unwanted intrusive thought. True, the fact that Hitler hated Jews probably also played its role and F. does seem to be afraid of being like Hitler, but it had to play a subordinate role to the fact that F. knew that you cannot be Hitler if you do not share all properties with Hitler. This is a very simple inference:

1) Hitler is ~J (Jew);
2) I am J;
3) Therefore, I am not Hitler.

This inference relies on an application of Leibniz's law and I do not see why one would deny that Mr F. realised that his thought is false by simply applying this law. He could have reasoned in the following way as well: Hitler hates Jews; I am a Jew and I do not hate them; therefore, I am not Hitler. Notice, he used the example of being a Jew but he probably would have made the same judgement if he were to discuss Hitler's height, age, or some other feature they do not share. Finally and most importantly, the fact that Mr F. uses the same line of reasoning in analysing his other thoughts also supports my interpretation. Moses and David did not hate Jews and F. was equally afraid of being Moses, David, and Hitler: being like King David was 'frightening' for him (J_5–F_6). The most obvious explanation is that he realised that he is not Moses or King David by applying Leibniz's law.

All in all, I do not deny that F. was motivated by fear to stop believing that he is Hitler or Satan. What I deny is that Mr F.'s fear was caused by the unpleasant nature of being (feeling) *like* Hitler or Satan. Rather, I argue that Mr F.'s fear was principally and perhaps even exclusively generated by the fact that *F. knew that he is neither of them*. Therefore, F. did not suffer from unwanted intrusive thoughts, which are unwanted by being inconsistent with the sufferer's cherished values and sense of self rather than by being recognized as false. People suffering from unwanted intrusive thoughts may continuously repeat that their thoughts are false but they do not actually realise that they are false. In contrast, Mr F. knew that he cannot be someone who is not a Jew (Hitler) or someone who is dead (Moses, David). His desire not to believe what (he thinks that) he believes is a consequence of his perfectly sound reasoning. Mr F. did not fail to reason correctly; he merely failed to apply his reasoning to his delusional thoughts in a way in which this would prompt revising the thoughts.

One may object that this is an unreasonable interpretation of the case. However, the same kind of thought-revision failure occurs in the general population as well. And, given that these cases are much less bizarre, it is easier to defend the general idea that judging that ~*p* entails neither believing that ~*p* nor unbelieving that *p*. Consider this case (Lackey 2007, 508).

> A racist, *Martin*, was called to serve on the jury of a case involving an African American on trial for raping a white woman. Even though Martin realises that the evidence that the defendant is innocent is compelling and that the defendant could not have committed the crime, he cannot help believing that the defendant is guilty. Moreover, Martin realises that it is his racism that leads him to this belief, which he nevertheless still holds with very high confidence. Shortly after leaving the courthouse, Martin bumps into a childhood friend who

> asks him whether the 'guy did it.' Despite the fact that he does not believe that the defendant in question is innocent, Martin asserts 'No, the guy did not rape her.'

Martin's credence in the proposition that the suspect is guilty remains high but Martin thinks that his conviction is epistemically unjustified: the evidence that the defendant is innocent is compelling. Furthermore, because he thinks that 'the defendant could not have committed the crime,' Martin most likely also believes that his racist belief is unjustified, perhaps even false, and that he should not believe it. Martin's belief is transparently irrational and he exhibits internal incoherence. Specifically, he appreciates that the evidence he has *is* evidence against his belief, he holds that the evidence against outweighs the evidence for his belief, and he thinks that the hypothesis supported by the totality of evidence should be believed, but he still believes that the guy did it (see Davidson 1985: 346). Martin is reasoning correctly, he just cannot utilise this awareness to revise his racist belief. The same explanation applies to akratic believing: one believes that *p* despite one's conscious recognition that one's belief is unjustified and irrational. So how to understand this phenomenon? I will focus on Mr F.'s condition.

3. Understanding Mr F.'s Condition

In this section, I will discuss Mr F.'s case by considering some theories of delusion. I begin by briefly addressing some interesting theories of delusion formation (Section 3.1). If I have convinced you that transparent delusion is a genuine phenomenon, then you will see this analysis as a way of discovering which theories of delusion are more and which are less successful. And if I did not convince you that transparent delusion is real, you may see this discussion as supporting your view further. In any case, this analysis will further explain the nature of Mr F.'s condition. Then, I proceed to consider some general theories about the nature of our mental states that might be used to explain transparent delusion (Section 3.2). In section 4, I will present a view that can easily explain not only this phenomenon but also the failures of thought-evaluation we find in transparently irrational beliefs and other, non-transparent, cases of delusion. Therefore, this should be our preferred theory of delusion based on a broader theory of thought-evaluation.

3.1 Aetiology of Delusion

Delusions are most commonly explained using two-factor frameworks, which posit impaired thought-formation (first factor) and impaired thought-evaluation (second factor).

Nevertheless, one-factor frameworks, which do not require impaired thought-evaluation, are not uncommon (most notably, Maher, e.g. 1999). Mr F.'s case, however, sits uneasily with one-factor proposals. They assume that the delusional response to abnormal experience is within the normal range for human psychology, even if it is an irrational one (the irrationality in play is not clinically significant), but Mr F. himself thinks that his beliefs are not only irrational but also undoubtedly false and he desperately tries to revise them. Mr F. is internally incoherent in a very serious sense, much more serious than people suffering from non-transparent delusions. We need an account of impaired thought-evaluation (the second factor) in order to explain the aetiology and maintenance of transparent delusion. In particular, we need to know why F's conscious judgement that his delusion is false is ineffective with respect to that very thought.

The account of the second factor depends on the theory in question, since the impaired thought-evaluation can be understood in various ways. For example, Langdon and Ward (2009) write that those suffering from delusion are not capable of approaching their thoughts from a third-person perspective; Parrott (2016) writes that delusional patients cannot correctly appraise the thought's epistemic possibility (the likelihood of something like that being possible); while Parrott and Koralus (2015) say that they are incapable of endogenously raising all relevant questions.[1] Each of these hypotheses explains the person's failure to detect that her thought is inconsistent with background knowledge. Therefore, they can tell us why delusions are normally not responsive to reasons but not why transparently delusional people have a correct insight into their condition and why this insight does not prompt a revision of the delusion.

According to Coltheart, Menzies, and Sutton (2010), delusional people accept the delusional hypothesis because the good fit between hypothesis and abnormal data trumps the overall implausibility of the hypothesis. The delusion is maintained because they are not giving sufficient weight to the contradictory data: counterevidence is interpreted in the light of the delusion. McKay (2012) objected to this model arguing that it is not rational to accept the delusional hypothesis (e.g., 'I am dead'), given that its probability before the abnormal data is rarely high. Instead, he proposes that the patients have a bias of explanatory adequacy; namely, they exaggerate observations in favour of a delusional hypothesis (they discount the

[1] Parrott (2016, 291–293) allows that there are more than two factors involved in the aetiology of some delusions, but this is tangential to my argument.

prior probability ratio). Neither of these proposals captures Mr F.'s condition: he neither interprets counterevidence in the light of the delusion nor discounts the prior probabilities. He knows that he cannot be Hitler et al.

Davies and Egan (2013) reason that delusional patients exhibit a failure of belief-compartmentalization as a result of which their delusion becomes fully integrated.[2] This integration of the thought, in turn, changes prior credences of other thoughts and thereby prevents the thought's revision – other beliefs now inferentially support the delusion. However, Mr F.'s delusions do not appear to be inferentially supported by any of his beliefs. Aimola-Davies and Davies (2009) appealed to impaired working memory that allows manipulation of information in a way in which an improbable hypothesis becomes the best explanation, but this proposal also sits uneasily with Mr F.'s behaviour. True, he did occasionally lose sight of reasons that would make him think that he is not Satan and so 'I am Satan' became the best explanation of his experience but, at other times, his thought that his thought is false *while* exercising it as true. Therefore, appealing to a compartmentalisation of beliefs or impaired working memory can only be a part of the explanation, but not the important part.

Turner and Coltheart (2010) posited that the impaired thought-evaluation consists of failures of unconscious and conscious processes that constitute two monitoring frameworks that check occurent thoughts. Even if the unconscious checking system 'passes' the delusion, the conscious checking system still may examine it. Our conscious monitoring framework consists of various abilities. One is the ability to make plausibility judgements, namely, the ability to judge whether thoughts are plausible in the context of everyday knowledge about the world; another is the ability to conduct reality monitoring, namely, checking whether a thought represents an externally derived experienced event or an imagined event; and so on. This theory can explain many cases of delusion but not Mr F.'s delusion: he knew that he was not Moses, David, Hitler, or Satan. He was able to detect that the thought is false but was unable to *do* anything about it.

Unlike other sufferers, transparently delusional subjects have optimal hypothesis-testing strategies; this is why they know their thoughts are false. They are just unable to apply them to their thoughts in a way in which the realisation that the thought is false will prompt a

[2] This not a complete account of their description of the second-factor but it is its relevant feature.

revision of the thought. One explanation could be that these strategies cannot be applied to transparent delusion simply because the thought is insufficiently reason-responsive and that, therefore, no second-factor impairment should be posited. Mr F.'s thought-evaluation mechanisms are not impaired; rather, his thoughts are not of the kind that can be revised. I now proceed to test this explanation by examining some proposals about the nature of delusional states.

3.2 Nature of Delusion

Let us consider some analyses according to which Mr F.'s doxastic situation would not require revising the delusion. It could be that F. was fluctuating between beliefs rather than simultaneously believing contradictions. This hypothesis, however, cannot easily explain F.'s persistent state of desperation – 'Can you kill me?' (F_9) – that eventually made him take his own life. Mr F. was desperate because he 'always [had] thoughts of grandeur in [his] head,' thoughts that he 'can *be* a very big person in this life' (F_3 and F_4). People who fluctuate between beliefs do not experience this kind of despair. Compare Mr F. to G.R, who was genuinely fluctuating between beliefs (Coltheart 2007, 1053–1054) (italics added).

> G.R.: The *lady* knows me way back. She couldn't say things that happened 40 years ago, and I wonder where she gets them from. And then I worked it out and I've wondered if it's M. all the time. It's nobody else. [M] keeps going back and then going somewhere, but where does she go? She disappears from me.
> M. (interjects): Probably into the garden patch.
> M.C. [examiner]: This is a good example. So since this person knows things from your past life, it must be M., because no-one else would know that.
> G.R.: Yes, I figured that out, professor, it couldn't be anybody else.
> But having expressed a rejection of the impostor belief, G.R. continued *immediately* with:
> G.R. And since I started that, I've noticed that many of the things in her name, it's like M. speaking all of the time. Like M. will get a list and say, 'I'm going shopping tomorrow G., do you want anything? I'm getting this, this, and this.' And then a minute later, she [the '*lady*', *not* M] will come in and say, 'I'm going shopping tomorrow, do you want this, G.?' […] And she [lady] goes away to get the shopping, comes back without it, and M.'s got it. And I don't know why she's [lady] not got it for me [M. has it]. Or where she's been. She never gets any money off us. And never asks us for any clothes or anything, *does she, M.? Yours are the same clothes as her, aren't you?*

At one point, G.R. realises that the lady is his wife but continues his story as if nothing happened; he even asks his wife about the lady: 'She … never asks us for any clothes or anything, *does she, M.*' In the first part of the interview, G.R. does not attend to the thought that the lady is *not* M and, in the second, he does not attend the thought that the lady *is* M.

We can see that Mr F. and G.R. experience two different kinds of bewilderment. G.R. wonders what is going on: he wonders who the lady is and how does she have all these features she should not have. In contrast, Mr F. wonders why he constantly has disturbing false thoughts that he should not be having. G.R. is fluctuating between beliefs; Mr F. is internally incoherent.

A plausible alternative is that Mr F. was simultaneously exhibiting contradictory dispositions, not beliefs. A phenomenal dispositionalist would say that F. is 'in-between' believing that *p* and that the question of what he believes has no determinate answer (Schwitzgebel 2002, 2012; Tumulty 2012). This proposal, however, does not tell us why this is the case and, more importantly, what this actually means. Saying that he has dispositions to think that he is David, feel like David, and think that he is not David brings no explanatory benefit because it tells us nothing new about F.'s condition. In addition, his despair seems particularly odd on this hypothesis. The claim is not that people that are in-between believing propositions should not exhibit distress but rather that Mr F.'s level of distress is exceptionally high for someone who is supposedly in-between beliefs. This is not a decisive consideration against the view but it is a good reason to seek a better one.

And indeed, one need not endorse a dispositionalist account in order to be able to explain simultaneously exhibited inconsistent behaviour. Quilty-Dunn and Mandelbaum's (2018) psychofunctional, representational theory of belief – combined with a plausible theory of the mind as compartmentalised (e.g., Egan 2008) – explains some situations in which a person simultaneously believes that *p* and that ~*p*. Consider Schwitzgebel's (2002, 260–261) example of a mother who asserts that her son does not smoke marijuana, and yet feels suspicious when he comes home red-eyed late at night and she is apt to tell her therapist that she is worried about her son's marijuana use. When this mother is reminded of teenagers smoking pot, Quilty-Dunn and Mandelbaum (2018) argue, she has two relevant, inconsistent beliefs activated simultaneously: she both believes that her child smokes pot and that her child does not smoke pot. The mother exhibits dissonance (the state she is in hurts) but the dissonance itself will not necessarily cancel one of the two beliefs. The mother has two options: she may *resolve* the dissonance (by abandoning one belief) or *expel* (assuage) it by focusing on something else, for instance.

This is an excellent explanation of many problematic cases but I do not think it can explain transparent delusion or transparently irrational beliefs. A parent who believes both that their child smokes pot and that their child does not smoke pot believes this because they think that they have good reason to believe both propositions *and* they have expelled their dissonance; the mother turns her head away from the distressing belief. In contrast, Mr F. thought that he should not believe that he is Hitler or David, since the former is not a Jew and the latter is not alive, and he desperately tried to resolve the chronic dissonance rather than expel it.

The fact that transparently delusional patients are unable to apply their rational reasoning only to their delusional thought in a way in which this triggers thought-revision may suggest that their delusions are insufficiently reason-responsive (they are neither beliefs nor sufficiently belief-like) and that, thus, they cannot be revised by way of reasoning. Many philosophers argue that delusions are not beliefs; rather, they might be imaginings (Currie 2000; Currie and Jureidini 2001; Currie and Ravenscroft 2002), faulty perceptual inferences similar to illusions (Hohwy and Rajan 2012; Hohwy 2013), default thoughts (subpersonally generated simulations similar to imagination) (Gerrans 2014; see 4.2), or even bimaginations (states 'in-between' belief and imagination) (Egan 2009). And indeed, it seems natural to think that patients who know that their thoughts are false but cannot 'unbelieve' them never believed them in the first place.

Taking it that Mr F. has mistaken his imagining for a belief seems like a nice explanation, but I think that it is incorrect. It is one thing to misidentify your imagining (illusion, or default thought) for a belief if you do not know that what you imagine is not true, but it is quite another to misidentify it for a belief while consciously judging that what you imagine is false. Generally, we think that to believe that *p* is false just entails believing that we should not believe that *p*. Therefore, knowing that *p* is false, Mr F. should at least think that he does not believe that *p* but that he merely imagines that *p*. Moreover, believing that you believe that *p* while consciously believing that *p* is false is as problematic as believing both that *p* and that your belief that *p* is false. And Mr F. did believe that he believes his delusions. To illustrate my point, let us consider the case of HS (Chatterjee and Mennemeier 1996, 226–227) who was anosognosic for hemiplegia for a week and was interviewed after the delusion has resolved (italics added). This is how a case of a transparently delusional subject whose delusions are insufficiently reason-responsive looks like.

E: What was the consequence of the stroke?
HS: The left hand here is dead and the left leg was pretty much.
HS: (later): I still *feel as if* when I am in a room and I have to get up and go walking … I just *feel like* I should be able to.
E: You have a belief that you could actually do that?
HS: *I do not have a belief, just the exact opposite*. I just have the feeling that sometimes *I feel like* I can get up and do something and *I have to tell myself 'no, I can't.'*

HS had the feeling that he could get up but he knew that this is not true, which is a point of similarity with the behaviour of Mr F., who, by being Satan, wanted to ask for forgiveness while knowing that he was not Satan. However, HS knows that he does not believe his delusion: 'I do not have a belief, just the exact opposite.' Furthermore, because his delusion is insufficiently belief-like, *HS can control it*: 'I have to tell myself "no, I can't [get up]".' In contrast, Mr F. was unable to control his condition; rather, he was *afraid* to think that I can be King David (F_4–F_5) even though he knew that he was not David. Notice that HS clearly uses '*like*' in the sense of 'everything is like *x* but it is not *x*.' This is far less clear in the case of Mr F.

I have no intention to defend the view that Mr F.'s delusions were fully-fledged beliefs or that transparent delusions are beliefs. I do think that some transparently delusional subjects do not believe their delusions. I already mentioned HS, his delusion is transparent but the thought is insufficiently belief-like, and the condition of John Nash seems like another good example. Nevertheless, I do reject the idea that Mr F.'s thoughts were insufficiently reason-responsive or belief-like. If this were the case, he would not have been bothered by the fact that these thoughts are false and he would have been able to control them, like Nash, or even stop them, like HS.

I will now present the view that nicely explains transparent delusion and can be generalised to explain similar thought-evaluation failures discovered in the general population, namely, it can explain the existence of transparently irrational beliefs.

4. Faulty Decontextualized Processing of a Thought

4.1 The Theory: Introduction

Two-factor theorists mainly argue that, in delusional patients, some of the relevant analytic reasoning processes are impaired; the patients cannot evaluate some domains of reality (the second factor). However, because transparently delusional subjects correctly see

that their thought is false and, because of that, want to abandon it, it must be that their relevant reasoning processes are working correctly and that the revision of the thought was initiated. What seems to have gone wrong is the way in which some of these reasoning processes were *used* in the revision of the delusional thought. In order to understand this error, we need a detailed account of the thought-evaluation process itself, not just of various abilities used in it. I suggest the following explanation.

Our thoughts are formed in contexts and therefore they come with various associations relevant to the contexts in which they are acquired. Therefore, in order to subject a particular thought to *reality testing* (i.e., analysing through different domains of reality), the thought must first be *decontextualized*, which is to say that all of the contextual associations must be (sort of) suspended or purged. After that, an organism can effectively reality test without the associated 'stuff' interfering. The final process of thought-evaluation, then, can be understood as having two steps. In *the first step*, the thought gets purged of its contextual and narrative associations (it gets decontextualized) and, in *the second*, it gets analysed through different domains of reality (it gets reality tested) (see, Gerrans 2014). Having this brief description of the decontextualized thought-processing in mind, let us consider transparent delusion.

Unlike other sufferers, transparently delusional patients recognize that their thought is false, which implies that the processes constitutive of thought-revision are working reliably and that their reasoning is not erroneous. And because these subjects are examining their thought from various perspectives (Mr F. considered what it is to be David or Satan), it must be that both steps of the thought's decontextualized re-evaluation have been initiated. However, because the thought was not revised notwithstanding the processing it was subjected to, it must be that the thought-revision was unsuccessful and that the thought was passed as correct. The general explanation is, I suggest, that *the initiated thought-revision was incomplete*. In particular, the thought was successfully purged of contextual and emotional cues, the first step was complete (otherwise F. would not be able to consider what it is like to be Satan), *but* those outputs of the analytic reasoning processes that signal that the thought is false have been left out when the thought was analysed through different domains of reality, the second step was incomplete. As a result, the transparently delusional thought was passed as correct even though the relevant analytic reasoning processes were producing outputs that contradict it.

The standard hypothesis that delusional patients do not initiate thought-revision can explain most cases of delusion but not the transparent delusion. Transparently delusional patients actually initiate thought-revision but they, nevertheless, cannot revise the thought. Notice that F. does not say '*I should not think* that I am David (et al.),' which would signal that the thought was at least tagged as suspicious; rather he says that it *is frightening to think this*: the thought was passed as correct by his thought-revision processing in the face of the judgment that F cannot be Hitler! From all of this, I infer that Mr F.'s dorsolateral prefrontal cortex is capable of detecting abnormalities in the thoughts generated by the ventromedial prefrontal cortex (see, Gerrans 2014; section 4.2) and that it initiates the thought-revision when this is deemed as necessary, but that it is unable to conduct the process thoroughly; specifically, *it fails to integrate into the thought-revision process* (the second step) *outputs of those reasoning processes that signal that the thought is false* in the first place. As a result, the thought was 'passed' as correct by the processing and the delusions remained.

An immediate concern that arises with respect to my hypothesis is that it seems odd to treat Mr F.'s occurrent judgment as just another input to the decontextualized thought-revision that is still only partially recovered. This processing is presumably in the service of what we might ordinarily call 'Mr F. making up his mind,' namely, revising his belief according to the evidence. However, the most common view is that, once he judges that he is not David or Moses, F. has already made up his mind. Therefore, treating either that completed judgment, or the act of so judging, as just another component of the thought-revision seems odd. That is, it seems odd for F. to treat the fact *that he himself judges that he is not Moses* as just an inconsistent-with-delusion but not useable-in-its-re-evaluation piece of information and that his delusion is not being tested with respect to the domain of reality responsible for this judgement (i.e., epistemic possibilities of states) in its re-evaluation.

In reply, I note that one of the main premises in my argument, and the main feature of *Mr F.* and *Martin*, is that *judging that p does not entail acquiring the belief that p* or even unbelieving that ~*p*. I am not sure whether this entails that judging that *p* is not identical to making up your mind with respect to *p* (it certainly is not making up one's mind in the sense in which this entails believing or unbelieving), but – contra the standard view – I do think that we have convincing reasons *not* to think that judging that *p* entails believing that *p* or unbelieving that ~*p*. If judging that *p* was identical to either of the latter two states, then not only that Mr F. would not think that he is not David et al. but also Martin would correctly

believe that the guy did *not* do it. However, Martin judges both that it is his racism that leads him to his belief and that the guy certainly did not do it, and yet he continues to believe that the guy did it.

What I argue, that is, is that my hypothesis straightforwardly follows from the cases discussed here. Is it not obvious that Martin is not applying his correct conscious reasoning to his belief? And is it not obvious that his correct reasoning *co-exists* with his epistemically and inferentially unjustified belief? And is it not obvious that this kind of failure is not caused by a lack of intelligence or failure of reasoning? I think that the answer to all of these questions is an unqualified 'yes,' Martin recognises his internal incoherence, and, if so, then my hypothesis straightforwardly follows from the cases. Martin knows that his belief should be revised and he probably wants to revise it. He just does not know how to do that. Plausibly assuming that the racist belief was not acquired on rational reasons, it does not seem right to think that Martin actually could revise it by appealing to reasons and reason.

I know that, from a theoretical perspective, my proposal that you can judge that *p* while believing that *~p* seems odd, but this intuition is generated by a reasoning fallacy. Evans (1982) influentially argued that I can answer the question of *whether I believe that p* by simply considering *whether p* (also, Peacocke 2000, 71). Shah (2003, p. 447) arrives at the same conclusion by analysing Transparency, the thesis that (italics added) '[t]he question of whether to believe that *p* is settled by, *and only by*, resolving whether *p* is true.' However, the view that the question of *whether I believe that p* is determinately answered by considering *whether p* is mistaken: from the plausible idea that I answer the question *whether to believe that p* by considering *whether p* it merely follows that, by answering the question *whether p*, I answer the question whether I *ought to* believe *p*; it does not follow that I actually believe *p* (similarly, Sullivan-Bissett and Noordhof 2019, p. 3).[3] The answer to *whether to believe p* is normative, it tells us when the belief is correct, whereas the answer to *whether p* is factual, and we cannot deductively draw factual conclusions from normative premises. Therefore, one may believe that *~p* when one thinks that one ought to believe that *p*.

Not only that my premise is not odd it is also not *ad hoc*. Mr F. and Martin are not the only examples. Akratic beliefs are another instance of internal incoherence in which judging is

[3] While Sullivan-Bissett and Noordhof (2019, p. 3) think that to answer *whether p* is to answer whether it is *permissible* to believe *p*, they (2019, n. 11) seem to think that one cannot judge that *p* while not believing it. Therefore, the existence of transparent delusion spells trouble for their view as well.

inconsistent with believing. Also, many delusional patients exhibit a kind of 'logical partitioning' similar, if not identical, to that in which one's judging that *p* does not cause believing that *p*. Recall Mr S. (Section 1), who admitted that it is virtually impossible that he has two families (he could not believe that he believes that) but who could not utilise this awareness to revise this delusion. Judging that he cannot have two families surely did not involve 'making up S's mind' in the sense that involves revising the delusion. Similarly, another patient admitted that it was 'logically impossible' that he has multiple heads and bodies but that 'the feeling was too real to have been a dream' (Weinstein et al. 1954). This man's preconscious feeling of rightness trumps the power of his conscious reasoning in the same way in which Martin's racist credences trump his conscious reasoning.

4.2 The Theory: More Detail

My proposal is a plausible refinement of the account of delusion put forward by Phil Gerrans (2014). According to Gerrans, delusions are not genuine beliefs but are rather *default thoughts*, subpersonally generated simulations similar to imaginings. Default thoughts are produced by the default mode network (DMN), which is a powerful simulation system that evolved to allow humans to simulate experiences. The DMN is behind 'mental time travel,' daydreaming, dreaming, and (according to Gerrans) delusion. Because the DMN supervises lower-level, more automatic processes that are not fixed by standards of consistency or empirical adequacy, it is supervised by higher-level, decontextualized processing, which revises subjective narratives (default thoughts) to fit reality by purging them from contextual cues and by analysing them through different domains of reality.

On this view, delusions arise when the default network simulates scenarios in response to salient abnormal input caused by a relevant (kind of) cognitive impairment. Once the relevant default thought has been generated, the thought and its associations are triggered on the next occurrence of the precipitating experience. What then happens is that the salience system, which determines which information stays in the background, allocates resources to the default processing of the delusion-related information; the thought assumes an active role in the DMN. Finally, because the decontextualized supervision is absent, reduced, or abnormal, the delusion runs unsupervised. In short, in delusion, the default cognitive processing is monopolized by hypersalient information because it is unsupervised by the decontextualized processing.

The immediate uptake of this account is that the falsity, fixity, and extraordinary aspects of delusion result from the nature of default thoughts rather than some reasoning failures. Default thinking is not intrinsically a reasoning process. Default thoughts are characterized by *subjective adequacy* rather than truth, accuracy, public acceptability, or verification (Gerrans 2014, 69, 74–75). Therefore, a delusional thought is unresponsive to evidence and other beliefs (a) because its nature is such that it need not respond to them and (b) because the system that should supervise it does not do this.

This theory easily explains many cases of delusion but it needs to be slightly modified to explain the case of Mr F. Gerrans (2014, 75) writers:

> To verify these subjectively adequate narratives [i.e., unsupervised default thoughts], the subject has to treat them as hypotheses about the nature of the world or the causes of experience and then engage in the process of confirmation or disconfirmation [reality test, the second step]. To adopt that perspective on the information represented in the story is to be able to *decontextualize* [the first step].

Mr F. is testing his delusions just as he would test any other descriptive or theoretically anchored belief; therefore, the above solution cannot be directly applied. From the features of the case, I infer that the process of decontextualized thought-revision (1) was initiated, (2) that it was incomplete, and (3) that, therefore, Mr F.'s relevant thoughts were passed as correct when they should have been revised. I assume that the error is in the second step of thought-evaluation because Mr F. analyses his thoughts purged of contextual or emotional cues. Finally, because default thoughts may become beliefs or sufficiently belief-like if the mind judges that they are reasonable, and Mr F.'s thoughts were passed as correct, it is safe to infer that they were sufficiently belief-like.

My modification of Gerrans's theory is small but important. According to Gerrans, delusions *bypass* the processes of rational belief fixation; they run unsupervised by the decontextualized thought-evaluation. According to my proposal, *transparent* delusions (but not all delusions) do not bypass the process of rational belief fixation; rather, they go through it unrevised (they '*pass*' it) when they should not have. Therefore, transparent delusions are supervised by the decontextualized processing, but incorrectly. This modification allows us to apply the view to beliefs, not just to default thoughts: If the thought is not supervised by the decontextualized processing, then it is still a default thought. However, if the thought is passed as correct by, then it becomes less simulation-like and more belief-like.

Importantly, the hypothesis that I defend is not put forward as a general account of delusion but rather as an addition to a comprehensive analysis of delusion and thought-evaluation in general. Two-factor theories should not posit only one kind of thought-evaluation failure when there are so many things that could go wrong. In fact, perhaps we do not even need a second factor to explain all cases of delusion – some delusions may indeed be within the normal range for human psychology – but, when we do, I hypothesise that thought-evaluation can go wrong for the following reasons:

1. Decontextualized supervision is absent or corrupted (Gerrans 2014),
2. Some analytic reasoning abilities are impaired (the standard second-factor proposal),
3. Thought-evaluation is initiated but it is unreliable and, as a result, it passes the thought when it should tag it as suspicions. This can happen for three reasons:
    a. The system was unable to adequately decontextualize the thought (i.e., purge contextual and emotional cues),
    b. The system decontextualized the thought successfully but it was unable to conduct the reality testing correctly: some reasoning abilities are working reliably (i.e., not a case of a number 2 failure) but, for some reason, they are not used in the thought-evaluation; therefore, the thought was not analysed through all relevant domains of reality,[4]
    c. Both (a) and (b) in some version,
4. A relevant combination of these factors.

Transparent delusion and transparently irrational beliefs are explained by the thought-evaluation failures of the third kind. Transparently irrational beliefs are insufficiently decontextualized and, because contextual cues obstruct the evaluation, the reality testing is ineffective (3a). Transparent delusions are reliably decontextualized but they are not analysed through all relevant domains of reality (3b). Neither of these thought-evaluation failures involves a reasoning error. The failures are of the *mechanisms that apply one's reasoning to the thought*. I will now apply my theory to the cases discussed in this paper.

4.3 Applying the Theory

Mr F. does appear to have been running his properly through different domains of reality – as in 'Hitler is not a Jew,' 'Satan did bad things [hence, asking for forgiveness],' and so on – without being able to use this ability to revise his delusions. My explanation is that even though his analytic reasoning processes were working reliably, the hypersalient information monopolized default cognitive processing and thus the delusional thoughts were regularly generated – the first factor. The thoughts were not revised because they were subjected to an

[4] The question 'Why they were left out of this process?' can only be answered by a future project.

incomplete decontextualized thought-revision processing that continued to pass them as correct – the second factor. The processing was incomplete because, even though a particular thought was properly decontextualized, the judgement that signals that the thought is false was left out of the reality testing. Because the thought-evaluation passed the delusion as correct and because Mr F.'s individual reasoning processes were working reliably, the delusion and the judgement that it is false were able to co-exist. Let us now reconsider a part of Mr F.'s interview. My analysis is added in brackets.

F_1: They want to kill me, I fear, because *I am* the Satan.

(The thought was passed as correct before F_1. He thinks that he is Satan and we even see an unelaborated secondary confabulation in F_{1a}).

J_1: *They* think you are the Satan, or *you* think that you are the Satan?

(The doctor is pointing out that the thought is false).

F_2: Sometimes *I think* I am the Satan …

(Prompted by J_1, Mr F. realises that he cannot be the Satan; he merely thinks he is. The relevant analytic reasoning processes are working reliably and the reasoning is sound).

[…]

J_7: How is it to feel like Satan?

(The doctor cleverly draws F.'s attention to the thought-evaluation procedure: 'go through everything relevant for determining the truth of your thought that you are Satan.' However, unbeknownst to both of them, the procedure is faulty).

F_8: I want to ask [for] forgiveness from the entire mankind. Yes, *I am* Satan.

(Mr F. conducted the faulty thought-evaluation procedure, the first statement presents one simulation from the 'Satan' model of the world that F. rendered while analysing the thought, and the result is 'pass').

J_8: What can you do *as* Satan?

(This question initiates the reasoning processes *used* in the thought-evaluation procedure; the thought was not analysed with respect to the domain of reality responsible for the conscious judgement. Notice, J.'s question *presupposes that F. is Satan*, an assessment that excludes this judgement).

F_9: To ask [for] forgiveness. Can you kill me?

(F_9 is extremely interesting and requires more effort.)

$F_9 = F_1 + F_2 + F_8$

'To ask [for] forgiveness.'

('I am the Satan.' See J_8 and my comment.)

'Can you kill me?'

(Asking this question *right after* affirming that he is Satan indicates that Mr F. surrenders to his condition. It is as if he says 'My thought is correct [I should ask forgiveness] but I am not the Satan [which means it's false]. But, how can the thought be both false and correct? Make it stop! ["Can you kill me?"]').

Mr F.'s thought-evaluation process can be reconstructed in four steps. In step I, he thinks that he is Satan (F_1). In step II, he realises that it is impossible that he is Satan. In step III, he initiates thought-revision, which he conducts during J_8. In step IV, the process passes the

thought as correct (F_{8a}: 'Yes') and, as a result, he consciously endorses the thought passed as correct (F_{8b}: '*I am* Satan') rather than unbelieving it. F_{8b} is identical to F_1; the vicious circle has been completed and Mr F. exhibits internal incoherence: his conscious belief-like thought is directly inconsistent with his conscious reasoning.

Transparently irrational beliefs also involve internal incoherence. Martin realises that the evidence he has *is* evidence against his belief (that the guy did it), he holds that the evidence against outweighs the evidence for his belief, and he thinks that the hypothesis supported by the totality of evidence should be believed, but he still believes that the guy did it (see, Davidson 1985, 346). Davidson (1982) would explain this condition by appealing to semi-autonomous structures in Martin's mind: one structure walls off Martin's judgement and the principle that he should believe propositions supported by evidence. However, my analysis allows us to accept Davidson's main premise – Martin is irrational from his own perspective – without partitioning the mind psychologically.

What explains transparently irrational beliefs is the fact that context is the defining feature of some thoughts and so these thoughts cannot be effectively decontextualized. As a result, the contextual cues will interfere with the reality testing and thus one's all-things-considered judgement will be ineffective with respect to the given thought. If the interference is strong enough, judging all-things-considered that *p* or that one ought to φ will not be sufficient for one to acquire the belief that *p*, revise the belief that ~*p*, or form the intention to φ even if one thinks that this should happen. Martin, for instance, genuinely approaches the situation as objectively as he can. He is unable to revise his belief that the guy did it in the light of his conscious judgement because the racist belief cannot be purged of its racist associations, and these associations are unresponsive to reasons.

Also, consider the case of the mother who supposedly believes both that her child smokes pot and that it does not. Quilty-Dunn and Mandelbaum (2018) write that, when a mother is reminded of teenagers smoking pot, she has two relevant, dissonant beliefs activated simultaneously. This explanation is plausible when the mother assuages her anxiety rather than resolves it. However, say that the mother realises that her belief that her son is not smoking pot must be false and thinks to herself that she should face the truth but that she still cannot bring herself to unbelieve it, which would make her belief transparently irrational. Quilty-Dunn and Mandelbaum's view struggles to explain this common situation, since the mother is trying to resolve the dissonance. On my view, the mother may judge 'My son

smokes pot' and fail to revise her belief 'My son does not smoke pot' if she fails to purge her belief of emotional cues. This does not seem at all unlikely, considering the subjective importance of the issue in question.

5. Concluding Remarks

In this paper, I defended the view that some cases of delusion are transparent – these sufferers recognize their thoughts as false beliefs, they want to revise their beliefs, but are unable to do so. The explanation is that, in transparent delusions, some reasoning processes are operational and these processes are the reason why the decontextualized thought-revision was initiated. However, they were not used in the decontextualized thought-revision, the process was initiated but it was incomplete, which explains why the thought was 'passed' as correct rather than revised. Because these reasoning processes are working reliably, the person is capable of correctly understanding reality (e.g., 'I cannot be Satan') but, since they are not used in the thought-revision, their outputs cannot be utilised so that the thought gets revised. Due to this incomplete thought-revision, the sufferer's judgement that ~*p* is rendered ineffective with respect to his delusion that *p*.

A satisfactory theory of delusion needs to recognize this kind of thought-evaluation failure. Transparent delusion is important not only because it gives another facet to delusion but also because it provides evidence against the popular but deeply mistaken view that judging that *p* entails believing that *p*. The insight that judging that *p* does not entail believing that *p* or unbelieving that ~*p*, in turn, makes many cases, not just those involving delusional patients, less problematic. In fact, in a closely related phenomenon, many non-pathological cases that appear as involving people who believe contradictions actually involve people judging the opposite of what they believe without abandoning their, what I call, transparently irrational belief. What explains this particular phenomenon is an error in an earlier stage of thought-revision: the believer was unable to purge successfully the belief of contextual or emotional cues, which, then, interfere with the revision of the thought.

REFERENCES

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

Amador, X.F., M. Flaum, N.C. Andreasen, D.H. Strauss, S.A. Yale, S.C. Clark, and J.M. Gorman. (1994). 'Awareness of Illness in Schizophrenia and Schizoaffective and Mood Disorders.' *Archives of General Psychiatry* 51, 826–836.

Aimola-Davies, A., and M. Davies. (2009). 'Explaining Pathologies of Belief.' In Broome M. and Bortolotti L. (Eds.), *Psychiatry as Cognitive Neuroscience: Philosophical Perspectives*, 285–323. Oxford: Oxford University Press.

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

Chatterjee, A. and M. Mennemeier. (1996). 'Anosognosia for hemiplegia: Patient retrospections.' *Cognitive Neuropsychiatry* 1, 221–237.

Coltheart, M. (2007). 'The 33rd Bartlett Lecture: Cognitive neuropsychiatry and delusional belief.' *Quarterly Journal of Experimental Psychology*, 60A, 1041–1062.

Coltheart, M., Menzies, P., and Sutton, J. (2010). 'Abductive inference and delusional belief.' *Cognitive Neuropsychiatry* 15, 261–287.

Cuffel B.J, J. Alford, E.P. Fischer, R.R. Owen. (1996). 'Awareness of illness in schizophrenia and outpatient treatment adherence.' *The Journal of nervous and mental disease* 184, 653–659.

Currie, G. (2000). 'Imagination, delusion and hallucinations.' In Coltheart M. and Davies M. (Eds.), *Pathologies of Belief*, 167–182. Blackwell.

Currie, G., and J. Jureidini. (2001). 'Delusion, rationality, empathy.' *Philosophy, Psychiatry & Psychology* 8, 159–162.

Currie, G., and I. Ravenscroft. (2002). *Recreative Minds*. Oxford: Oxford University Press.

Davidson, D. (2004/1982). 'Paradoxes of Irrationality.' In his *Problems of Rationality*, 169–188. Oxford: Clarendon Press.

———. (1985). 'Incoherence and Irrationality.' *Dialectica* 39, 345–354.

Davies, M., and A. Egan. (2013). 'Delusion, Cognitive Approaches: Bayesian Inference and Compartmentalisation.' In K.W.M. Fulford et al. (Eds.), *The Oxford Handbook of Philosophy and Psychiatry*, 689–727. Oxford: Oxford University Press.

Egan, A. (2008). 'Seeing and Believing: Perception, Belief Formation and the Divided Mind.' *Philosophical Studies* 140, 47–63.

———. (2009). 'Imagination, delusion, and self-deception.' In Bayne T. and Fernández J. (Eds.), *Delusion and Self-Deception: Motivational and Affective Influences on Belief-Formation*, 263–280. New York: Psychology Press.

Evans, G. (1982). *The Varieties of Reference*. Oxford: Clarendon Press.

Gerrans, P. (2014). *The Measure of Madness: Philosophy of Mind, Cognitive Neuroscience, and Delusional Thought*. Cambridge US: The MIT Press.

Hohwy, J. (2013). 'Delusions, Illusions and Inference under Uncertainty.' *Mind & Language* 28, 57–71.

Hohwy, J. and V. R. (2012). 'Delusions as Forensically Disturbing Perceptual Inferences.' *Neuroethics* 5, 5–11.

Lackey, J. (2007). 'Norms of Assertion.' *Noûs* 41, 594–626.

Landa, Y., S.M. Silverstein. F. Schwartz, and A. Savitz. (2006). 'Group Cognitive Behavioral Therapy for Delusions: Helping Patients Improve Reality Testing.' *Journal of Contemporary Psychotherapy* 36, 9–17.

Langdon, R. and P.B. Ward. (2009). 'Taking the Perspective of the Other Contributes to Awareness of Illness in Schizophrenia.' *Schizophrenia Bulletin* 35, 1003–1011.

Maher, B.A. (1999). 'Anomalous experience in everyday life: Its significance for psychopathology.' *The Monist* 82, 547–570.

McKay, R. (2012). 'Delusional Inference.' *Mind & Language* 27, 330–355.

Parrott, M. and P. Koralus. (2015). 'The erotetic theory of delusional thinking.' *Cognitive Neuropsychiatry* 20, 398–415.

Parrott, M. (2016). "Bayesian Models, Delusional Beliefs, and Epistemic Possibilities." *The British Journal for the Philosophy of Science* 67, 271–296.

Peacocke, C. (2000). 'Conscious Attitudes, Attention, and Self-Knowledge.' In C. Wright, B. C. Smith, and C. Macdonald (Eds.), *Knowing Our Own Minds*, 63–98. Oxford: Oxford University Press.

Quilty-Dunn, J. and E. Mandelbaum. (2018). 'Against dispositionalism: belief in cognitive science.' *Philosophical Studies* 175: 2353–2372.

Schwitzgebel, E. (2002). 'A phenomenal, dispositional account of belief.' *Noûs* 36, 249–275.

———. (2012). 'Mad Belief?' *Neuroethics* 5, 13–17.

Shah, N. (2003). 'How Truth Governs Belief.' *The Philosophical Review* 112, 447–482.

Sullivan-Bissett, E. and P. Noordhof. (2019). 'The transparent failure of norms to keep up standards of belief.' *Philosophical Studies* (First Online): 1–15.

Tumulty, M. (2012). 'Delusions and Not-Quite-Beliefs.' *Neuroethics* 5, 29–37.

Turner, M., and M. Coltheart. (2010). 'Confabulation and Delusion: A Common Monitoring Framework.' *Cognitive Neuropsychiatry* 15, 346–376.

Zislin, J., V. Kuperman, and R. Durst. (2011). '"Ego-Dystonic" Delusions as a Predictor of Dangerous Behavior.' *Psychiatric Quarterly* 82, 113–120.