# Delusional Evidence-Responsiveness*

Carolina Flores

January 2021

*This is a post-peer-review, pre-copyedit version of an article published in Synthese. The final authenticated version is available online at: https://doi.org/10.1007/s11229-021-03070-2.*

**Abstract**

Delusions are deeply evidence-resistant. Patients with delusions are unmoved by evidence that is in direct conflict with the delusion, often responding to such evidence by offering obvious, and strange, confabulations. As a consequence, the standard view is that delusions are not evidence-responsive. This claim has been used as a key argumentative wedge in debates on the nature of delusions. Some have taken delusions to be beliefs and argued that this implies that belief is not constitutively evidence-responsive. Others hold fixed the evidence-responsiveness of belief and take this to show that delusions cannot be beliefs. Against this common assumption, I appeal to a large range of empirical evidence to argue that delusions are evidence-responsive in the sense that subjects have the capacity to respond to evidence on their delusion in rationally permissible ways. The extreme evidence-resistance of delusions is a consequence of powerful masking factors on these capacities, such as strange perceptual experiences, motivational factors, and cognitive biases. This view makes room for holding both that belief is constitutively evidence-responsive and that delusions are beliefs, and it has important implications for the study and treatment of delusions.

## 1 Introduction

Delusions are deeply evidence-resistant. Consider, for example, Esmé Weijun Wang's first-personal description of Cotard delusion:

> "In the beginning of my own experience with Cotard's delusion, I woke my husband before sunup…"I'm dead," I said, "and you're dead, and Daphne [the dog] is dead, but now I get to do it over. Don't you see? I have a second chance. I can do better now." C. said, gently, "I think

*Acknowledgements: Thanks to Elisabeth Camp, Andy Egan, August Gorman, Sofia Jeppsson, Hannah Read, Susanna Schellenberg, Ernest Sosa, Christopher Willard-Kyle, and three anonymous referees for helpful comments on previous versions of this paper. Thanks to Anita Avramides for first spurring my interest in this topic. Thanks also to participants in the Rutgers Epistemology Seminar and the Rutgers Junior Seminar for helpful discussion.

> you're alive." But this statement, of course, meant nothing. It was his opinion, and I had my solid belief. I can state that the sky is green, but will you see it as such? (Wang 2019, 148)[1]

Deep evidence-resistance is close to being definitional of delusions. The Diagnostic and Statistics Manual of the American Psychiatric Association defines delusions as "not amenable to change in light of conflicting evidence" (Association 2013). Patients with delusions are unmoved by evidence that is in direct conflict with the delusion, often responding to such evidence by offering obvious, and strange, confabulations.

As a consequence, the standard view is that delusions are intractable, that is, not evidence-responsive. This claim has been used as a key argumentative wedge in debates on the nature of delusions. Some have taken delusions to be beliefs and argued that this implies that belief is not constitutively evidence-responsive. Others hold fixed the evidence-responsiveness of belief and take this to show that delusions cannot be beliefs (§2).

Against this shared assumption, I appeal to a large range of empirical evidence to argue that delusions are evidence-responsive in the sense that subjects have the capacity to adjust their take on the content of the delusion in rationally permissible ways in response to relevant evidence (§3). This makes room for holding both that belief is constitutively evidence-responsive and that delusions are beliefs, and it has important implications for the study and treatment of delusions and of people with delusions (§4.1).

## 2 The Intractability Assumption

When a person has a delusion, they seem to lose their grip on reality. They make claims such as "I am dead" (the Cotard delusion (Cotard 1880, Young and Leafhead 1996)), "My partner has been replaced by an impostor" (the Capgras delusion (Capgras and Reboul-Lachaux 1994, Pandis et al. 2019)), "Someone else's thoughts are being inserted into my mind" (the thought insertion delusion (Mullins and Spence 2003)). Even when delusions have fairly mundane content—"My partner is cheating on me," (Othello syndrome (Todd and Dewhurst 1955)), "People have it out for me" (persecutory delusion (Freeman 2007)) or "That person is in love with me" (erotomania (Jordan and Howe 1980))—they share the same feature of appearing disconnected from the available evidence.

More specifically, delusions exhibit extreme evidence-resistance. As the DSM 5 notes, delusions "are not amenable to change in light of conflicting evidence" (Association 2013).[2] Patients maintain their delusions despite being surrounded by strong

1. Wang mostly discusses her experiences with the Capgras and Cotard delusion, which are fairly unusual. These delusions arose, in her case, in the context of schizoaffective disorder, which is a diagnosis often accompanied by a wide range of (common and unusual) delusions.
2. The DSM's definition of delusion has been subject to vigorous criticism, both for failing to distinguish delusions from non-delusional beliefs and for wrongly excluding some delusions from the category (Bortolotti 2018, Coltheart 2007). But few contest the claim that deep evidence-resistance is at least present in the vast majority of delusions.

counter-evidence to them and tend to hold on to their delusions with tenacity, rejecting, dismissing, or explaining away what looks like decisive counter-evidence.

The ways in which people with delusions interact with counter-evidence to their delusions can look deeply puzzling. For example, consider the patient who claimed that his hand belonged to his doctor and who answered, "Ever see a man with three hands?" with: "A hand is the extremity of an arm. Since you have three arms it follows that you must have three hands."(Bisiach 1988, 469). Or consider the patient who held on to her delusion that an acquaintance was in love with her even after he told her on the phone that, not only was he not in love with her, but in fact could barely remember who she was (Jordan and Howe 1980). Patients might receive counter-evidence as one would take up "a bedtime story" (Wang 2019, 158). They may claim to be an exception to all known facts: since one is dead, feeling one's heartbeat and other physical sensations is not evidence that one is alive, in just this one special case (Young and Leafhead 1996, 158). They may recognize that what they are claiming is "unbelievable," and have the sense that "something is not quite right" when they make delusional claims while holding on to them nonetheless (Alexander et al. 1979, 335).

The evidence-resistance of delusions is so extreme, and so puzzling, that it has led many to doubt whether delusions are beliefs. It is hard to make sense of attitudes that are so deeply evidence-resistant as aiming at the truth, or as hooked to a shared reality that they seek to represent, as beliefs are standardly taken to be. As Andy Egan puts it, "If we think that a certain sort of evidence responsiveness is essential to belief, then, in many cases, we'll be reluctant to say that delusional subjects genuinely believe the contents of their delusions" (Egan 2008, 266). We can summarize this argument as follows:

**The Anti-Doxasticism Argument**

1. The Evidence-Responsiveness View of Belief: Belief is constitutively evidence-responsive.
2. Intractability: Delusions are not evidence-responsive.[3]
3. Anti-Doxasticism: Therefore, delusions are not beliefs.[4]

The Evidence-Responsiveness View of Belief is the orthodox view.[5] It helps make good on the epistemic role of belief as an attitude that aims at truth (Williams 1970, Velleman 2000) and is constitutively subject to epistemic standards (Burge 2010). If delusions do not meet the evidence-responsiveness benchmark, the anti-doxasticist holds, they cannot be beliefs, and therefore have some other non-doxastic status. Perhaps patients' utterances are empty speech acts (Berrios 1991). Or maybe delusions are attitudes of some different type: imaginings (Currie and Ravenscroft 2002), acceptances (Frankish 2012, Dub 2017), hybrids between belief and desire or belief and imagination (Egan 2008), or attitudes towards mental states (Stephens and Graham 2004).

3. I owe the term "intractable" to Reimer (2010).
4. Bortolotti and Miyazono (2015) helpfully center this argument in discussing the philosophical literature on belief.
5. See Helton (forthcoming) for illuminating discussion.

But one person's *modus ponens* is another's *modus tollens*. If one accepts the standard, DSM-endorsed view that delusions are beliefs, their evidence-resistance *prima facie* puts pressure on the idea that beliefs are constitutively evidence-responsive. From this perspective, we get the following argument:

**The Anti-Responsiveness Argument**

1. **Doxasticism**: Delusions are beliefs.
2. **Intractability**: Delusions are not evidence-responsive.
3. **Anti-Responsiveness**: Therefore, belief is not constitutively evidence-responsive.

In a series of influential works, Lisa Bortolotti has forcefully pushed this argument (Bortolotti 2005a, Bortolotti 2005b, Bortolotti 2009). Other defenders of the doxastic conception of delusions (Bayne and Pacherie 2005, Reimer 2010) also "deny that there is a constitutive connection between belief and evidence" (Bayne and Pacherie 2005, 183).

Doxasticists offer compelling reasons to hold that delusions are beliefs, in addition to the fact that delusions are both intuitively and in clinical practice classified as beliefs.[6]

Delusions appear to be continuous with ordinary believing, which makes it unnatural to think that they are an entirely different kind of attitude. Run-of-the-mill non-delusional beliefs (for example, political and ideological beliefs, and beliefs that play an important role in justifying one's habits or practices) can also be deeply evidence-resistant. And there are plenty of borderline cases between stubborn, but non-delusional, beliefs on the one hand and delusions on the other. For example, how should we classify the belief that Hillary Clinton and prominent democrats run a pedophile ring (LaFrance 2020), among other conspiracist beliefs? Such beliefs seem as deeply disconnected from shared reality as many clinical delusions, yet its proponents are not held to be delusional in the clinical sense. The doxastic view elegantly accounts for the continuity between delusions and ordinary beliefs.

Additionally, doxastic views have explanatory power: they account for the fact that we can often interpret subjects' delusion-related behavior by ascribing beliefs. This is most salient in cases where delusions lead to deeply disconcerting behavior, such as stalking someone the patient believes to be in love with them, or attacking a partner who they believe to be an impostor (as 18% of the 260 Capgras patients in Foerstl et al. (1991)'s review did). But belief ascriptions also have explanatory power when the delusion does not lead to extreme content-congruent behavior: they explain assertions of the content of their delusion, distress and difficulties with coping, and confabulations in response to evidence. The large majority of patients with delusions who are admitted to a psychiatric institution behave in some ways that ascribing a

6. Note that one can accept doxasticism without holding that a full theory of delusions will be limited to studying delusional beliefs. As phenomenological approaches note, a full understanding of delusions involves understanding the subjective experience of delusions, which in turn may require studying more pervasive changes in the patient's experience of, and perspective on, the world. See Bovet and Parnas (1993), Sass (1994), Sass et al. (2011), Sass and Pienkos (2013) for more on such approaches.

belief in the delusion helps explain (Wessely et al. 1993), supporting the claim that delusions are beliefs.

The debate appears to be at a standstill. At play are two different conceptions of belief and delusions. Doxasticists sever traditionally-held connections between belief and evidence-responsiveness so as to hold the attractive view that delusions are beliefs. Anti-Doxasticists reject that view so as to protect the orthodox connection between belief and evidence-responsiveness. [7]

Despite these deep disagreements, both sides agree on one thing: the intractability assumption, that is, the claim that delusions are *not* evidence-responsive. And both the Anti-Responsiveness and Anti-Doxasticism arguments rely crucially on this claim. As we have seen, this claim is *prima facie* very plausible: it explains delusion maintenance in the face of counter-evidence and the strangeness of patients' reactions to such evidence. Plausible as it may be, I will argue that we ought to reject it. Though delusions are evidence-resistant, they are not intractable.

## 3 Delusions Are Evidence-Responsive

In this section, I will argue against the view that delusions are intractable. Specifically, I will argue that delusions are evidence-responsive in the sense that subjects have the capacity to rationally respond to evidence bearing on the content of their delusion. At the same time, delusions are not *exclusively* regulated in response to evidence; motivational and affective factors (among others) are also involved. This explains why delusions are evidence-resistant, i.e. often maintained in the face of seemingly decisive counter-evidence.[8]

To make the case for this view, I will start by outlining some key facts about mental capacities in general, and capacities to rationally respond to evidence in particular (§3.1). This will yield criteria for the possession of such capacities and put me in a position to argue, by appealing to empirical evidence on delusions, that people with delusions have the relevant capacities (§3.2). I will then explain why, though people with delusions have such capacities, they so often fail to use them properly, i.e. why they fail to abandon the delusion in response to counter-evidence (§3.3).

Before starting, an important clarification is in order. I want to argue that all (or at least the vast majority of) delusions are evidence-responsive. But *prima facie* delusions form a highly heterogeneous class. As we have seen, their content can vary widely, from bizarre claims ("I am dead") to completely ordinary ones ("My partner

7. There are parallel debates about other candidate constitutive features of belief, such as inferential integration, action guidance, and reason-giving. On one side, doxasticists use delusions to argue that these are not constitutive features of belief; on the other side, defenders of traditional conceptions of belief appeal to these features to argue that delusions are not beliefs. Discussing these debates is beyond the scope of this paper. See Bortolotti and Miyazono (2015) and Bortolotti (2018) for overviews.

8. Gerrans (2001) argues for the related view that delusions are performance failures of the subject's capacity for pragmatic rationality, i.e. the ability to apply rules of rationality in context. In his view, subjects with delusions have a capacity for pragmatic rationality, and just fail to apply it in a wide range of circumstances. Though this is similar to my view, we are concerned with different capacities: the capacity to respond to evidence, in my case, and the capacity to apply rules of reasoning in context, in his. Further, I am not equating delusions with failures to exercise such capacities, as Gerrans does. Instead, I explain why delusions are evidence-resistant in terms of masking factors on the subject's capacities.

is cheating on me"). Delusions also differ in their effects on action, affect, and the subject's overall vision of the world. While some delusions lead subjects to act out, sometimes in extreme ways (consider the Capgras patient who, taking his father to have been replaced by a robot, decapitated him to find the batteries (Blount 1986, Silva et al. 1989)), many are more behaviorally circumscribed, with the subject continuing their life more-or-less as before despite their delusional claims. At the inferential level, some delusions have a reduced effect on the subject's other beliefs, while others are elaborated into florid delusional theories (Davies and Coltheart 2000). Finally, delusions are highly heterogeneous in their accompanying psychiatric diagnosis: they appear in cases of schizophrenia (Coltheart 2007), localized brain damage (Young et al. 1992, Ellis and Lewis 2001), or dementia (Flynn et al. 1991), among other diagnoses, or even no psychiatric diagnosis (Freeman 2006). And this matters to features of the delusion. Delusions in the context of schizophrenia and related disorders tend to be polythematic and elaborated, that is, they tend to cover multiple interrelated themes and have more noticeable effects on inference, affect, and action, whereas delusions arising after localized brain damage tend to be monothematic and circumscribed, that is, focused on a single theme and comparatively insulated from the rest of the patient's cognition, affect, and action (Davies and Coltheart 2000).

Given such heterogeneity, one may doubt whether delusions are all in the province of the same cognitive mechanisms. If they are not, then one cannot (for example) generalize from studies about delusions in schizophrenia to claims about delusions in the context of localized brain damage, or perhaps even from claims about Capgras delusion to claims about Cotard delusion.

I am not in a position to decisively refute the possibility that different delusions are in the province of radically different cognitive mechanisms, or to consider the whole range of variation within the category of delusions. That said, we should not overstate the threat of heterogeneity. There are compelling models of delusions that propose that they are in the province of the same range of cognitive mechanisms (Appelbaum et al. 1999, Bell et al. 2008) while allowing for variation in the weights assigned to different factors. Indeed, the framework I will articulate for explaining delusion maintenance in the face of counter-evidence is of this sort.

To vindicate the claim that all or the vast majority of delusions are evidence-responsive, I will present data supporting the evidence-responsiveness of delusions with a range of different contents, degrees, and kinds of circumscription, and in the context of different diagnoses. This suffices to make an inductive case for the claim that all (or at least the vast majority of) delusions are evidence-responsive, though, of course, such a case would be bolstered by further research on a wide range of delusions.[9]

## 3.1 The capacity to respond to evidence

I will argue that delusions are evidence-responsive in the following sense:

> **Evidence-Responsiveness**: *S*'s attitude towards *p* is *evidence-responsive* just in case *S* has the capacity to rationally respond to evidence bearing

9. Thanks to an anonymous referee for pressing me to clarify the scope of my claim.

on $p$.

In this sub-section, I will put forward key facts about the nature of capacities to rationally respond to evidence. This will put me in a position to explore whether people with delusions have such capacities with respect to their delusions.

Rationally responding to evidence consists in changing one's attitude in epistemically permissible ways when one receives evidence: for example, reducing one's degree of belief or abandoning a belief when one receives counter-evidence or increasing one's degree of belief when one receives supporting evidence (to an epistemically permissible extent). I will here remain neutral on what the epistemically permissible responses to evidence are, a subject of considerable disagreement in epistemology.

In the case of many delusions, there is a broad consensus that many patients' responses are *not* epistemically permissible: patients often fail to respond to counter-evidence to their delusions in epistemically permissible ways. For example, there is wide agreement that one ought to abandon, or at least become less certain in, the claim that one's partner was replaced by an impostor when given evidence that the impostor looks exactly like their partner and can remember many shared life experiences, which Capgras patients often fail to do. It is on the basis of the claim that many of patients' responses to counter-evidence on their delusion are not epistemically permissible that we claim that delusions are evidence-resistant.

I will grant the standard view of what the epistemically permissible responses are in these cases, that is, I will grant that patients often respond in ways that are not epistemically permissible. This is a concessive move to proponents of the intractability view. If one were to claim that patients' interactions with seeming counter-evidence are epistemically permissible, it would follow trivially that delusions are evidence-responsive. In fact, they would count as evidence-responsive in the strong sense that patients *actually appropriately respond* to relevant evidence when they have such evidence.

For the purposes of this paper, I will adopt a factive conception of evidence (Williamson 2002). On this conception, one's non-factive mental states (i.e. mere appearances) do not count as evidence. In other words, I will show that subjects with delusions have the capacity to respond to factive counter-evidence on their delusions. This is also a concessive move to the intractability view. Patients with delusions receive strange perceptual input, which plays a role in causally explaining their delusions (as I will discuss in §3.3). It is therefore widely agreed (though rarely emphasized) that patients with delusions have the capacity to respond to non-factive evidence. Showing that they have the capacity to respond to factive evidence bearing on their delusions is more challenging. It is also more relevant to the debate about whether delusions are beliefs: the capacity to respond to non-factive evidence does not suffice to hook the delusion to shared reality, and therefore is not especially relevant to whether delusions are beliefs.

So far, I have elucidated what rationally responding to evidence bearing on $p$ involves. I will now focus on the most important element of this conception of evidence-responsiveness: the appeal to *capacities*.

I will start with some general facts about what having a capacity involves.[10] Hav-

10. My discussion here draws heavily on Schellenberg (2018)'s discussion of capacities.

ing the capacity to Φ does not imply that one Φs whenever one engages in the relevant activity, or whenever one tries to Φ. For example, having the capacity to run 10k in under 40 minutes does not imply that one always runs at that pace, and having the capacity to score a goal does not mean that every shot at the goal goes in. Indeed, as these examples illustrate, having a capacity does not even require one to succeed reliably, i.e. most of the time that one exercises that capacity.

Instead, having the capacity to Φ involves successfully Φ-ing in specific conditions that suit that capacity. For example, what matters to whether one has the capacity to run a 40-minute 10k is whether one does so when exerting serious effort, not injured, well-rested, highly motivated, and so on—even if one would fail to do so if a single one of these conditions is not met. Having the capacity to Φ is a matter of satisfying counterfactuals of the form "If special conditions $C$ were in place, then the subject would successfully Φ".

Applying these points to the capacity to rationally respond to evidence $e$, we can conclude that having that capacity does not require always responding to $e$ when one has that evidence. Instead, having the capacity to respond to $e$ involves satisfying the following counterfactual: if one were to receive evidence $e$ in some set of special conditions, one would respond to $e$ in an epistemically permissible way.

What are those special conditions? We are here dealing with a mental capacity: one which operates on mental states, and whose successful exercise consists in yielding a new set of mental states. As such, one would expect that factors internal to one's cognitive system could interfere with the successful exercise of such a capacity. For this reason, the special conditions in which one will rationally respond to evidence require the right sort of internal environment, one without internal tampering factors. Salient candidates for such tampering include exhaustion, cognitive biases, motivational factors such as the desire to hang on to a belief that is central to one's identity, and affective factors that cloud one's judgment. The special conditions under which someone who has the capacity to rationally respond to evidence $e$ would actually do so are ones where such interfering mental factors are absent. Importantly, these conditions may routinely fail to be met in the actual world. The subject may be too tired or overwhelmed most of the time, or their motivational structure may make them hold extremely tightly to some of their beliefs.

The central upshot is the following: an attitude can both be evidence-responsive in the sense that the subject has the capacity to rationally respond to evidence bearing on its content, and evidence-resistant in the sense that the subject does not change their mind most of the time they receive counter-evidence. That will be the case whenever, for a given individual and attitude, the right conditions for the exercise of the capacity rarely occur in the actual world.

I will argue that this is precisely what we find with delusions: patients with delusions have the capacity to rationally respond to counter-evidence to their delusion but are rarely in the right (internal) conditions to respond to it. To do so, I will, in §3.2, appeal to empirical evidence to argue that subjects with delusions would appropriately change their mind in response to evidence bearing on the delusion in a range of conditions, and, in §3.3, explain—by appeal to factors that interfere with the exercise of such capacities—why they fail to rationally respond to counter-evidence in many real-world circumstances.

## 3.2 In support of the presence of evidence-responsiveness capacities in delusions

In this sub-section, I will give reasons in favor of thinking that subjects have the capacity to rationally respond to evidence bearing on the content of their delusion. My focus will be on showing that that have the capacity to respond to *counter*-evidence. Their capacity to respond to supporting evidence is not commonly questioned.

### 3.2.1 Characteristic interactions with counter-evidence

As discussed in §2, much of the case for the intractability of delusions is based on patients' puzzling responses to evidence. I will now argue that these puzzling responses do not indicate the lack of the capacity to rationally respond to such evidence; in fact, they offer moderate support to the claim that patients have that capacity.

First, note that patients do not *dismiss* the evidence, but *incorporate* it, that is, make efforts to inferentially integrate the evidence with their delusion and background beliefs (Garety et al. 2001). This is often occluded in the philosophical literature on the topic, which highlights the fact that patients hold on to their delusion with tenacity while omitting or minimizing the fact that patients make efforts to accommodate the evidence.

Efforts to accommodate evidence can take different forms. Patients may state that they are an exception to generalizations, bite the bullet on implausible conclusions, or contrive stories that explain away the evidence. This case is illustrative:

> We asked her during the period in which she claimed to be dead whether she could feel her heart beat, whether she could feel hot or cold and whether she could feel whether her bladder was full. J.K. said that since she had such feelings even though she was dead they clearly did not represent evidence that she was alive. She said she recognised this was a difficult concept for us to grasp and one which was equally difficult for her to explain, partly because the experience was unique to her and partly because she could not fully understand it herself. We then asked J.K. whether she thought we would be able to feel our hearts beat, to feel hunger and so on if we were dead. J.K. said that we wouldn't and that this experience was unique to her. (Young and Leafhead 1996, 158)[11]

This patient receives what looks like decisive evidence that she is not dead: her heart is beating, she can feel hot or cold, she can feel whether her bladder is full. She does not change her mind in response. But she does engage with the evidence. She recognizes that, in normal cases, this would entail that a person is dead. But, from her point of view, she *knows* that she is dead; so death must look different from usual in her case.

11. This is, of course, just one case. I do not here present data on what fraction of patients with delusions inferentially integrate the evidence in relevantly similar ways. The role of presenting case studies is to make vivid why certain kinds of behavior are indicative of the capacity to respond to counter-evidence. The pervasiveness of those kinds of behavior is a different matter. Thanks to an anonymous referee for pressing me to articulate the role of case studies in the argument.

This pattern of reasoning is recognizable. If you observed, or were told about, something that looks like a violation of the laws of nature but were *certain* of what you saw or heard, you might find yourself reasoning in an analogous way. You know what you saw (or that someone else saw it), so the laws of nature must have been violated in that instance. Arguably, belief in miracles is often supported roughly in these ways. More generally, the ways in which patients integrate counter-evidence through confabulations are similar to those of ordinary believers when they receive counter-evidence to cherished beliefs.

Such responses are rationally impermissible (or, at least, I will grant that they are). The fact that patients integrate counter-evidence with their beliefs in rationally impermissible ways does not imply that they have the capacity to integrate such evidence in rationally permissible ways. However, the similarities between these interactions with counter-evidence and others we witness in non-delusional cases are suggestive. To the extent that a capacity to respond to counter-evidence is present in the latter cases, it is plausible to think it is present in cases of delusion too.[12]

Further, patients often understand how the evidence they receive bears on the content of their delusions, that is, what the epistemically permissible responses would be. Here is a striking example:

> E: Isn't that [two families] unusual?
>
> S: It was unbelievable.
>
> E: How do you account for it?
>
> S: I don't know. I've tried to understand it myself and it was virtually impossible.
>
> S: What if I told you I don't believe it?
>
> E: 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. (Alexander et al. 1979)

This interaction shows that *S* understands how the evidence bears on their delusions, i.e. what the rationally permissible responses would be. If, as Ryle noted, "execution and understanding are merely different exercises of knowledge of the tricks of the same trade" (Ryle 1949, 55), then the same capacity is employed to judge responses to evidence and to produce such responses. Assuming such a connection between understanding and execution, the fact that the subject appropriately judges responses to evidence indicates that they have the capacity to respond appropriately.

A third characteristic aspect of how subjects with delusions interact with counter-evidence consists in evidence-avoidance, motivated by strong discomfort provoked

12. Thanks to an anonymous referee for pressing me to clarify this point.

by counter-evidence.[13] As Esme Weijun Wang writes about her stint with the Cotard delusion:

> Being dead butted up against the so-called evidence of being alive, and so I grew to avoid that evidence because proof was not a comfort; instead, it pointed to my insanity. (Wang 2019, 157)

This behavior looks remarkably similar to the evidence-avoidance we find in run-of-the-mill confirmation bias, i.e. the ordinary tendency to seek out evidence friendly to one's current beliefs, and to avoid evidence against those beliefs (Klayman and Ha 1987, Klayman 1995, and Nickerson 1998). For example, how different is it from avoiding AIDS testing when one fears one might test positive (Dawson et al. 2006, Lerman et al. 2002)? In both cases, the subject avoids gathering evidence where the beliefs they would end up with, if they revised in accordance with that evidence, would be unpleasant ones (namely, beliefs that they have a serious illness).

One upshot is that delusion maintenance in the face of *available* evidence is sometimes a result of the patient failing to gather that evidence, and not of the patient failing to rationally respond to evidence they have. In other words, once we take into account evidence-avoidance, fewer instances of evidence-resistance—i.e. of failing to rationally respond to evidence one has—are left to explain.

Further, this kind of evidence-avoidant behavior is a sign of the capacity to respond to evidence. The reason is the following. Wanting to avoid revising beliefs one would end up revising if one acquired certain evidence is a good reason to avoid gathering it. This is sometimes explicitly recognized from a first-person perspective. For example, in his account of living with (and overcoming) grandiose and paranoid delusions, Robert Chapman writes that he "was afraid to check with reality for fear that [his] ideas might be deflated and [his] sense of having a useful and meaningful direction in pursuing these might be demolished." (Chapman 2002, 547). In contrast, if one were not even able to respond to the evidence, gathering it would make no difference to what one believes. Evidence-avoidance would be puzzling for intractable attitudes. [14]

13. Though it is hard to get statistics on just how common evidence-avoidance is in delusion patients, empirically well-supported models (e.g.Freeman et al. (2001)'s model, which focuses on persecutory delusions) ascribe evidence-avoidance a significant role in delusion maintenance. This suggests that it is a pervasive feature of delusions.

14. Note two complications. First, there is an alternative explanation that is compatible with full evidence-insensitivity: perhaps subjects avoid gathering evidence because they don't want to receive evidence of their own evidence-insensitivity (thanks to Andy Egan for suggesting this alternative explanation). But this explanation is inferior to the one I propose. It imputes to subjects beliefs about the degree of evidence-sensitivity of their own beliefs. And it does not match the phenomenology of confirmation bias, which is one of wanting to avoid being forced into a view one dislikes, or first-personal descriptions like the ones above. Second, this does not establish that patients have the capacity to respond to evidence to a rationally permissible extent; perhaps they have the capacity to respond by adjusting to *some* extent in the right direction, but not to a rationally permissible extent (thanks to an anonymous referee for pointing this out). However, it does indicate that counter-evidence can have some substantial effects on patients, which sits poorly with the intractability assumption. I will shortly discuss some reasons why patients with delusions may fail to respond to evidence to a sufficient extent, while still adjusting in the right direction.

### 3.2.2 The bias against disconfirming evidence (BADE)

A problem for the view that delusions are evidence-responsive is that people with delusions display a bias against counter-evidence (once they have gathered it). In this subsection, I will consider studies on this phenomenon and argue that this bias against counter-evidence is compatible with the capacity to rationally respond to counter-evidence.

In a series of studies, Todd Woodward and collaborators have shown that people with schizophrenia (Moritz and Woodward 2006, Woodward et al. 2008) and with schizotypal traits (Woodward et al. 2007), who are more likely to experience delusions, display a *bias against disconfirming evidence* (BADE): they adjust their beliefs to counter-evidence much less than other subjects.[15] This bias is restricted to strongly held beliefs (Woodward et al. 2008), but domain-general, that is, not restricted to delusional topics. The authors of these studies hypothesize that a BADE is a contributing factor to the maintenance of delusions (Woodward et al. 2006).

These findings may seem hard to square with the capacity to rationally respond to such evidence. Indeed, the authors of these studies describe people with a bias against disconfirming evidence as "less *able* to revise false convictions in general" (Moritz and Woodward 2006, 158) and "generally *impaired* in their ability to integrate disconfirmatory evidence" (Woodward et al. 2008, 268) (my italics).

In my view, these are inaccurate descriptions of the implications of these findings. A bias against disconfirming evidence does not imply lacking the capacity to respond to such evidence, as I will now argue.

First, given the domain-generality of the bias, and the fact that it is found in the non-clinical population, if a bias against disconfirming evidence implies lacking the capacity to respond to such evidence, lacking capacities to respond to counter-evidence would be widespread. People with schizophrenia and schizotypal traits, most of whom do not have delusions, would lack the capacity to respond to counter-evidence to all of their strongly held beliefs. This is a very strong, and *prima facie* implausible, claim.

Further,other facts about how these subjects interact with evidence are hard to square with lacking the capacity to respond to counter-evidence. They do as well as other subjects when it comes to responding to supporting evidence. They adjust their beliefs to some extent—just an insufficient one—in light of disconfirming evidence in the studies. And, in other contexts, they show a greater tendency to over-adjust to disconfirmatory evidence (Moritz and Woodward 2005), so that a bias against disconfirming evidence is only manifest in some contexts. These facts sit awkwardly with the idea that they lack the capacity to rationally respond to counter-evidence. Instead, they suggest that these subjects have the capacity to respond to (both supporting and disconfirming) evidence on their beliefs, where that capacity is masked in *some* instances in which they receive counter-evidence to strongly held beliefs.

In fact, other work by the same team suggests a plausible explanation for the bias against disconfirming evidence that does not impute lacking the capacity to respond to counter-evidence. All of us—subject to this bias or not—are motivated to find ways of accommodating counter-evidence to our strongly held beliefs without abandoning them (Gilbert 2006, Cooper 2007, Harmon-Jones and Harmon-Jones 2007). Now, one

15. Thanks for an anonymous referee for bringing these studies to my attention.

would expect interpersonal and contextual variation in how many ways of accommodating counter-evidence one comes up with, i.e. in how many alternative explanations for that evidence (other than the falsity of one's cherished beliefs) one constructs. The more alternative ways of accommodating counter-evidence one can come up with, the less that evidence will seem to discredit one's strongly held beliefs.

We have reason to think that the very same subjects who display a bias against disconfirming evidence come up with more ways of accommodating counter-evidence. This is because they display a *liberal acceptance bias*: they tend to give high plausibility ratings to a wide range of alternative views, including views that common sense would immediately dismiss (Moritz and Woodward 2004, Moritz and Woodward 2005).[16] [17]

Let's put these two points together. Like all of us, people who display a bias against disconfirming evidence are motivated to generate alternative explanations for that evidence. Due to their liberal acceptance bias, they will accept more of those as plausible than other members of the population. This leads them to adjust their beliefs to counter-evidence less, i.e. to display a bias against disconfirming evidence. From their perspective, the counter-evidence does not look decisive because it can be explained away in a wide range of ways.

If this is right, then we have a good explanation for the bias against disconfirming evidence that is compatible with people with this bias having the capacity to respond to counter-evidence to strongly held beliefs in rationally permissible ways. People with a BADE have the capacity to rationally respond to counter-evidence, but it is masked by (1) their motivation to hold on to those beliefs, which leads to generating more alternative explanations for the counter-evidence, and (2) a liberal acceptance bias, which makes them fail to rule out some of those explanations which common sense would exclude.

More strongly, one may think that, taking into account the wider range of alternative explanations they consider, people with a BADE respond rationally to counter-evidence. Specifically, if one thinks that how confident one should be in a belief in the face of relevant evidence depends on the space of alternative explanations for that evidence of which one is aware, then displaying a BADE may be rational given the wider range of alternative explanations considered.

### 3.2.3 Delusion remission and cognitive behavioral therapy

So far, we have been looking at how the behavior of people with delusions when they do *not* change their mind in *prima facie* epistemically permissible ways in response to counter-evidence contains markers of the capacity to respond to evidence. But

16. One might worry that this just pushes intractability one step back, to the intractability of these alternative views: why do they accept views that common sense would immediately dismiss? But liberal acceptance does not require failing to respond to counter-evidence to these implausible alternatives. More plausibly, patients simply fail to *gather* such evidence, in that they may fail to retrieve it from memory and are likely not to receive such counter-evidence from their environment at the moment of acceptance.
17. These findings cohere with phenomenological descriptions of delusion in schizophrenia, which highlight disturbances in background or bedrock certainties Rhodes and Gipps (2008), "a change in the totality of understandable connections"(Jaspers 1963, 97), or "a mutation of the ontological framework of experience itself"(Sass and Pienkos 2013, 633) as factors in the formation and maintenance of delusions. One would expect such disturbances to be reflected in accepting options that common sense dismisses.

patients sometimes abandon their delusions in response to counter-evidence or come to hold them less strongly. In this section, I will discuss such instances and consider their significance for whether delusions are evidence-responsive.

First, in some cases, the strength with which a delusion is held wanes in response to counter-evidence, with the subject eventually abandoning the delusion:

> LU was asked whether she had ever seen a dead person before, and if so how she had known that the person was dead. LU responded that after her grandmother's death she had viewed her grandmother, and that she knew her grandmother was dead because her eyes were closed and she was motionless. LU acknowledged that the fact that she herself was moving and talking was inconsistent with the typical characteristics of dead people, and she subsequently expressed some uncertainty about her beliefs. Within a week of the initial neuropsychological assessment, her delusion appeared to have completely resolved. (McKay and Cipolotti 2007, 353)

Giving LU evidence against the claim that she is dead leads her to, first, reduce the degree to which she felt certain of that claim, and then to abandon it altogether. This delusion is evidence-responsive: LU *actually responds* to counter-evidence to the delusion, and therefore trivially has the capacity to do so.

The fact that patients' confidence in their delusions varies also indicates that patients have and exercise capacities to rationally respond to counter-evidence. Conviction can vary even over the course of one day, with a predictor for reduced confidence being interactions with other people (Myin-Germeys et al. 2001). This may be because others provide counter-evidence to the delusion, suggesting that people with delusions are at least sometimes swayed by counter-evidence.[18][19]

Further, delusions are typically not chronic conditions: they tend to be eventually abandoned, at least when they do not arise in the context of neurodegenerative disease. Once the delusion is in remission, subjects exercise (and therefore have) the capacity to rationally respond to evidence. It is unlikely that they lose this capacity during the delusional period and then re-acquire it. In such cases, "it is reasonable to assume that the neuroanatomical basis of… competence is unimpaired" (Gerrans 2001, 166).

Notably, first-personal descriptions of coming to abandon a delusion often emphasize the effects of considering counter-evidence. Consider these two cases, the first one of Capgras delusion remission, and the second one of remission of a range of persecutory delusions:

> I've started going through it, and seeing what could possibly happen and what couldn't happen. That was wrong, that couldn't happen. Even though it has happened it couldn't. Mary couldn't suddenly disappear from the room, so there must be an explanation for it. The lady knows me way back. She could say things that happened 40 years go, and I

18. This study considers only patients with schizophrenia.

19. As seen in our discussion of the bias against disconfirming evidence, they are typically swayed to a lesser extent than people without schizophrenia. But, as discussed there, this is no objection to the claim that they have and exercise the capacity to rationally respond to evidence in such cases.

> wonder where she gets them from. … And then I worked it out and I've wondered if it's Mary all the time. It's nobody else. (Turner and Coltheart 2010, 371)

> When a delusion is stacked against a conscious awareness of reality and rationality, the delusion falls apart. While the delusive ideas disintegrate, their pretense is revealed. I put on my detective cap. I would test out arguments. I tried to develop the strongest arguments possible against the falsehoods. I made a list of all the rational alternatives that I could think of. I looked for evidence for what really was happening and what really wasn't happening. I asked myself, "How do I know this?" Did I actually see it or just a "sign" of it? Did I really hear it, or could I have misinterpreted what I heard? Did I smell, taste, or feel it? Did someone tell me this? Is most of my evidence beyond my senses or interpretations of signs and symbols? I tried to test reality in terms of the here and now transactions with other people rather than assuming what their supposed intentions were or predicting what would happen.(Chapman 2002, 551)

In cases of remission like these, patients respond to evidence (from their senses, testimony, and memory). Considering counter-evidence leads them to abandon their delusion. They have the capacity to respond to evidence, and successfully exercise it to abandon the delusion.

These first-personal descriptions are indicative of the fact that offering counter-evidence has the potential to make the patient abandon the delusion. This lies behind the effectiveness of cognitive-behavioral therapy (CBT), which provides a particularly compelling case for the view that delusions are evidence-responsive. Cognitive-behavioral therapy directly guides the patient toward collecting and assessing evidence bearing on their delusion. It centers on changing patients' beliefs about events they have experienced so as to change their responses to those events. More specifically, it focuses on *cognitive restructuring*, i.e., on examining and challenging maladaptive thought patterns and helping the patient establish more adaptive ones (Dozois and Dobson 2010, 11) and on *reality testing*, that is, assessing whether beliefs fit the evidence and replacing distorted beliefs with more realistic ones (Kendall and Bemis 1983). The second quote on delusion remission above is illustrative of the process (though, remarkably, the patient himself designed and conducted the process).

Though the philosophical significance of cognitive-behavioral therapy has been insufficiently explored, doxasticists (Bortolotti 2009 and Bayne and Pacherie 2005) have appealed to it in arguing for the claim that delusions are beliefs. They argue that our theories of delusion ought to be consistent with best clinical practice, in which "the therapist treats the delusional patient as a believer of $p$, and he or she gently invites the patient to question whether $p$ is the thing that ought to be believed" (Bayne and Pacherie 2005, 185). Consistency with such practice, they argue, requires classifying delusions as beliefs, as opposed to claiming that the therapist is mistaken in their ascription.

Note that such a practice also involves the therapist assuming that the patient has the capacity to respond to counter-evidence: the therapist assumes that the patient

can "question whether $p$ is the thing that ought to be believed"(Bayne and Pacherie 2005, 185) and abandon their belief that $p$ if it is not. By the same token, then, doxasticists ought to accept the evidence-responsiveness of delusions. This is significant because, as we saw in §2, they endorse the intractability assumption.

The effectiveness of cognitive-behavioral therapy can be marshaled in an additional argument for delusional evidence-responsiveness. Specifically, the best explanation for successful instances of cognitive behavioral therapy involves the target attitudes being evidence-responsive. Successful CBT requires the patient to correctly assess the significance of the evidence they receive, and adjust their delusion and related beliefs accordingly. Properly guided by the therapist, the patient *actually responds* to counter-evidence, which establishes that they have the capacity to do so at that moment. The best explanation for this fact is that they had that capacity before entering the therapeutic context, but were failing to exercise it properly.[20]

CBT is effective as a treatment for delusions. Recent meta-analyses concur that "targeted individualized CBT for delusions and hallucinations is effective" (Lincoln and Peters 2018; see also Gaag et al. (2014)). These studies cover both delusions in the context of schizophrenia and other clinical contexts.[21] They establish that therapeutic approaches with a "focus on cognitive reframing and reality testing" (Lincoln and Peters 2018, 67)—i.e., on getting patients to assess the content of the delusion based on evidence—are effective, exhibiting significant benefits on "subjective strength of conviction, reactions to and acting on beliefs" (Lincoln and Peters 2018, 67). In other words, patients have been found to respond to evidence against their delusions by abandoning or reducing their credence in their delusions.

We have seen that, for cognitive-behavioral therapy to work on an attitude, that attitude must be evidence-responsive, and that CBT works on many delusions. The upshot is that many delusions are evidence-responsive.[22]

## 3.3 Explaining delusional evidence-resistance

In the last sub-section, I presented a wide range of reasons to think that delusions are evidence-responsive. Still, patients hold on to their delusions with a tenacity that seems hard to understand. They do not employ their capacities to respond to evidence often or easily. This requires explanation: what is keeping them from successfully exercising these capacities? Three families of factors are promising candidates for explaining evidence-resistance in delusions. These are continued strange experiences reinforcing the delusion, motivational factors, and automatic reasoning biases.

20. The alternative is that they acquire this capacity in therapy, but lacked it beforehand. This alternative implies deep changes to the patient's cognitive architecture over the course of a small number of therapy sessions, which is implausible.

21. For reviews that focus specifically on schizophrenia, and establish that CBT is effective in treating delusions in schizophrenia, see Sarin et al. (2011) Wykes et al. (2008) Turkington et al. (2006).

22. This result should not be over-stated. CBT is *relatively* effective as a treatment for delusions, but this is in part because delusions are still poorly understood, and, as a result, we lack highly effective treatments. Indeed, most of the effect sizes found are in the small-to-medium range (Gaag et al. 2014), and there are many cases of delusions where this treatment does not work at all. That said, when it works, it is in part because the delusion is evidence-responsive. When it does not, this is likely due to the kinds of factors I will outline in §3.3.

In this section, I will show that these factors and their interaction with evidence-responsiveness capacities can explain a wide range of evidence-resistant behavior in patients with delusions.

Take strange experiences. Many delusions are formed in response to strange experiences. For example, the best explanation for Capgras delusion ("My partner has been replaced by an impostor") formation appeals to a deficit in the visual processing of faces that is the mirror image of prosopagnosia.[23] In prosopagnosia, people lose the ability to recognize faces but still experience the affective responses they would usually have in response to the faces of loved ones, as measured by, for example, skin conductance tests (Bruce and Young 1986). Conversely, Capgras patients' facial recognition systems are intact, but they lack the usual affective responses they would have upon seeing loved ones (Ellis and Young 1990, Ellis and Lewis 2001). This generates the sense that something is deeply wrong. It appears to the patient that something has changed in the person, and this needs accounting for. The view that they have been replaced by an impostor is a response to this abnormal experience.[24]

Though the strange experience does not suffice to explain delusion formation, it is widely agreed that it is a crucial causal factor.[25] Further, this is not a one-off experience. It is sustained by cortical damage that does not just go away. As a result, the patient's unsettling sense of unfamiliarity when looking at the face of a loved one is continually reinforced: whenever the Capgras patient looks at their loved one's face, they experience the same disturbing lack of affect. They *constantly* receive apparent evidence that supports their delusion. This comes into conflict with the actual evidence they get from talking with other people or recalling facts they know. Responding to that evidence in rationally permissible ways is difficult in the face of such strange private experiences. In other words, these strange experiences explain why exercising the capacity to respond to counter-evidence does not result in the abandonment of the delusion.[26]

To get a clear sense of the difficulty in abandoning a delusion in such circumstances, consider the following case of remission from anosognosia (the delusional

23. This explanation paradigmatically applies to cases of Capgras that occur without a schizophrenia diagnosis, but Coltheart et al. (2007) argue that it also extends to such delusions in the context of schizophrenia. Similar explanations in terms of localized brain damage causing perceptual distortions have also been explored for other monothematic delusions.

24. This brief explanation omits many points of disagreement about Capgras formation. For example, some accounts hold that the first step in delusion formation is not a conscious experience but the mere lack of an autonomic response (Coltheart 2005). And some hold that the patient *endorses* the experience of seeing their partner as an impostor (Bayne and Pacherie (2004)), instead of adopting it as an *explanation* for the experience that something is off in the interaction (Maher (1999), Stone and Young (1997)). Others still think that the perceptual effects we find in Capgras are the result of a top-down disturbance (Campbell (2001): the experience is the result of the delusion, and not the other way around. Which of these views is correct is an interesting question, but it does not affect the point that delusion formation involves a strange experience, which is the key point for my purposes.

25. Most theorists agree that there are other factors involved in the formation of the delusion, such as a deficit in hypothesis evaluation (Coltheart 2007), reasoning biases such as the tendency to jump to conclusions (Garety and Freeman 1999), a liberal acceptance bias (Moritz and Woodward 2004, Moritz and Woodward 2005), or a bias toward privilege explanatory adequacy (i.e. privileging how well the experience is explained by the hypothesis over how probable the hypothesis antecedently is (McKay 2012)).

26. See Corlett et al. (2009) for an account of delusion maintenance that emphasizes the role of such reinforcement.

failure to acknowledge illness or impairment (Davies and Davies 2009)):

> 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'. (Chatterjee 1996, 227)

The patient believes that his left leg is paralyzed: he knows he had a stroke that left him paralyzed on the left side of the body. That said, he still experiences an inclination towards the delusional claim that he can move around as before the stroke. This inclination seems to partly result from strange, illusory perceptual experiences of their paralyzed limb moving as intended, where these experiences result from the impairment of sensation, attention, and motor control (Davies and Davies 2009). It takes continued suppression and monitoring efforts for the patient not to be taken in by this powerful feeling. Given this fact, it is not surprising that patients sometimes temporarily become less certain of their delusion when given evidence against it, but then return to strongly believing it.

Following the philosophical literature on delusions, I have so far focused on the role of strange experiences in monothematic delusions that result from localized brain damage. But the same point applies to delusions characteristic of schizophrenia. They also appear to be partly formed and sustained by abnormal perceptual experiences. Due to dopamine deregulation, patients with schizophrenia attribute salience inappropriately to internal and external stimuli (Kapur 2003). This results in a strange experience of the world writ large (as the phenomenological tradition has long recognized), one characterized by a sense of tension and hidden meaning (Fuchs 2005, Sass and Pienkos 2013). Again, this sense of strangeness and significance is continually reinforced in each and every perceptual experience, coming into conflict with evidence received via testimony or memory.

Strange perceptual experiences, then, are an important factor in explaining why people with delusions maintain them in the face of counter-evidence, despite having the capacity to respond to such evidence. That said, not all people with delusions have such strange experiences (Bell et al. 2008). Such experiences cannot be the only factor masking evidence-responsiveness capacities.

Motivational factors (such as the desire to hang on to a pleasant delusion or a compelling explanation for one's strange experiences) are another such factor. The role of motivational factors in delusion has long been recognized. Indeed, traditional psychodynamic theories of delusion formation held that these were formed exclusively due to motivational factors. Such theories claim, for example, that delusions of persecution are part of a defense mechanism to avoid attributing negative events to oneself, and that Capgras delusion arises to make sexual desire for an inappropriate target (e.g. a parent) acceptable.

These views do not hold up to scrutiny, given evidence of brain damage, anomalous experiences, and reasoning biases and deficits in delusions. Nonetheless, it remains very plausible that motivational factors play some role in the maintenance of delusions, helping explain why subjects' evidence-responsiveness capacities do not lead to revision in the face of counter-evidence.

This is especially plausible when it comes to delusions with a content that is self-aggrandizing—such as delusions of grandeur, or erotomania—or simply more positive than evidence warrants—such as anosognosia or the reverse Othello syndrome (the delusion that one's partner continues to be faithful, despite overwhelming evidence to the contrary). Sophisticated motivational accounts of such delusions have been developed (e.g. Ramachandran (1996)'s account of anosognosia). Similarly, persecutory delusions seem to be maintained in part because attributing negative events to external causes allows the patient to avoid negative beliefs about the self (Bentall et al. 2001).

The basic idea here—explored in detail in accounts that see self-deception and delusion as overlapping—is that holding on to such views supports one's self-esteem or self-image, motivating the subject to maintain them.[27] This inclines the subject to process evidence in biased ways, confabulating (Turnbull et al. 2004) and more generally finding elaborate ways of fitting counter-evidence by adjusting some of their other beliefs.

The role of motivational factors is not restricted to positive delusions. As cognitive dissonance theory (Festinger et al. 1956, Devine 1994, Cooper 2007, Harmon-Jones and Harmon-Jones 2007) tells us, receiving counter-evidence to our beliefs (positive or not) has negative valence. We are motivated to alleviate cognitive dissonance by re-establishing coherence among our beliefs, now including the counter-evidence. How we do so depends on the centrality of the beliefs under attack: the more central or strongly held they are, the more likely we are to alleviate dissonance while maintaining those beliefs.

Two facts about delusions, regardless of whether their content is positive, should make us expect patients to be highly motivated to maintain them. First, delusions provide an explanation for an unusual experience, without which the patient would feel at sea in the world.[28] Second, in the case of delusions, admitting that one is wrong is admitting that something has gone seriously amiss with one's cognition, a conclusion that would likely ravage one's self-esteem. For these reasons, as seen in the discussion of the bias against disconfirming evidence above, we should expect patients to expend significant effort at generating alternative explanations for counter-evidence to their delusions. In this way, motivational factors would contribute to delusion maintenance by leading patients to come up with a wide range of ways of accommodating counter-evidence.

Finally, successfully exercising one's capacity to respond to evidence can be effortful, requiring one to overcome or suppress automatic reasoning biases. Where such biases are operative, one will fail to rationally respond to evidence, despite having

27. See the essays in Bayne and Fernández (2010) for discussion of the relationship between delusion and self-deception.

28. Freeman and Garety (2004) argue that this is an important factor in the maintenance of persecutory delusions.

the capacity to do so. Reasoning biases function as masks on the subject's cognitive capacities: they are "thinking distortions and processing preferences rather than performance deficits and limitations of mental capacity" (Moritz and Woodward 2007, 619). In other words: exhibiting reasoning biases does not imply that one lacks the capacity to reason rationally, only that exercising that capacity will require effort to overcome or suppress those biases.

As seen, people with delusions exhibit several reasoning biases (including attribution biases, the tendency to jump to conclusions, and a bias towards observational adequacy) to a greater extent than control subjects (Broome et al. 2007, Garety and Freeman 1999, McKay 2012). To respond to evidence, they will need to suppress these biases, which is cognitively effortful and requires motivation. Indeed, training subjects to identify such cognitive biases in their thinking, and to come up with strategies for avoiding them (metacognitive training), is an effective treatment for delusions (Moritz and Woodward 2007, Moritz et al. 2014). This indicates that such cognitive biases are a factor in delusion maintenance, functioning as masks on evidence-responsiveness capacities. And it indicates that these are not permanent masks, but can be removed with sufficient effort and motivation, allowing patients to successfully exercise their evidence-responsiveness capacities.

In sum, we have reason to think that subjects with delusions have the capacity to respond to counter-evidence on their delusion (§3.2). This capacity can be masked by continued strange experiences, motivational factors, and cognitive biases, in which case the delusion stays in place, or does not sufficiently change, in the face of counter-evidence (§3.3). However, when they employ that capacity in the absence of masking factors, the result is a rationally permissible change of mind in response to the evidence. This is what we see when subjects become less confident in the content of their delusion in the face of counter-evidence, when CBT succeeds, and in cases of evidence-responsive remission more generally.

# 4 Implications of the View that Delusions are Evidence-Responsive

## 4.1 Implications for the anti-responsiveness and anti-doxasticism arguments

The claim that delusions are evidence-responsive bears directly on the debate on the nature of delusions and belief outlined in §2. As we saw there, both proponents of the Anti-Responsiveness and Anti-Doxasticism arguments hold that delusions are intractable, that is, not evidence-responsive, and rely on this claim as a central argumentative hinge. The discussion in §3 shows that this view is false: delusions are evidence-responsive. We can therefore simultaneously hold the two following views:

> **The Evidence-Responsiveness View of Belief**: Belief is constitutively evidence-responsive, i.e. if $S$ believes that $p$, then $S$ has the capacity to rationally respond to evidence bearing on $p$.
>
> **Doxasticism**: Delusions are beliefs.

It follows that the Anti-Responsiveness Argument does not go through: delusions do not provide a counterexample to the claim that belief is constitutively evidence-responsive. Similarly, the Anti-Doxasticism Argument fails: the evidence-responsiveness of beliefs does not militate against the claim that delusions are beliefs. We can have our evidence-responsive cake and eat it too.

One might argue that this is too quick—for two reasons. The first has to do with the scope of the claim that delusions are evidence-responsive: to dissolve the debate I would need to show conclusively that all delusions are evidence-responsive, and I have not done so. The second has to do with the sense of evidence-responsiveness at stake: participants in the initial debate, this objector claims, meant evidence-responsiveness in a stronger sense. I will consider these in turn.

First, the scope objection. As I mentioned at the beginning of §3, I cannot decisively show that *all* delusions are evidence-responsive, but only make an inductive case for that claim by considering data on a wide range of delusions. However, given delusional heterogeneity (in etiology, accompanying diagnoses, degree of circumscription, and content, among others), the worry remains that the results I present do not generalize to all delusions. Now, suppose that I have established that many or most delusions are evidence-responsive, but that there remain some that are not. Why not think that the very same debate arises again for this smaller subset of intractable delusions, with the Anti-Responsiveness side arguing that they provide a counterexample to the view that belief is constitutively evidence-responsive and the Anti-Doxasticism side arguing that they are not beliefs? And, if so, how does my argument in §3 help with this debate?[29]

Against this, given that I present data supporting the evidence-responsiveness of delusions with a wide range of different contents, degrees, and kinds of circumscription, and accompanied by a range of different psychiatric diagnoses, I think that we can legitimately conclude (through an inductive argument) that all delusions are evidence-responsive. Delusional heterogeneity in actual evidence-responsiveness is the result of different masks on evidence-responsiveness capacities.

Of course, the claim that all delusions are evidence-responsive relies on an inductive argument, so there may be delusions to which these findings do not generalize. But the burden of proof is on the opponent to present such cases. In other words, it is not clear if there is space left for the same debate to arise over a smaller subset of delusions. There is, to be sure, space for debates about how to classify imaginary cases, and to use thought experiments to probe the limits of belief. But this is a different debate, one that runs parallel to debates on delusions. This paper is not meant to refute every possible counter-example to the evidence-responsiveness view of belief resulting from thought experiments.

But let's grant the objector that there are some intractable delusions. It is not clear that this would generate the same debate we saw. This is because few if any doxasticists hold that *all* delusions are beliefs: for instance, as Elisabeth Pacherie and Tim Bayne note, they "certainly do not intend to argue that all delusional states are beliefs" (Bayne and Pacherie 2005, 179). For this reason, it is not clear that doxasticists would feel compelled to claim that these marginal cases are beliefs and use them to

29. Thanks to Christopher Willard-Kyle for pressing me on this point.

argue against the evidence-responsiveness view of belief. Indeed, the existing debate in the literature focuses precisely on the kinds of cases I consider in §3, so it is fair to say that my discussion in that section advances the debate by showing that those are cases where we detect evidence-responsiveness (against the assumption made by both sides).

The second objection is that, in claiming that delusions are evidence-responsive in the capacities sense articulated in §3.1, I have just changed the subject. Participants in the initial debate meant something else by "evidence-responsiveness", namely, rationally responding to counter-evidence most of the time, and it is no resolution of the debate to change the topic.

In response, grant that many participants in the debate spell out evidence-responsiveness along such lines. Most explicitly, Lisa Bortolotti's arguments (in Bortolotti (2005a), Bortolotti (2005b), and Bortolotti (2009)) have as their target what she calls "the background view" of rationality constraints on belief, according to which belief requires an overall background of rationality. On the background view, "if deviations from norms of rationality are too widespread, then the ascription of beliefs is compromised" (Bortolotti 2005a, 20). Applying this view of rationality constraints to evidence-responsiveness, her target is the view that people rationally respond to counter-evidence on their beliefs most of the time.

I agree that we ought to reject this view. As Bortolotti (2009) notes, we do not even need to appeal to delusions to establish this point. We often fail to rationally revise many ordinary beliefs—including political beliefs, beliefs one cherishes, and prosaic beliefs about the goodness of our own everyday decisions—in the face of counter-evidence.[30]

However, this does not settle things in favor of the Anti-Responsiveness side. Proponents of the Anti-Responsiveness Argument take themselves to be attacking the most promising version of a connection between evidence-responsiveness and belief.[31] If I am right, then they are not attacking the most promising version of this idea. The version I propose is more promising, given that, unlike the ones they consider, it can accommodate data on delusion maintenance and provides a framework in which to theorize about the interaction between evidence-responsiveness, perceptual experience, motivational factors, and cognitive biases.

Similarly, proponents of the Anti-Doxasticism Argument aim to "protect the idea of essentially rational belief from attack;"(Currie and Jureidini 2001, 161), and not specifically the idea that beliefs adjust to the evidence in rational ways most of the time. The appeal to capacities for evidence-responsiveness offers a new tool to defend that general idea. The Evidence-Responsiveness View of Belief, spelled out in terms of capacities, makes belief essentially rational in the sense that it is part of the nature of belief that beliefs are underwritten by evidence-responsiveness capacities; in other words, that, in the absence of interfering factors, beliefs are rationally updated in response to counter-evidence. This is compatible with a range of powerful interfering factors to such updating, as we saw in §3.3. For this reason, as I argued in §3, belief being essentially rational in this sense is compatible with delusions being beliefs.

30. See Mandelbaum (2019) for compelling defense of this claim.
31. Indeed, sometimes they explicitly consider other versions of the idea. See Bortolotti 2009, 18–21.

The notion of evidence-responsiveness I propose therefore allows us to make compatible two attractive and otherwise jointly untenable positions: the traditional view that there is some connection between believing and rationally responding to evidence, and the claim that delusions are beliefs. This dissolves the debate we saw in §2, which hinges on the claim that delusions are not evidence-responsive. And it opens the door to holding that delusions are evidence-responsive beliefs.

As discussed in §2, the doxastic conception of delusions has many advantages. Centrally, the doxastic view does justice to intuitive and clinical classifications of delusions as beliefs, respects the continuity between non-delusional beliefs and delusions, and accounts for the value of ascribing delusional beliefs in explaining patients' behavior. These advantages are preserved in a view that claims that delusions are *evidence-responsive* beliefs.

Though I favor the view that delusions are evidence-responsive *beliefs*, this can be seen as an optional add-on to my central claim in this paper, namely, that delusions are evidence-responsive. You can accept that claim while rejecting a doxastic analysis of delusions. Delusions might fail to be beliefs because they fail to meet other necessary conditions on belief, such as characteristic connections to action, inference, and affect. This is more than a bare possibility: there are standing debates on whether delusions meet such necessary conditions on belief.[32]

Independently of how such debates turn out, and of whether we ultimately ought to classify delusions as beliefs, showing that the Evidence-Responsiveness View of Belief does not rule out classifying delusions as beliefs is significant. It dissolves a long-standing debate in the literature. And it shows that the Evidence-Responsiveness View of Belief need not be overly rationalist, or require us to idealize beliefs. Evidence-responsiveness is compatible with the pervasive interference of non-rational factors, to an extent that can lead to delusion.

## 4.2 Implications for the study of delusion and belief maintenance

In §3, I argued that delusions are underwritten by evidence-responsiveness capacities that can be masked in a range of ways, including by abnormal perceptual experiences, motivational factors, and cognitive biases. This involved surveying and piecing together a wide range of empirical literature on delusions. One can view the result as an an integrated framework on which to think about delusion maintenance. Instead of, for example, seeing views that focus on how delusions are responsive to evidence (Maher 1974) as competitors to views that center motivational influences (like traditional psychodynamic views), we can factor in contributions of these different factors into a single model.

This paper provides only a framework, not a full model that can make concrete predictions about the maintenance of specific delusions. As our empirical understand-

32. My argument in this paper suggests a novel argumentative strategy for dealing with these debates. Instead of asking whether subjects with delusions act, infer, and feel in belief-characteristic ways most of the time, we should investigate whether subjects with delusions have the relevant capacities (e.g. for acting on a belief, drawing inferences from it, and experiencing corresponding emotional reactions). Focusing on capacities is a promising strategy for accommodating pervasive irrationality on the one hand, and theoretical and practical roles that belief is called upon to play on the other. Exploring this possibility is beyond the scope of this paper.

ing of the contribution of different factors to delusion maintenance grows, we might be able to construct a detailed model, one which assigns weights to these different factors for specific delusions, helping us understand the causal mechanisms behind delusion maintenance and yielding recommendations for interventions that are likely to lead to delusion remission in specific cases.

The framework I propose may also help us understand *belief* maintenance, not just delusion maintenance. If one embraces the orthodox view that belief is constitutively evidence-responsive, then beliefs are in the province of evidence-responsiveness capacities—as are delusions. Further, the masking factors to which I appealed to explain delusional evidence-resistance (non-veridical perception, motivational factors, and cognitive biases) also apply to beliefs. This suggests that beliefs and delusions are regulated by at least some of the same cognitive mechanisms.

Consequently, if my view of delusions is right, we can look to delusions to better understand beliefs, and vice-versa. This fits with the methodology of cognitive neuropsychiatry, which uses data from "people with acquired disorders of cognition to constrain, develop, and test theories of normal cognitive structures and processes" (Davies and Davies 2009, 288), and models of normal cognition to investigate the causes of such disorders.[33]

This approach has been successfully applied to the study of delusion formation. It underlies the account of Capgras formation discussed in §3.3, which relies on comparing Capgras patients' responses to faces to a model of normal face recognition (Ellis and Young 1990). Building on such work, two-factor theorists of delusions (who think that there is an additional factor on top of perceptual abnormalities involved in delusion formation and maintenance) have approached delusion formation and maintenance through comparisons with our best models of such processes in subjects without delusions (Stone and Young 1997, Davies and Coltheart 2000, Davies and Davies 2009, Coltheart 2007, Coltheart et al. 2011). The framework that emerges from seeing delusions as evidence-responsive can be seen as a contribution to this research project.

## 4.3 Implications for treatment

The view that delusions are evidence-responsive in the capacities sense has implications for treatment. Specifically, it encourages us to take seriously the view that offering counter-evidence to the delusion—when conditions for getting the patient to exercise their evidence-responsiveness capacities are in place—is a promising path to delusion remission. In practical terms, the advice is to invest more in cognitive-behavioral therapy coupled with interventions that remove masking factors.

This fits current treatment practices. As seen in §3.2, cognitive-behavioral therapy is recommended for delusions. Further, within cognitive-behavioral therapy, at-

33. There is substantial unclarity about what normality consists in: is it the statistically normal case (here, of belief maintenance)? The case where the systems involved in belief maintenance meet their function, whatever that may be? The case that satisfies norms of rationality? This is an important unresolved methodological question in cognitive neuropsychiatry. I am here just assuming that there is some thin (non-moral) notion of normality that allows us to study the functioning and malfunctioning of cognitive mechanisms. Thanks to August Gorman for discussion.

tention is given to ensuring that background conditions for the patient to respond to evidence are in place. For instance, cognitive-behavioral therapy includes building a relationship of trust between the patient and therapist, without which it is unlikely to be effective (Greene 2005). Trust is a condition for the patient to accept the evidence the therapist provides, and it reduces the likelihood of defensive reactions to it. In other words, it increases the likelihood that the patient will exercise their evidence-responsiveness capacities without the interference of motivational factors that manifest themselves in defensive reactions.[34]

There is growing recognition of the role of motivational factors in mediating the effectiveness of CBT. Specifically, when patients respond defensively to their delusions being directly challenged—a clear sign of the influence of motivational factors via cognitive dissonance—CBT is unlikely to succeed. In such cases, meta-cognitive therapy (discussed in §3.3) is recommended. Meta-cognitive therapy takes a 'back-door' approach, not directly addressing the delusion but instead enabling the patient to do so at a later stage (Moritz and Woodward 2007), in a way that avoids motivational factors taking up such a large role in responding to the evidence. In addition to setting up the stage for interacting with evidence in a way that puts less weight on motivational factors, meta-cognitive therapy trains patients to avoid common reasoning biases. As such, it can be understood as removing masks of a non-motivational kind, namely, reasoning biases.

Finally, antipsychotic medication can be understood as operating on perceptual masks on the patient's evidence-responsiveness capacities. It primarily operates on the patient's dopaminergetic neurotransmission (Gardner et al. 2005). As seen in §3.3, abnormal perceptual experiences caused by dopamine deregulation, which affects perceptual salience, are one of the factors that explain delusion maintenance in the face of counter-evidence. Antipsychotic medication operates on this masking factor through regulating dopamine transmission.

In addition to helping account for current treatment practices, the view that delusions are evidence-responsive helps de-stigmatize delusions, which is important for successful treatment. Delusions are not bizarre attitudes outside the space of reasons, or which place patients into a realm apart from people without psychosis. Instead, they are the result of (heightened forms of) cognitive, motivational, and perceptual factors that also cause and maintain ordinary beliefs. Normalizing and reducing stigma increases treatment motivation in patients (Lüllmann and Lincoln 2013). My view suggests that de-stigmatizing delusions is not merely practically helpful in treatment, but is also accurate to the cognitive mechanisms behind delusions.

## 4.4 Ethical implications

As the point about stigma foreshadows, the view that delusions are evidence-responsive beliefs has important ethical upshots. Delusions are often treated as grounds on which to exclude people from the moral and epistemic community. The view that delusions are evidence-responsive offers tools to argue that such exclusion is inappropriate, and not just for moral reasons.

34. For more on the role of disturbances in trust and communication in the maintenance of delusions, see Fuchs (2015) and Fuchs (2020).

The justification for excluding people with delusions from the moral and epistemic community goes something like this. People with delusions are unmoored from shared reality, beyond the reach of reasons or evidence. At least insofar as the delusion is concerned, there is no point attempting to rationally engage with them. Further, the strangeness and tenacity of many delusions easily make it easy to slip into thinking that there is something so deeply wrong that attempting to rationally engage on *any* topic is unwarranted. As a result, a kind of blanket objectifying treatment comes to cover all interactions with the patient.

In this view, a person with delusions is properly treated in *the objective stance*, as an object to be handled as opposed to a person to reason with. Indeed, people with delusions have been taken to be *paradigmatic* candidates for the objective stance. As Strawson noted in introducing the concept of the objective stance, seeing an agent "as one whose picture of the world is an insane delusion…tends to inhibit ordinary interpersonal attitudes in general, and the kind of demand and expectation which those attitudes involve; and tends to promote instead the purely objective view of the agent as one posing problems simply of intellectual understanding, management, and control"(Strawson 1974, 16–17).

Now, one may think the objective stance is appropriate given the kind of attitude delusions are, yet oppose taking it up based on ethical considerations. For example, one may object to this treatment because it is disrespectful, or because marginalizing and stigmatizing people with delusions makes their lives worse and reduces the chances of delusion remission, while holding that people with delusions are outside the space of reasons, at least insofar as the delusion is concerned. This position makes adopting a non-objective stance toward people with delusions a useful fiction.

In contrast, my view of delusions implies that it is taking up the objective stance that is grounded in a (pernicious) fiction. It is not that it is morally better to act as if people with delusions are within the reach of reasons. They genuinely are within the reach of reasons. In other words, my view reconstitutes how it is appropriate to look at people with delusions, and opens up space for attitudes "of involvement or participation in a human relationship"(Strawson 1974, 9).

There is substantive debate about what these attitudes look like in the case of severe mental illness. The traditional Strawsonian view upholds a dichotomy between the objective stance and the participant stance, where the participant stance includes blame and resentment (in addition to attitudes such as "gratitude, forgiveness, anger, or the sort of love which two adults can sometimes be said to feel reciprocally, for each other"(Strawson 1974, 9)). However, perhaps blame and resentment are inappropriate in the case of delusions. Philosophers of psychiatry have challenged the claim that it is appropriate to blame patients while rejecting the view that we ought to shift into the objective stance. This can be achieved in two main ways: either by severing the participant stance and the appropriateness of blame (Pickard and Ward 2013, Pickard 2015) or by introducing new non-objective stances, such as "the nurturing stance" (Brandenburg 2018).[35]

Outlining precisely what attitudes are appropriate in the case of delusions is beyond the scope of this paper. The key point for my purposes is that the objective

35. Thanks to Sofia Jeppsson for bringing the nurturing stance to my attention.

stance is not appropriate given the nature of the delusion, against the standard view outlined above. In particular, at least some attitudes and ways of interacting that fall sharply outside the objective stance, such as reasoning with someone and holding them to normative standards for how they ought to respond, are appropriate.

The fact that there is space for such attitudes does not on its own imply that we ought, all things considered, to take them up in every interaction. This is a familiar point. Taking up the objective stance is "a resource" we use "as a refuge…from the strains of involvement; or as an aid to policy; or simply out of intellectual curiosity"(Strawson 1974, 9–10). This is true in interacting with anyone, not just with people with delusions. In the latter case, the strains of involvement may prove themselves more burdensome, so that taking up the objective stance when interacting on the topic of the delusion may still often be the better option.

As the point above illustrates, the question of what stance to take up when interacting with people with delusions in specific circumstances is a difficult one. Addressing it would require taking into account many factors beyond the evidence-responsiveness of the delusion. Nevertheless, establishing that there is room for taking up a non-objective stance when interacting on the topic of the delusion is significant, especially given the marginalization, exclusion, and stigmatization that comes with the objective stance.[36]

# 5 Conclusion

Delusions have been taken to pose a hard challenge for the orthodox view that belief is constitutively evidence-responsive. They are typically classified as beliefs, yet they are deeply evidence-resistant, seeming to constitute a counterexample to the orthodox view. In the opposite direction, proponents of the orthodox view have argued that the deep evidence-resistance of delusions implies that they cannot be beliefs.

Against this, I appealed to empirical evidence on delusions and their remission to argue that delusions are evidence-responsive in the sense that people with delusions have the capacity to rationally respond to evidence bearing on their delusions. Their extreme evidence-resistance is a consequence of masking factors on these capacities such as strange perceptual experiences, motivational factors, and cognitive biases.

Once we see that delusions are evidence-responsive, we can dissolve the apparent tension between the view that delusions are beliefs and the evidence-responsiveness view of belief. Further, the claim that delusions are evidence-responsive yields a unified framework on which to study delusions and belief updating, and has significant implications both for treatment and for how to interact with patients with delusions.

36. Thanks to August Gorman for helpful discussion of the ethical implications of the view.

## Works Cited


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

Paul S Appelbaum, Pamela Clark Robbins, and Loren H Roth. Dimensional approach to delusions: comparison across types and diagnoses. *American Journal of Psychiatry*, 156(12):1938–1943, 1999.

American Psychiatric Association. *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub, 2013.

Tim Bayne and Jordi Fernández. *Delusion and self-deception: Affective and motivational influences on belief formation*. Psychology Press, 2010.

Tim Bayne and Elisabeth Pacherie. Bottom-up or top-down: Campbell's rationalist account of monothematic delusions. *Philosophy, Psychiatry, & Psychology*, 11(1): 1–11, 2004.

Timothy J. Bayne and Elisabeth Pacherie. In defence of the doxastic conception of delusions. *Mind and Language*, 20(2):163–88, 2005. doi: 10.1111/j.0268-1064.2005.00281.x.

Vaughan Bell, Peter W Halligan, and Hadyn D Ellis. Are anomalous perceptual experiences necessary for delusions? *The Journal of nervous and mental disease*, 196(1): 3–8, 2008.

Richard P Bentall, Rhiannon Corcoran, Robert Howard, Nigel Blackwood, and Peter Kinderman. Persecutory delusions: a review and theoretical integration. *Clinical psychology review*, 21(8):1143–1192, 2001.

G. Berrios. Delusions as 'wrong beliefs': A conceptual history. *British Journal of Psychiatry*, 159:6–13, 1991.

Eduardo Bisiach. Language without thought. In L. Weiskrantz, editor, *Thought without Language*, pages 464–84. Oxford University Press, 1988.

G. Blount. Dangerousness of patients with capgras syndrome. *Nebraska Medical Journal*, 71(207), 1986.

Lisa Bortolotti. Delusions and the background of rationality. *Mind and Language*, 20 (2):189–208, 2005a. doi: 10.1111/j.0268-1064.2005.00282.x.

Lisa Bortolotti. Intentionality without rationality. *Proceedings of the Aristotelian Society*, 105(3):385–392, 2005b. doi: 10.1111/j.0066-7373.2004.00119.x.

Lisa Bortolotti. *Delusions and Other Irrational Beliefs*. Oxford University Press, 2009.

Lisa Bortolotti. Delusion. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2018 edition, 2018.

Lisa Bortolotti and Kengo Miyazono. Recent work on the nature and development of delusions. *Philosophy Compass*, 10(9):636–645, 2015. doi: 10.1111/phc3.12249.

Pierre Bovet and Josef Parnas. Schizophrenic delusions: a phenomenological approach. *Schizophrenia bulletin*, 19(3):579–597, 1993.

Daphne Brandenburg. The nurturing stance: Making sense of responsibility without blame. *Pacific Philosophical Quarterly*, 99:5–22, 2018.

MR Broome, LC Johns, I Valli, JB Woolley, P Tabraham, C Brett, L Valmaggia, E Peters, PA Garety, and PK McGuire. Delusion formation and reasoning biases in those at clinical high risk for psychosis. *The British Journal of Psychiatry*, 191(S51):s38–s42, 2007.

Vicki Bruce and Andy Young. Understanding face recognition. *British journal of psychology*, 77(3):305–327, 1986.

Tyler Burge. *Origins of objectivity*. Oxford University Press, 2010.

John Campbell. Rationality, meaning, and the analysis of delusion. *Philosophy, Psychiatry, & Psychology*, 8(2):89–100, 2001.

J Capgras and J Reboul-Lachaux. L'illusion des' sosies' dans un délire systématisé chronique. *History of Psychiatry*, 5(17):119–133, 1994.

Robert K Chapman. First person account: Eliminating delusions. *Schizophrenia Bulletin*, 28(3):545–553, 2002.

Anjan Chatterjee. Anosognosia for hemiplegia: Patient retrospections. *Cognitive Neuropsychiatry*, 1(3):221–237, 1996.

Max Coltheart. Conscious experience and delusional belief. *Philosophy, Psychiatry, & Psychology*, 12(2):153–157, 2005.

Max Coltheart. Cognitive neuropsychiatry and delusional belief. *Quarterly journal of experimental psychology (2006)*, 60(8):1041–1062, 2007.

Max Coltheart, Robyn Langdon, and Ryan McKay. Schizophrenia and monothematic delusions. *Schizophrenia Bulletin*, 33(3):642–647, 2007.

Max Coltheart, Robyn Langdon, and Ryan McKay. Delusional belief. *Annual review of psychology*, 62:271–298, 2011.

Joel Cooper. *Cognitive dissonance: 50 years of a classic theory*. Sage, 2007.

Philip R Corlett, John H Krystal, Jane R Taylor, and Paul C Fletcher. Why do delusions persist? *Frontiers in human neuroscience*, 3:12, 2009.

J. Cotard. Du délire hypochondriaque dans une forme grave de la méleancolie anxieuse. *Annales medico-psychologiques*, 4:168–174, 1880.

Gregory Currie and Jon Jureidini. Delusion, rationality, empathy: Commentary on martin davies et al. *Philosophy, Psychiatry, & Psychology*, 8:159–162, 01 2001. doi: 10.1353/ppp.2001.0006.

Gregory Currie and Ian Ravenscroft. *Recreative Minds: Imagination in Philosophy and Psychology*. Oxford University Press, 2002.

Anne M. Aimola Davies and Martin Davies. Explaining pathologies of belief. In Matthew Broome and Lisa Bortolotti, editors, *Psychiatry as Cognitive Neuroscience: Philosophical Perspectives*. Oxford University Press, 2009.

Martin Davies and Max Coltheart. Pathologies of belief. In Martin Davies and Max Coltheart, editors, *Pathologies of belief*. Wiley-Blackwell, 2000.

E. Dawson, K. Savitsky, and D. Dunning. "don't tell me, i don't want to know": Understanding people's reluctance to obtain medical diagnostic information. *Journal of Applied Social Psychology*, 36:751–768, 2006.

A.J. Elliot & P.G. Devine. On the motivational nature of cognitive dissonance: Dissonance as psychological discomfort. *Journal of Personality and Social Psychology*, 67 (3):384–394, 1994. doi: 10.1037/0022-3514.67.3.382.

David Dozois and Keith Dobson, editors. *Handbook of Cognitive Behavioral Therapies*. Guilford Press, 02 2010.

Richard Dub. Delusions, acceptances, and cognitive feelings. *Philosophy and Phenomenological Research*, 94(1):27–60, 2017. doi: 10.1111/phpr.12220.

Andy Egan. Imagination, delusion, and self-deception. In Tim Bayne and Jordi Fernandez, editors, *Delusion and Self-Deception: Affective and Motivational Influences on Belief Formation (Macquarie Monographs in Cognitive Science)*. Psychology Press, 2008.

Hadyn D Ellis and Michael B Lewis. Capgras delusion: a window on face recognition. *Trends in cognitive sciences*, 5(4):149–156, 2001.

Hadyn D Ellis and Andrew W Young. Accounting for delusional misidentifications. *The British Journal of Psychiatry*, 157(2):239–248, 1990.

L. Festinger, H. W. Riecken, and S. Schachter. *When Prophecy Fails*. University of Minnesota Press, 1956.

Frederick G Flynn, Jeffrey L Cummings, and Jeffrey Gornbein. Delusions in dementia syndromes: investigation of behavioral and neuropsychological correlates. *The Journal of neuropsychiatry and clinical neurosciences*, 1991.

Hans Foerstl, Osvaldo P Almeida, Adrian M. Owen, Alistair Burns, and Robin Howard. Psychiatric, neurological and medical aspects of misidentification syndromes: a review of 260 cases. *Psychological medicine*, 21 4:905–910, 1991.

Keith Frankish. Delusions, levels of belief, and non-doxastic acceptances. *Neuroethics*, 5(1):23–27, 2012. doi: 10.1007/s12152-011-9123-7.

Daniel Freeman. Delusions in the nonclinical population. *Current psychiatry reports*, 8(3):191–204, 2006.

Daniel Freeman. Suspicious minds: the psychology of persecutory delusions. *Clinical psychology review*, 27(4):425–457, 2007.

Daniel Freeman and Philippa A Garety. *Paranoia: The psychology of persecutory delusions*. Psychology Press, 2004.

Daniel Freeman, PA Garety, and E Kuipers. Persecutory delusions: developing the understanding of belief maintenance and emotional distress. *Psychological medicine*, 31(7):1293, 2001.

Thomas Fuchs. Delusional mood and delusional perception–a phenomenological analysis. *Psychopathology*, 38(3):133–139, 2005.

Thomas Fuchs. The intersubjectivity of delusions. *World Psychiatry*, 14(2):178, 2015.

Thomas Fuchs. Delusion, reality, and intersubjectivity: a phenomenological and enactive analysis. *Philosophy, Psychiatry, & Psychology*, 27(1):61–79, 2020.

Mark Gaag, Lucia Valmaggia, and Filip Smit. The effects of individually tailored formulation-based cognitive behavioural therapy in auditory hallucinations and delusions: A meta-analysis. *Schizophrenia Research*, pages 30–37, 04 2014. doi: 10.1016/j.schres.2014.03.016.

David M Gardner, Ross J Baldessarini, and Paul Waraich. Modern antipsychotic drugs: a critical overview. *Cmaj*, 172(13):1703–1711, 2005.

Philippa A. Garety and Daniel Freeman. Cognitive approaches to delusions: A critical review of theories and evidence. *British Journal of Clinical Psychology*, 38(2):113–154, 1999. doi: 10.1348/014466599162700.

Phillip A Garety, Elizabeth Kuipers, David Fowler, Daniel Freeman, and PE Bebbington. A cognitive model of the positive symptoms of psychosis. *Psychol Med*, 31(2): 189–195, 2001.

Philip Gerrans. Delusions as performance failures. *Cognitive Neuropsychiatry*, 6(3), 2001.

D. Gilbert. *Stumbling on happiness*. Knopf, 2006.

Judy A Greene. Cognitive-behavioral therapy. *Psychiatric Services*, 56(9):1161–1162, 2005.

Eddie Harmon-Jones and Cindy Harmon-Jones. Cognitive dissonance theory after 50 years of development. *Zeitschrift für Sozialpsychologie*, 38(1):7–16, 2007.

Grace Helton. If you can't change what you believe, you don't believe it. *Noûs*, forthcoming. doi: 10.1111/nous.12265.

Karl Jaspers. *General Psychopathology*. University of Chicago Press, 1963.

Harold W. Jordan and Gray Howe. De clerambault syndrome (erotomania): A review and case presentation. *Journal of the National Medical Association*, 72(10), 1980.

Shitij Kapur. Psychosis as a state of aberrant salience: a framework linking biology, phenomenology, and pharmacology in schizophrenia. *American journal of Psychiatry*, 160(1):13–23, 2003.

P.C. Kendall and K.M. Bemis. Thought and action in psychotherapy: The cognitive-behavioral approaches. In *The clinical psychology handbook*, pages 562–592. Pergamon, 1983.

Joshua Klayman. Varieties of confirmation bias. In *Psychology of Learning and Motivation*, volume 32, pages 385–418, 1995. doi: 10.1016/S0079-7421(08)60315-1.

Joshua Klayman and Young-won Ha. Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, 94(2):211–228, 1987. doi: 10.1037/0033-295X.94.2.211.

Adrienne LaFrance. Q-anon is more important than you think. *The Atlantic*, June 2020. URL https://www.theatlantic.com/magazine/archive/2020/06/qanon-nothing-can-stop-what-is-coming/610567/.

C. Lerman, R.T. Croyle, K.P. Tercyak, and H. Hamman. Genetic testing: Psychological aspects and implications. *Journal of Consulting and Clinical Psychology*, 70:784–797, 2002.

Tania Lincoln and Emmanuelle Peters. A systematic review and discussion of symptom specific cognitive behavioural approaches to delusions and hallucinations. *Schizophrenia Research*, 203:66–79, 01 2018. doi: 10.1016/j.schres.2017.12.014.

Eva Lüllmann and Tania M Lincoln. The effect of an educating versus normalizing approach on treatment motivation in patients presenting with delusions: An experimental investigation with analogue patients. *Schizophrenia research and treatment*, 2013, 2013.

Brendan A Maher. Delusional thinking and perceptual disorder. *Journal of individual psychology*, 30(1):98, 1974.

Brendan A Maher. Anomalous experience in everyday life: Its significance for psychopathology. *The Monist*, 82(4):547–570, 1999.

Eric Mandelbaum. Troubles with bayesianism: An introduction to the psychological immune system. *Mind and Language*, 34(2):141–157, 2019. doi: 10.1111/mila.12205.

Ryan McKay. Delusional inference. *Mind & Language*, 27(3):330–355, 2012.

Ryan P. McKay and Lisa Cipolotti. Attributional style in a case of cotard delusion. *Consciousness and Cognition*, 16:349–359, 2007.

Steffen Moritz and Todd S Woodward. Plausibility judgment in schizophrenic patients: evidence for a liberal acceptance bias. *German Journal of Psychiatry*, 7(4):66–74, 2004.

Steffen Moritz and Todd S Woodward. Jumping to conclusions in delusional and non-delusional schizophrenic patients. *British Journal of Clinical Psychology*, 44(2):193–207, 2005.

Steffen Moritz and Todd S Woodward. A generalized bias against disconfirmatory evidence in schizophrenia. *Psychiatry research*, 142(2-3):157–165, 2006.

Steffen Moritz and Todd S Woodward. Metacognitive training in schizophrenia: from basic research to knowledge translation and intervention. *Current opinion in psychiatry*, 20(6):619–625, 2007.

Steffen Moritz, Christina Andreou, Brooke C Schneider, Charlotte E Wittekind, Mahesh Menon, Ryan P Balzan, and Todd S Woodward. Sowing the seeds of doubt: a narrative review on metacognitive training in schizophrenia. *Clinical psychology review*, 34(4):358–366, 2014.

Simon Mullins and Sean A Spence. Re-examining thought insertion: Semi-structured literature review and conceptual analysis. *The British Journal of Psychiatry*, 182(4): 293–298, 2003.

Inez Myin-Germeys, NA Nicolson, and Philippe AEG Delespaul. The context of delusional experiences in the daily life of patients with schizophrenia. *Psychological medicine*, 31(3):489, 2001.

Raymond Nickerson. Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2:175–220, 06 1998. doi: 10.1037/1089-2680.2.2.175.

Charalampos Pandis, Niruj Agrawal, and Norman Poole. Capgras' delusion: A systematic review of 255 published cases. *Psychopathology*, 52:1–13, 07 2019. doi: 10.1159/000500474.

Hanna Pickard. Psychopathology and the ability to do otherwise. *Philosophy and Phenomenological Research*, 90(1):135–163, 2015.

Hanna Pickard and Lisa Ward. Responsibility without blame: Philosophical reflections on clinical practice. *Oxford handbook of philosophy of psychiatry*, pages 1134–1154, 2013.

Vilayanur S Ramachandran. The evolutionary biology of self-deception, laughter, dreaming and depression: some clues from anosognosia. *Medical Hypotheses*, 47 (5):347–362, 1996.

Marga Reimer. Only a philosopher or a madman: Impractical delusions in philosophy and psychiatry. *Philosophy, Psychiatry, and Psychology*, 17(4):315–328, 2010. doi: 10.1353/ppp.2010.0028.

John Rhodes and Richard GT Gipps. Delusions, certainty, and the background. *Philosophy, Psychiatry, & Psychology*, 15(4):295–310, 2008.

Gilbert Ryle. *The Concept of Mind*. University of Chicago Press, 1949.

Freddy Sarin, Lennart Wallin, and Birgitta Widerlöv. Cognitive behavior therapy for schizophrenia: a meta-analytical review of randomized controlled trials. *Nordic journal of psychiatry*, 65(3):162–174, 2011.

Louis Sass, Josef Parnas, and Dan Zahavi. Phenomenological psychopathology and schizophrenia: contemporary approaches and misunderstandings. *Philosophy, Psychiatry, & Psychology*, 18(1):1–23, 2011.

Louis A. Sass. *The Paradoxes of Delusion: Wittgenstein, Schreber, and the Schizophrenic Mind*. Cornell University Press, 1994.

Louis A Sass and Elizabeth Pienkos. Delusion: The phenomenological approach. *KWM Fulford*, 2013.

Susanna Schellenberg. *The unity of perception: Content, consciousness, evidence*. Oxford University Press, 2018.

J.A. Silva, G.B. Leong, R. Weinstock, and C.L. Boyer. Capgras syndrome and dangerousness. *Journal of the Americal Academy of Psychiatry and the Law Online*, 17(1): 5–14, 1989.

G. Lynn Stephens and George Graham. Reconceiving delusions. *International Review of Psychiatry*, 16:236–241, 2004.

Tony Stone and Andrew W Young. Delusions and brain injury: The philosophy and psychology of belief. *Mind & Language*, 12(3-4):327–364, 1997.

Peter Frederick Strawson. *Freedom and resentment, and other essays*, volume 595. Egmont Books (UK), 1974.

John Todd and Kenneth Dewhurst. The othello syndrome: a study in the psychopathology of sexual jealousy. *The Journal of nervous and mental disease*, 122(4):367–374, 1955.

Douglas Turkington, David Kingdon, and Peter J Weiden. Cognitive behavior therapy for schizophrenia. *American Journal of Psychiatry*, 163(3):365–373, 2006.

Oliver H Turnbull, Sarah Jenkins, and Martina L Rowley. The pleasantness of false beliefs: An emotion-based account of confabulation. *Neuropsychoanalysis*, 6(1):5–16, 2004.

M. Turner and M. Coltheart. Confabulation and delusion: A common monitoring framework. *Cognitive Neuropsychiatry*, 15:346–376, 2010.

David Velleman. On the aim of belief. In David Velleman, editor, *The Possibility of Practical Reason*, pages 244–81. Oxford University Press, 2000.

Esmé Weijun Wang. *The Collected Schizophrenias*. Graywolf Press, 2019.

Simon Wessely, A Buchanan, J Cutting, B. Everitt, Philippa Garety, and Pamela Taylor. Acting on delusions. i: Prevalence. *The British journal of psychiatry : the journal of mental science*, 163:69–76, 08 1993. doi: 10.1192/bjp.163.1.69.

Bernard Williams. Deciding to believe. In Bernard Williams, editor, *Problems of the Self*, pages 136–51. Cambridge University Press, 1970.

Timothy Williamson. *Knowledge and its Limits*. Oxford University Press on Demand, 2002.

Todd S Woodward, Steffen Moritz, Carrie Cuttler, and Jennifer C Whitman. The contribution of a cognitive bias against disconfirmatory evidence (bade) to delusions in schizophrenia. *Journal of Clinical and Experimental Neuropsychology*, 28(4):605–617, 2006.

Todd S Woodward, Lisa Buchy, Steffen Moritz, and Mario Liotti. A bias against disconfirmatory evidence is associated with delusion proneness in a nonclinical sample. *Schizophrenia bulletin*, 33(4):1023–1028, 2007.

Todd S Woodward, Steffen Moritz, Mahesh Menon, and Ruth Klinge. Belief inflexibility in schizophrenia. *Cognitive neuropsychiatry*, 13(3):267–277, 2008.

Til Wykes, Craig Steel, Brian Everitt, and Nicholas Tarrier. Cognitive behavior therapy for schizophrenia: effect sizes, clinical models, and methodological rigor. *Schizophrenia bulletin*, 34(3):523–537, 2008.

A. Young and K. Leafhead. Betwixt life and death: Case studies of the cotard delusion. In Peter Halligan and John Marshall, editors, *Method in madness: Case studies in cognitive neuropsychiatry*. Psychology Press, 1996.

Andrew W Young, Ian H Robertson, DJ Hellawell, KW De Pauw, and B Pentland. Cotard delusion after brain injury. *Psychological Medicine*, 22(3):799–804, 1992.