The Certainties of Delusion

Appearing in Moretti, L. and Pedersen, N. J. L. L. (eds) (2021) *Non-Evidentialist Epistemology*. Leiden: Brill.

1. Introduction

> [Rhodes] was talking to his patient Bryan while he (Bryan) performed Tai-Chi-style movements. Through the window, the sun could be seen to be setting. Bryan spontaneously said that he was 'setting the sun.' He appeared to fully believe that he was doing this through moving his body. (Gipps and Rhodes 2008: 302)

Bryan is delusional. His belief that he is able to "set the sun" and indeed doing so is a nice example of what is known as a bizarre delusion. It is so far out of the ordinary that it is apparent that something must be wrong with Bryan. Bizarre delusions are the most striking subset of an entire class of mental states, delusions, which play an important role as a diagnostic tool in psychiatry. Delusions are one of the marks of mental illness. But compare Bryan's case with the following passage from Wittgenstein's *On Certainty*:

> Men have believed that they could make rain; why should not a king be brought up in the belief that the world began with him? And if Moore and this king were to meet and discuss, could Moore really prove his belief to be the right one? (Wittgenstein 1969: §92)

The king's belief is equally bizarre as Bryan's, but for all we know the king—let us call him Arthur—is utterly sane. His environment is just very peculiar.

Chinese and Roman Emperors thought of themselves as gods. Wittgenstein takes Arthur's belief to be of a special kind: it is a *hinge* or a *certainty.* Indeed, Bryan's delusion and Arthur's belief are strikingly similar on several aspects, even if we put aside their bizarreness for a while. In this chapter I will argue that this is not a mere coincidence: delusions are a kind of certainty. This demystifies delusions: they are simply errant cousins of a type of belief that everyone has.

The key similarity is that both beliefs are extraordinarily resistant against contrary evidence. It is the mark of delusions that they are immune against contrary evidence. "Delusions are fixed beliefs that are not amenable to change in light of conflicting evidence." (DSM-5 2013: 87) Similarly Arthur would not accept any amount of evidence, for example the discovery of ancient ruins in the ground.

The only thing that might make Bryan snap out of it would be some transformative event, be it therapy or treatment with antipsychotics.[1] Similarly, Wittgenstein thinks that Moore—the champion of common sense—could not bring Arthur to abandon his young-earth-belief by reason: "I do not say that Moore could not convert the king to his view, but it would be a conversion of a special kind; the king would be brought to look at the world in a different way." (Wittgenstein 1969: §92)

---

[1] However, delusions tend to reappear when medication is laid off, e.g. (Munro 1999: 94).

I consider certainties to be distinct from hinges: certainties are beliefs of which we are so confident that no evidence will change that belief. Hinges are a subclass of the class of certainties; they are so central to an agent's belief system that changes in hinges would imply a wide-ranging shift in the belief system. Compare this also to Crispin Wright's notion of a "cornerstone" (2004).

John Campbell (2001) originally proposed that delusions are dysfunctional hinges, and generated some debate under the label of the "framework account" of delusions. In this chapter, I take an explicitly epistemological approach to the hypothesis—that is, I am less focussed on the psychiatric and therapeutic aspects of delusion. Rather, I am interested in what epistemology can learn from delusions. They constitute a breakdown of a certain type of convictions—certainties—and the systems with which we manage them. We can learn a lot about what a thing is by looking at when and how it stops working.

I will proceed by introducing and illustrating more precisely what delusions and certainties are. I then show in which aspects delusions resemble certainties to both establish their compatibility and the plausibility of the claim that delusions are certainties. Next, I present an argument for why we should think of delusions as a kind of certainty. Finally, I defend this position against different criticisms and I compare it to competing views.

2. What are delusions?

When I talk about delusions I do not mean the folk notion or the kitchen psychological diagnosis for some very irrational beliefs which I consider to be

more harmful than useful. I mean the above-mentioned symptom of mental illness. Delusions can have a wide range of contents. They reach from the banal to the bizarre as this incomplete anthology shows. (See also DSM-5 2013: 87)

Litigious delusion refers to beliefs that the patient has suffered some great injustice and must set all institutions into motion to even it out. (Munro 1999: 130–136) Morbid jealousy is an obsessive preoccupation with a romantic partner's infidelity. (Kingham and Gordon 2004) Then there is the great classic, delusion of persecution, colloquially known as paranoia.[2] But there are also more peculiar delusions such as delusion of infestation, where a patient for example believes that their skin is crawling with insects (Munro 1999: 85–86) or delusions of grandeur in which the patient thinks herself to be a great or important person (Munro 1999: 140–142). Finally, there are bizarre delusions like the Capgras delusion that persons close to the patient have been replaced by identical *doppelgängers* (ICD-11 2018: MB26.0B) or the Cotard delusion that the patient is dead (ICD-11 2018: MB26.0A).

I shall take the definition from the ICD-11 as a starting point for my account of delusions. Although I do have some misgivings about some aspects that will become clear in the course of this paper, I think that it is correct at its core.

---

[2] "Paranoia" was a concept specified by Emil Kraepelin in the late 19th Century to designate what is nowadays known as delusional disorder. (Munro 1999: 11–13) I shall not be using the term to avoid confusion.

> Delusion: A belief that is demonstrably untrue or not shared by others, usually based on incorrect inference about external reality. The belief is firmly held with conviction and is not, or is only briefly, susceptible to modification by experience or evidence that contradicts it. The belief is not ordinarily accepted by other members or the person's culture or subculture (i.e., it is not an article of religious faith). (ICD-11 2018: MB26.0)[3]

3. What are certainties and hinges?

Meanwhile, there is no ICD for epistemology and consequently no official definition of hinges and certainties. Indeed, the question is quite contested (*cf.* Coliva and Moyal-Sharrock: 2016). I will therefore argue for my own fairly minimal definitions of certainty and hinge.[4]

(C) A belief that P is a *certainty* for an agent S if and only if S holds it with such a high degree of conviction that no other belief or evidence would change the degree of conviction with which it is held.

(C) is a notion of psychological certainty that I take to be fairly common-sense. Certainty here is characterised by what claims of being (absolutely) certain

---

[3] Compare this to the DSM-5 definition (2013: 819).

[4] Note that it is not my goal to do an exercise in Wittgenstein-exegesis, examining what the author of *On Certainty* would have argued about delusions. Rather, I take certainties and hinges to be independently interesting epistemological phenomena.

mean. I am for example certain that Santa Claus does not exist—no amount of evidence would convince me otherwise. This certainty is internalist. I do not take certainty to be factive, a kind of knowledge-gold as in (Beddor: 2020). If I come across contrary evidence to such a certainty, I will reject the evidence as either false or misleading.[5]

Some certainties are about propositions that are so fundamental and central in our belief system that they inform the structure of all other beliefs. They influence what counts as evidence for what, and what sort of things there are. These certainties are called *hinges*, because they are the fix points or axes around which the rest of our belief system rotates. Danièle Moyal-Sharrock (2004: 52, 72) also makes this distinction but calls the *certainties* of (C) “subjective certainties”, while *hinges or hinge certainties* as defined in (H) below are the non-propositional content of “objective certainties”:

(H) A certainty that P is a *hinge certainty* or a *hinge* for an agent S if and only if a change in the conviction whether P would imply changes for many other beliefs of S; be they changes in their evidential support, or changes in the nature of the entities and properties contained in the beliefs.

For example, ‘the world began at my birth’ is Arthur’s hinge certainty. If Arthur stopped believing this hinge, this would have far-reaching implications about the

[5] Certainty would correspond to a credence of 1 which can never be updated to <1, at least in classical Bayesian calculus.

age of most objects and people. Ruins in the ground suddenly would become evidence for their being older rather than equally old as Arthur. And 'being a grandmother' would undergo a profound shift in meaning, given that there would now be women who actually gave birth to Arthur's parents.

As mentioned, hinges have been introduced by Ludwig Wittgenstein in his *On Certainty* (1969). Wittgenstein observed, how we just take some things for granted, we will not question them. Imagine you heard that bunny rabbits were vicious beasts and that they'd attack and kill you if you turned your back on them. You would reject that claim. You are certain that bunnies are harmless animals. Now, consider how you would react to seeing a bunny rabbit viciously attacking and harming someone. You would refuse to believe that that was a bunny rabbit—you're certain that they are too small and weak to do that. You would think you're hallucinating or being tricked. Arguably, it is a hinge about leporids and thereby bunnies, that they are not vicious and dangerous animals. Hinge certainties can fulfil this function *because* they are certainties.

The interesting thing about hinges is that we cannot do without them. Our epistemic households necessarily have loose ends—we cannot support everything with evidence. That is the lesson we can and should draw from sceptical arguments. Hinges tie in these loose ends. They fix them as certain and beyond evidence. There are hinges at all levels: logical axioms, principles of metaphysics and physics, rules of language, ideas about what and who we are.

They are the frame within or the background in front of which we interpret our evidence. The frame gives us starting points from which to deal with our evidence and it is not responsive to the evidence. Scepticism tries to undermine our knowledge by showing that nothing supports our hinges, i.e. the frame itself. Hinge-epistemology argues that this is misunderstanding how our epistemology works—already the notion of evidential support only makes sense within an evidence-independent frame that fixes these evidential relations. The frame holds, because we are *certain* of it.

Note that this is a strongly internalist notion of hinges and certainties, it focuses exclusively on the agent's point of view. Note also, that my notion of hinges is highly descriptive and thereby minimal: there is no normativity or teleology involved in it. Hinges and certainties are simply an essential feature of belief systems as we possess them. They play an analogous role to axioms and theorems in a logic—it leaves the normative status of the axiom outside of the logic undetermined.

Hinge-epistemology as a field has only recently taken off. Nevertheless, it has already given rise to a thriving literature. (e.g. Moyal-Sharrock 2004; Wright 2004; Coliva and Moyal-Sharrock 2016; Pritchard 2016) The main point about certainties and hinges for this paper is their functional role: they are independent of evidence and cannot be dislodged by any amount of reasoning or evidence.

4. Does one believe delusions and hinges?

I have defined certainties as beliefs. There are arguments about whether beliefs are essentially sensitive to evidence—there may be an epistemic state that is subject to such restrictions, but I will operate with a broad notion of 'belief' or doxastic state. I am committed to what is known as epistemicism in hinge-epistemology and as doxasticism in the debate about delusions, namely the thesis that these states are a kind of belief. I do not wish to delve into this argument and leave the defence of the assumption to others. (Wright 2004; Bayne and Pacherie 2005; Kusch 2016)

Still, in this chapter I do rely on both delusions and hinges being beliefs or doxastic states, therefore I will nevertheless explain why I treat both delusions and certainties as such. By a 'doxastic state' I mean a representational state whose content is taken to be actually the case. That is, it is essential for doxastic states that the subject accepts their content to be true. Clearly, neither delusions nor hinges are like more regular beliefs that are acquired by relying on evidence—nevertheless both represent things being or behaving in a certain way and they involve an endorsement of this being so. This means that they are doxastic states or beliefs in the broad sense.[6]

[6] Bayne and Pacherie (2005) also point to beliefs' distinctive phenomenology, however that is only the case with occurrent beliefs.

Additionally, as mentioned, I am interested in the epistemology of delusions. Treating delusions and certainties as doxastic states straightforwardly makes them epistemic states and subject to epistemic constraints. Meanwhile, if we took them to be some other kind of mental state, it would take a considerable amount of conceptual manoeuvring to examine their epistemological role, if it is possible at all.

Finally, this view is in a sense more light-weight. It takes a certain amount of presuppositions about how our mind works and what beliefs are to get the result that evidence-insensitive states cannot be doxastic. Meanwhile, it is quite common to see delusions and certainties naively described as 'beliefs' where arguably their doxasticity is at issue. Note specifically, that the DSM and the ICD ascribe belief and that we naively ascribe beliefs with certainties and hinges. We will touch upon doxasticism again later on.

5. How do delusions resemble certainties?

The thesis of this paper is that delusions are *defective certainties*. Is this plausible at all? I will begin by mentioning some of the striking similarities between delusions and certainties. Consider the ICD-11 definition of delusions again:

> A belief that is demonstrably untrue or not shared by others, usually based on incorrect inference about external reality. The belief is firmly held with conviction and is not, or is only briefly, susceptible to modification by experience or evidence that contradicts it. The belief

> is not ordinarily accepted by other members or the person's culture or subculture (i.e., it is not an article of religious faith). (ICD-11 2018: MB 26.0)

This looks like an individualist subset of the definition of certainties that I gave:

(C) A belief that P is a *certainty* for an agent S if and only if S holds it with such a degree of conviction that no other belief or evidence would change the degree of conviction with which it is held.

That is, from the subject's egocentric point of view, delusions and certainties just look the same. Both are beliefs which cannot be influenced by the environment. If the subject gains some evidence that seems to contradict a certainty or a delusion, then she will reinterpret the evidence as somehow misleading. The evidence is defeated by the very fact that the evidence appears to contradict a certainty or a delusion:

> The distinction *between a delusion and a strongly held idea* is sometimes difficult to make and depends in part on the degree of conviction with which the belief is held despite clear or reasonable contradictory evidence regarding its veracity. (DSM-5 2013: 87)

There are differences between my account of certainties and the ICD-11 definition though. First there is the bit that delusions are demonstrably false. Delusions do not necessarily need to be false. The FBI may keep even the paranoiac under surveillance, someone may have a delusion of infestation and still catch scabies.

The other difference is the ICD's individualism—the definition claims that disagreement is essential for delusions. I think that disagreement is at best a diagnostic tool; but it cannot be an essential trait of delusions. Because necessary disagreement would exclude the possibility of groups of individuals happening to share the same delusion even by accident.[7] Additionally, the definition in this form requires disagreement—why could I not suffer from delusions, if I lived in complete social isolation? I therefore argue that the incompatibility with the environment's belief set is more a heuristic than characteristic of delusion.

You might think that the possibility of groups sharing a delusion is quite far-fetched and would never occur in reality. However, delusions are contagious—they can be shared like certainties. We know of *folie à deux* or "induced delusional disorder" (ICD-10 2016: F24)[8], where a primary delusional patient

---

[7] If we sent a group of patients suffering from a delusion of infestation on holiday somewhere in isolation, this would not make them any less delusional.

[8] The diagnosis has been subsumed under the label of "Delusional Disorder" with others to simplify the diagnoses in the DSM-5 and ICD-11. (Biedermann and Fleischhacker 2016: 352) Note that delusional disorder itself is treated as a sort of leftover category from schizophrenia and transient psychosis. However there is still research on the topic of induced delusions, e.g. (Vigo *et al.* 2019; Jerrom *et al.* 2020). There is an important *caveat*: transmitted delusions may be indicative of independent mental health issues in the secondary patient.

transmits his or her delusion to a secondary patient, who did not at first harbour any delusions but is emotionally close to the primary patient. Usually, the primary and the secondary patient are socially isolated. The secondary patient adopts the delusion and can become as stable in his or her induced delusion as the primary patient. Often the secondary patient is a child who adopts his or her parent's delusion.

Some psychiatrists believe that cults represent cases of induced delusion in groups. It is hypothesised that cult leaders often suffer from delusions. The social isolation of cult structures as well as the strong emotional bonds within a cult may then lead to the spread of the delusion. (Munro 1999: 186) The ICD-definition would explicitly exclude even the possibility of this phenomenon—I therefore take its social dimension to only function as a diagnostic tool.

Folie à deux also shows another striking parallel between delusions and certainties: the way that they are adopted. We tend to adapt our outlook on the world to the one of people close to us; we establish a common ground—especially as children. That is, we adopt others' certainties and hinges without evidence —this is how we learn. *Folie à deux* would then just be a delusion hijacking the very natural mechanism by which we epistemically conform to our immediate environment.

Taking the insensitivity to evidence as a starting point, John Campbell (2001) has proposed a *hinge* view of delusion. His "framework approach" however is less influenced by epistemology than by reflections on philosophy of language. Campbell tries to preserve the delusional subject's rationality and agency—a

shift in hinges would explain the patient's beliefs and the behaviour that appears irrational to us.

However, according to Campbell, what comes along with this shift in hinges is a deep shift in language and world view. The delusional subject, *qua* delusional hinge, changes language—semantics and grammar shift along with the hinge. Consequently, the appearance that we use the same language as the patient is merely superficial. To someone with delusions of persecution 'following me' must mean something completely different—similarly 'love' must mean something else to someone with erotomanic delusions.

Campbell's framework view has been rightly criticised for this radical conclusion. (Bayne and Pacherie 2004) It would hardly be helpful to consider a delusional patient as someone to whom we have lost all means of access and with whom communication is a mere illusion. Indeed, it would be deeply problematic—the alienation would be total and all hope for both understanding and therapeutic treatment except by drugs would have to be given up! Only by self-inducing the delusion, i.e. adopting their hinge and losing grip of our community's language, could we reach the delusional patient.[9]

I believe that delusions are misguided certainties, not hinges. They therefore can remain much more superficial and conservative. The delusional subject

---

[9] Clearly, it can be difficult to communicate with such patients—but this is only partially due to their having delusions and also owned to emotional volatility and the fact that patients may easily be distracted.

simply adopts a few, maybe only one, additional *certainty* on top of all the hinges and certainties they already possessed before.[10] That is the subject's world view remains essentially preserved—she is simply now certain of one particular thing where this particular conviction may be emotionally highly charged and therefore a constant issue due to the syndrome.

That is, the patient's language remains essentially the same as does her world view: delusional subjects are usually aware that their delusion is extraordinary, they do not expect to be believed just like that. Additionally, if patients are asked how they would react to someone else making the same delusional claims as them they may recognise those to be delusional. (Bell, Halligan and Ellis 2003: 4; Bayne and Pacherie 2004: 9) This indicates that they haven't transformed their entire world-view and language.

However, there may be space for Campbell's analysis of delusions. If I am correct that delusions are certainties and hinges are a subset of the set of certainties, then it seems clearly within the remit of possibility that a delusion becomes a hinge. This would mean that the subject's delusional certainty would have far-reaching implications in her belief system, although not necessarily as extreme as predicted by Campbell. There are delusions that may satisfy this hinge criterion: some delusions are considered to be bizarre due to their impossible-seeming content. (ICD-11 2018: MB 26.0) Namely, it seems a

---

[10] Obviously, if for example a schizophrenia progresses the patient might accumulate delusions.

plausible hypothesis that bizarre delusions would be delusions caused by delusional hinges, while non-bizarre delusions are mere certainties. A version of this reading has been proposed in (Klee, 2004).

6. Why delusions are certainties

All of this, at best, plausibilises the contention that delusions are certainties. But there is more than just *prima facie* similarity. My argument for the view that delusions are certainties runs as follows.

P1 Delusions are doxastic states.

P2 Delusions are not controlled by evidence.

P3 Certainties are doxastic states that are not controlled by evidence.

P4 Apart from certainties there is no other class of doxastic states that are not controlled by evidence and that is compatible with subsuming delusions.

Therefore

C5 Delusions are a subset of the class of certainties.

The first three premises are relatively unproblematic. P1 and P2 simply arise from the ICD-11 definition of delusion, while P3 follows from my definition of certainties. As mentioned above, P1 and the doxastic status of certainties in P3

are somewhat controversial—however I will take this for granted. I.e. I will ignore the larger debate about doxasticism and epistemicism.

One might counter against P2 that delusional subjects clearly take certain things to be evidence for their delusion; the DSM-5 also states that delusions are based on inference. In delusions of reference for example each newspaper headline is taken to be further evidence for the delusion that the newspapers write about the patient. Similarly, in delusions of persecution, patients interpret patterns such as how the cars are parked in a street as secret messages between their persecutors; they interpret it as evidence for their persecution.

However, these examples illustrate that the evidence for the delusional acceptance already itself relies on the delusion in order to function as evidence. Take the example of a patient inferring from seeing a row of marble tables that the world is ending. (Campbell 2001: 95) Without some delusional certainty, this bizarre inference could not be made. This is the very same sense in which certainties are not controlled by evidence—they are what makes certain things into evidence in the first place. Without the given certainties, there would be no evidence for them.[11]

Clearly P4 is the most questionable premise of my argument. I will therefore have to defend this claim more extensively. There are several ways to deny P4:

---

[11] An exception would be delusions that are partially supported by hallucinations. But also interpreting, for example, acoustic hallucinations as alien radio emissions would require a delusional certainty.

either one claims that delusions form their own class of doxastic state or one argues that there is some other kind of doxastic state to which delusions belong. I find both alternatives lacking.

I have two objections against the claim that delusions form their own class: first, it is not very parsimonious. Given that hinge epistemology and the notion of certainty I presented have independent standing, why would we need to create an entire separate doxastic class for delusions? What distinguishes delusions so much from certainties that we would need a separate class? And what do we gain in terms of explanatory power by treating delusions as a class apart? It would just be restating that delusions exist.

Second, I do not think that delusions are the right kind of thing to form its own class of doxastic states. Delusions are epistemological defects—something went wrong with a subject's epistemology when they hold a delusion. It seems odd to grant a defect its own class just *in virtue of* its defectivity. A defect always depends on some function that is not fulfilled—without some function, no defect. The only option to treat delusions as its own kind of doxastic state would then be to claim that they constitute the class of defective doxastic states, but this cannot be: there are many more kinds of defective doxastic states, most generally false beliefs, but also self-deception or the results of sloppy reasoning. There seems to be no positive reason to grant delusions a status independent of certainties.

The other objection against P4 is that delusions belong to another class of doxastic states that are not controlled by evidence, but which are no certainties.

I shall consider three alternatives: the first is that delusions are derailed assumptions, the second that delusions are misinterpreted imaginations, and the third that delusions are uncontrolled memories.

Assumptions are evidence-insensitive in the sense that they are adopted without any need for evidence. They bridge a lack of evidence we sometimes have, but where we are forced to decide. However, assumptions are not immune to evidence in every sense: neither would we adopt nor maintain an assumption against which we have defeating evidence or against which defeating evidence has come up. This distinguishes assumptions from delusions—given a delusion, contrary evidence is itself defeated. Delusions do not bridge missing information, they are independent of other information.

The second option is that the delusions are a kind of imagining. This account has been suggested by Currie (2000), taking evidence of how schizophrenic patients appear to process information. The imaginings in question, I guess, are the sort of thing we do in thought experiments: “suppose that an articulate voice were heard in the clouds, […] that the words spoken from the sky were not only meaningful but conveyed some instruction” as for example David Hume (1993: 54) suggested. This way of imagining isolates the imagined proposition from contrary evidence—in that sense it is insensitive to evidence.

Currie (2000: 174–175) suggests that delusions arise because the subject cannot distinguish between her imaginings and her beliefs. That is, she will treat both states in the same way. I do not think that this diagnosis of delusions holds up. Imaginings are isolated from evidence *because* their content is not

endorsed as true. Currie argues that in delusion imaginings become more similar to beliefs. Note that Currie is taken to be a non-doxasticist—that is the orthodox reading of his proposal goes against P1 of my argument and falls outside the scope of this paper. For a careful critique of the non-doxasticist approach to imagination as delusion, please refer to Bayne and Pacherie's (2005).

I believe that delusions essentially involve an endorsement of their content as true. Indeed, their truth is felt with great intensity—that is part of what gives delusions such power over the patient. Take an example: if I suffer the delusion that someone is trying to poison me, then I take there to be a real danger that my food or drink will kill me. I am actually scared of what my food might do to me and I act accordingly—as can be seen for example from Kurt Gödel's fate. He did not merely pretend or imagine that he would be poisoned, he genuinely feared that it would happen. If imaginings became such states, they would not be imaginings anymore—Currie (2000: 178) concedes this point. Imagination can only be an aetiological source of delusions, not delusion itself.

As mentioned, imaginings are evidentially isolated *because* we do not endorse their content as true. If we lose this non-endorsement, then the imagining would lose its evidential isolation. Arguably, this representational state would even stop being an imagining and instead become more of a consideration of a proposition or if loss of non-endorsement means endorsement, an irrational belief.

A further problem with the imagining as delusion approach is that, to my knowledge, delusional subjects do not lose their ability to imagine things.[12] We imagine things of all sorts all the time. If a delusional patient confused his imaginings with beliefs, his delusion would become truly florid. They'd be 'believing' one thing now and the contrary a few seconds later, it would be a true kaleidoscope or firework of 'beliefs'. There may be syndromes like that, but this hardly yields a sufficiently general account of delusions. I will therefore disregard imaginings as an account for delusions.

The third option originates in the psychiatric literature: delusions are false memories that have taken hold. (Gibbs and David 2003; Moritz and Woodward 2006) Apparently, delusional patients have a lower evidential threshold for when to accept or reject a belief. This may lead to patients' misremembering imaginings or confabulations as real. This delusional memory is then reinforced through emotionally facilitated recall.

This account is somewhat orthogonal to the epistemological approach—in a sense *any belief* that is not based on current perceptions or inferences is a memory. This would hardly be informative epistemologically speaking. Indeed, Gibbs and David argue that delusion is grounded in episodic memories that are not controlled for whether they were actual events. (Gibbs and David 2003: 171–172)

---

[12] Maybe patients with schizophrenia do, but schizophrenic delusion is a subset of the field of delusion.

I am sceptical of this approach. First, if there is a lowered threshold for accepting some idea as true in delusion—why does the patient not continually change her memories and beliefs? Low evidential thresholds would, it seems, rather have a destabilising effect on our belief system than the doxastic ossification that we can observe in delusion. After all, delusions are characterised by that they are *not* dismissed but instead that they are maintained against all evidence. What would make them so special? The simple fact that they happen to be more frequently recalled? Under any circumstance, nothing in this account of delusions explains why they are immune against evidence and reasoning.

The second problem is that I am not sure whether delusions always are the kind of beliefs that can be grounded in episodic memories. What sort of memory would be at the bottom of a delusion of reference? Presumably, only few delusional subjects have actually had a news item about themselves, and even if so how would a single memory lead to the patient's belief that *all* news items are about them? Or what about the admittedly more peculiar case of Cotard syndrome? Cotard patients, to my knowledge, do not report that they remember dying. In sum, I think that memories, however imaginary or outlandish, cannot have the right kind of content in order to constitute generalised delusions. What may be is that delusions are supported and reinforced by distorted memories—but the memories are not the delusions themselves or their immediate source.

I would not know what other avenues there might be to deny P4. Meanwhile, all of the above-mentioned alternative suggestions to classify delusions were

unsatisfactory. Any other account of delusions would have to be dealt with in a similar manner. A strategy I didn't take is to explain how some alternative solution is, itself, an instance of a certainty. That is, certain suggestions made to deny P4 might actually themselves confirm P4.

7. How do delusional certainties remain confined?

Bayne and Pacherie have made a careful critique of Campbell's (2001) framework account of delusions. One of the objections they make against Campbell might gain traction against my proposal: delusions frequently are "encapsulated". That is, the delusional subject does not draw the appropriate conclusions from his or her delusion or always behave as if it were the case. Instead, it remains quite insulated and does not structure the patient's entire thinking and acting—something that hinges arguably would do, given their role as the fundamental presuppositions for epistemic and practical agency. (Bayne and Pacherie 2004: 9)

I already distinguished between hinges and certainties. Notably, I argued that delusions are certainties but not necessarily hinges. Certainties may remain fairly isolated beliefs. They do not need to have far-reaching consequences—an advantage over Campbell's view that I already hinted at above. Certainties may not always be salient, thus you may be implicitly certain of something[13] without being aware of all of its consequences. This differentiates certainties from

---

[13] Just as you may implicitly believe something.

hinges, which by their nature will influence the content of other beliefs. Thus the restricted nature of certainties may explain the encapsulation of delusions.

I also considered the possibility that there may also be delusional hinge certainties as proposed by Campbell (2001) and Klee (2004), which would then yield bizarre delusions. How might one counter Bayne and Pacherie's (2004) criticism in that case? I guess one would have to argue that delusional hinges are remarkably specific in their content. Delusions usually ascribe some singular property to some particular object or a restrained class of objects: my body/an organ is X in somatic delusions, I am not alive in Cotard's, my thoughts are being manipulated with thought insertion. Consequently, they tend to only have consequences in relation to that specific content. Possibly, delusional hinges are in fact more specific in their content than a patient's utterances would seem to imply. To take the example of Cotard's this would mean that 'dead' has shifted in its meaning given the patient's delusional hinge, while other beliefs remain unchanged.

This however contravenes the notion of 'hinge' as defined in (H). A possible solution might be to consider 'hinge' to be a graded notion. The more implications a certainty has, the 'hingier' it is. Delusions about too consequential hinges would imply undiagnosable, utter alienation.

Encapsulation is then explained by the fact that delusional hinges and certainties frequently seem to be simply planted on top of a patient's otherwise intact belief system. These delusional certainties do not replace much, except for certainties whose direct negation they are such as 'I am alive' in Cotard's.

That is, most of the delusional patient's epistemic structure will actually be preserved and still guide them in many respects. Encapsulation would arise from the manner in which patients attempt to integrate their delusional certainty or hinge into their otherwise intact world-view.

8. Where do delusions go wrong?

I have argued that delusions are somehow defective certainties, but my argument left open what makes them defective. The traits that the ICD-11 proposes as characteristic of delusions won't do: neither do delusions need to be false, nor is immunity to evidence and reasoning exclusive to delusions. Also the disagreement between a delusional subject and her environment is more diagnostic than definitory—entire groups may be delusional or individuals may harbour a delusion that is accepted by their environment. The DSM recognises this difficulty by pointing out that delusions and firmly held beliefs may be hard to distinguish. (DSM-5 2018: 87)

What other avenues are there to distinguish delusions from other certainties? One suggestion has the vice of being remarkably uninformative but possesses the virtue of being probably true: the aetiology of how the delusion was adopted and is maintained. Although there is no tell-tale evidence of what goes wrong in delusions neurologically speaking, there is some atypicality. (Kunert, Norra and Hoff 2007) This suggestion is unhelpful in several senses. A philosopher stating that a neurologist ought to be able to find something won't help anyone. But also epistemologically speaking, this would be a capitulation—we ought to be

able to say more than that a patient is delusional if and only if her brain is not working as it should. After all, we take the delusion itself as evidence for that. It is circular as a characterisation, but probably true because there may very well be some neurological atypicality that coincides with delusional beliefs. Additionally, anti-psychotics are a tested, successful treatment for delusions—thus neurological mechanisms appear to play a role in the formation of delusions.

A more interesting differentiation is that of functionality. We possess certainties and hinges because they enable us to act and think—they guarantee our practical and epistemic agency, that is their function. Delusions reduce our agency: they hinder us from pursuing practical and theoretical projects, they distract us from what is at hand and so on. Delusions may be of the same *doxastic* type as certainties, but they fail to fulfil their functional or *epistemic* role. To make a medical metaphor: delusions are dysfunctional certainties just as leukaemia cells are dysfunctional leukocytes.

A specific corollary of this function-account would be that delusions are dysfunctional in that they are harmful to the subject and his or her environment, while regular hinges are there to keep us out of harm's way. Clearly, also regular certainties can have harmful consequences, for example sexism and racism may be grounded in certainties, but delusional certainties are much more immediately harmful. In cases of Capgras for example, patients are much more frequently violent than healthy subjects. (Bayne and Pacherie 2004: 6)

An admittedly speculative difference between delusional and regular certainties may be that delusional certainties express emotions. The content of most delusions seems to channel deep-seated fears (of poisoning, infestation, social exclusion) or hopes (grandeur, romantic delusions). I would guess that delusional beliefs anchor themselves because their content is correlated with deeply felt emotions. This does not imply that each mention or elaboration of the delusion will cause these emotions—it may very well be that delusions can also detach themselves from their emotional growing ground. Under any circumstance, regular certainties do not seem to express emotions as strongly. One exception may be moral certainties, if you are an expressivist.

9. An alternative framework account

John Campbell's framework theory of delusion has not been the only one. Rhodes and Gipps (2008) have proposed a very interesting alternative framework account. Instead of arguing that delusions are positive certainties, they suggest that they are the symptom of missing hinges. Indeed, often we recognise delusions by their bizarreness—and bizarreness of a belief arguably arises when we could not see how someone would believe *that*. That is, a belief appears as bizarre if it goes against our standing hinges. Consequently, a delusional subject must be lacking those hinges. For example, Bryan from the opening of this paper clearly lacks the certainty that we cannot influence the sun by simply moving our body. We recognise delusion because it is incompatible with the hinges we possess. (Rhodes and Gipps 2008: 300–301)

This approach generates similar difficulties as Campbell's: delusional subjects become absolutely inaccessible to us. The fact that they do not share our hinges leads to a completely different world-view that is deeply incompatible to ours—a delusional subject should consequently not be able to recognise how extraordinary their claims are, something they occasionally do. As said, Rhodes and Gipps argue that the divergent framework is what makes delusion diagnosable, but as Bortolotti rightly points out, hinge beliefs are not set in stone, for example they change through time. (Bortolotti 2011: 81–82) I.e. not everyone disagreeing about our framework is delusional. Additionally, there are very ordinary, "pedestrian" (Klee 2004), delusions that seem not to be caused by missing hinges.

On a different note, nothing in Rhodes and Gipps' approach explains why delusions are immune to all sorts of contrary evidence or reasoning. It only explains why *we* cannot reach delusional subjects and convince them: because they do not share the certainties we would use to convince them. Meanwhile this does not explain why their delusions are so fixated, and as Bortolotti points out, from missing certainties we would expect delusions to be flourishing all over the place. (Bortolotti 2011: 83) I therefore do not think that missing hinges explain delusions.

10. Conclusion

What are then delusions to the epistemologist? They often seem like deeply irrational beliefs—so irrational that some have denied that they even are beliefs.

I have taken a different approach: I argued that delusions are certainties, something that all of us possess. Given that we understand how certainties work, we can extrapolate how delusional subjects deal with their delusion. They take their delusion to be certain. Anything that seems contradictory to it will be brushed aside or if they are confronted with a contradiction, this will generate considerable resistance and discomfort, i.e. cognitive dissonance. We experience the same if our standing certainties are put into question. Thus delusions are nothing profoundly mysterious, they do not render patients utterly alien. Rather, some neural or other cognitive defect makes them be absolutely certain of something.

In the other direction, classifying delusions as certainties puts the latter category under scrutiny. If certainties as a class contains both the hinges on which algebra is built and the delusion that I am President of the US, then we need to pay careful attention to certainties. It raises the question which certainties are acceptable, and which are not. What are the criteria for warranted certainties and how do we recognise a certainty as warranted? Additionally, we may wonder whether there are certainties, with a non-pathological aetiology that are as problematic as delusions. Some instances of *folie à deux* may belong to this problematic category.

11. Bibliography

American Psychiatric Association (ed.). 2013. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*. Washington, D.C.: American Psychiatric Publishing.

Bayne, T., and Pacherie, E. 2004. "Bottom-Up or Top-Down: Campbell's Rationalist Account of Monothematic Delusions," *Philosophy, Psychiatry, & Psychology* 11(1): 1–11.

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

Beddor, B. 2020. "New Work for Certainty," *Philosophers' Imprint* 20(8): 1–25.

Bell, V., Halligan, P., and Ellis, H. 2003. "Beliefs about delusions," *Psychologist* 16(8): 418–422.

Biedermann, F. and Fleischhacker, W. W. 2016. "Psychotic disorders in DSM-5 and ICD-11," *CNS Spectrums* 21(4): 349–354.

Bortolotti, L. 2011. "Continuing Commentary: Shaking the Bedrock," *Philosophy, Psychiatry, & Psychology* 18(1): 77–87.

Campbell, J. 2001. "Rationality, Meaning, and the Analysis of Delusion," *Philosophy, Psychiatry, & Psychology* 8(2): 89–100.

Coliva, A., and Moyal-Sharrock, D. (eds.). 2016. *Hinge Epistemology*. Leiden: Brill.

Currie, G. 2000. “Imagination, delusion and hallucinations,” *Mind and Language* 15(1): 168–183.

Gibbs, A. A., and David, A. S. 2003. “Delusion formation and insight in the context of affective disturbance,” *Epidemiologia e Psichiatria Sociale* 12(3): 167–174.

Gipps, R. G. T., and Rhodes, J. 2008. “The Background Theory of Delusion and Existential Phenomenology,” *Philosophy, Psychiatry, & Psychology* 15(4): 321–326.

Hume, D. 1993. “Dialogues Concerning Natural Religion.” In J. C. A. Gaskin (ed.), *Dialogues Concerning Natural Religion and The Natural History of Religion,* 29–130. Oxford: Oxford University Press.

Jerrom, R. *et al.* 2020. “A case series of shared delusional infestation: folie à deux revisited,” *Clinical and Experimental Dermatology* 45(4): 414–416.

Kingham, M., and Gordon, H. 2004. “Aspects of morbid jealousy,” *Advances in Psychiatric Treatment* 10(3): 207–215.

Klee, R. 2004. “Why Some Delusions Are Necessarily Inexplicable Beliefs,” *Philosophy, Psychiatry, & Psychology* 11(1): 25–34.

Kunert, H. J., Norra, C., and Hoff, P. 2007. “Theories of delusional disorders: An update and review,” *Psychopathology* 40(3): 191–202.

Kusch, M. 2016. “Wittgenstein on Mathematics and Certainties.” In A. Coliva, and D. Moyal-Sharrock (eds.), *Hinge Epistemology*, 48–71. Leiden: Brill.

Moritz, S., and Woodward, T. S. 2006. “Metacognitive control over false memories: A key determinant of delusional thinking,” *Current Psychiatry Reports* 8(3): 184–190.

Moyal-Sharrock, D. 2004. *Understanding Wittgenstein’s On Certainty*. Basingstoke: Palgrave Macmillan.

Munro, A. 1999. *Delusional Disorder*. Cambridge: Cambridge University Press.

Pritchard, D. 2016. *Epistemic Angst*. Princeton: Princeton University Press.

Rhodes, J., and Gipps, R. G. T. 2008. “Delusions, Certainty, and the Background,” *Philosophy, Psychiatry, & Psychology* 15(4): 295–310.

Vigo, L. *et al.* 2019. “Shared psychotic disorder in children and young people: a systematic review,” *European Child and Adolescent Psychiatry* 28(12): 1555–1566.

WHO (ed.). 2016. *International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10)*. Geneva: World Health Organization. URL: http://apps.who.int/classifications/icd10/browse/2016/en

WHO (ed.). 2018. *ICD-11 for Mortality and Morbidity Statistics (ICD-11)*. Geneva: World Health Organization. URL: https://icd.who.int/browse11/l-m/en

Wittgenstein, L. 1969. *Über Gewissheit - On Certainty*. G. E. M. Anscombe, and G. H. von Wright (eds.). Oxford: Blackwell.

Wright, C. 2004. “Warrant for Nothing (and Foundations for Free)?,” *Aristotelian Society Supplementary Volume* 78(1): 167–212.