

Why are we certain that we exist?

please quote the published version <http://onlinelibrary.wiley.com/doi/10.1111/phpr.12113/abstract>

Alexandre Billon

Univ Lille Nord de France, UdL3, STL (CNRS, UMR 8163)

If I had merely ceased thinking, even if everything else I had ever imagined had been true, I should have had no reason to believe that I existed. (Descartes, 1985, I, 127)

Abstract

Descartes was certain that he was thinking and he was accordingly certain that he existed. Like Descartes, we seem to be more certain of our thoughts and our existence than of anything else. What is less clear is the reason *why* we are thus certain. Philosophers throughout history have provided different interpretations of the cogito, disagreeing both on the kind of thoughts it characterizes and on the reasons for its cogency. According to what we may call the *empiricist interpretation* of the cogito, I can only claim to be certain of having *experiences*, and this certainty, as well as that of my own existence, stems from their phenomenal and subjective character. According to *rationalist interpretations*, on the other hand, I am certain of having some self-reflexive *propositional attitudes*, and this certainty derives from their rational features. Psychiatric patients suffering from acute forms of depersonalization or of the Cotard syndrome often doubt that they think and exist, and might even believe that they don't. I argue that their study allows us to favor the empiricist interpretation of the cogito.

Keywords : cogito, certainty, phenomenal consciousness, subjective character, reflexivity, first-person, self-awareness, rationality, depersonalization, Cotard syndrome.

Descartes claimed that the only things he could be certain of were that he was thinking and that he existed. By imagining being tricked by an evil genius, he could doubt almost anything, including the deliverances of his senses. He could not, however, imagine being deceived into believing falsely that he thought or existed.

Even if it hasn't occurred to us yet, and even it doesn't have for us the significance it might have had for him, we seem to share Descartes' certainty in our thoughts and our existence. Of course, some philosophers have challenged this certainty, sometimes even claiming that they do not themselves think or exist (Unger (1997), who entitled a paper "I do not exist", is a clear example).¹ These "self-doubters" agree, however, that it takes substantial philosophical theorizing to raise such doubts and that we are at least *prima facie* certain that we think and exist: after some consideration, we find that the proposition 'I think and I exist' is certain and that it passes some tests that are criterial for certainty. It passes, in particular, what may be called Descartes' evil genius test.² What such self-doubters deny is only that this *prima facie* certainty can withstand deep philosophical scrutiny and can, accordingly, be claimed to be a form of *ideal* certainty.

I won't defend the cogito argument against those philosophical self-doubters here. I will also ignore the exegetic questions surrounding Descartes's own formulation and defense of the cogito. My aim will be to figure out why we are *prima facie* certain that we think and exist—call this the question of the *cogito interpretation*. Philosophers throughout history have provided different interpretations of the *cogito* and have disagreed on the reasons for its cogency. They have concurred that there is a specific class of thoughts that we can be certain to have and that give rise to the certainty that we think and exist—call these the *cogito-like thoughts*. But they

¹ Notice that it is not obvious that philosophers who raise doubts about the existence of "selves" (whatever that is) would accordingly welcome doubts about their own thoughts and existence (see Kennedy and Graham (2006) for a very useful discussion).

² It is common to distinguish, after Descartes, moral from absolute certainty by stipulating that while the latter can be defeated by any *rational* doubt, the former can only be defeated by *reasonable* doubts. As should be clear, I am only concerned here with absolute certainty.

have disagreed both on the extension of those cogito-like thoughts, and on the features of those thoughts that explain our certainty.

It is customary to distinguish two broad kinds of feature of our thoughts. If, for example, I suddenly fear that we might lose the war, there will be something it is like for me to have my thought. We say that my thought will have *phenomenal features*. If I have this thought, I should also be in a position to judge that I belong to a group that is engaged in a war, that (being at war) we are not at peace, etc. That is, my thought will have certain *rational features*, which give it its rational role. Rational and phenomenal features are at least conceptually distinct. Phenomenal features are characteristic of experiences, and arguably individuate them (no difference in experience without a phenomenal difference and vice versa). Rational features are characteristic of propositional attitudes, which are at least partly individuated by them (different propositional attitudes will have different rational features). Corresponding to these two aspects of thoughts we can distinguish two interpretations of the cogito. According to what we may call the *empiricist interpretation*, I can only claim to be certain about my experiences, and this certainty, as well as that of my own existence, stems from their phenomenal character. According to *rationalist interpretations*, on the other hand, I can only be certain about some of my token propositional attitudes, and this certainty, as well as that of my existence, derives from their rational, rather than from their phenomenal features.

Psychiatric patients suffering from acute forms of depersonalization or of the Cotard syndrome often doubt that they think and exist, and might even believe that they don't. I will argue that the study of those *pathological self-doubters*, who express doubts about their own existence, allows us to favor the empiricist interpretation of the cogito.

Here is the outline of my argument. There are basically three plausible interpretations of the cogito in the literature, one empiricist and two rationalist (§ 1). Different kinds of interpretation make different empirical predictions about subjects who might be *prima facie* uncertain whether they think and exist. The empiricist interpretation predicts that their *experience* should be abnormal (§ 2). Rationalist interpretations predict that the relevant subjects should fail to be “minimally rational” in a sense that I shall soon make explicit. Pathological self-doubters are the only people who are *prima facie* uncertain whether they think and exist (§ 3). There are independent reasons to consider that they are minimally rational (§ 4.1) and that their experience is abnormal in just the way the empiricist interpretation predicts (§ 4.2). Paying attention to patients’ reports furthermore allows us to understand why the rationalist interpretations of the cogito are wrong; why, that is, one can be minimally rational and yet uncertain about one’s thoughts and one’s existence (§ 5). The case of patients suffering from depersonalization thus provides evidence for the empiricist interpretation of the cogito and against its rationalist rivals.

After more than eighty years of quasi-hibernation, the literature on the Cotard syndrome and depersonalization is burgeoning again. Even if I believe that focusing on patients’ uncertainty about their own existence could yield an interesting contribution to this literature, I must stress that my aim here is not to defend a novel account of these disorders.³ In particular, I take it that everything I will say about the Cotard syndrome here is consistent with most interpretations that construe it as a minimally rational reaction to abnormal experiences.

³ I pursue this task elsewhere (Billon, 2014).

1 The *cogito* and its interpretations

I know of only three plausible interpretations of the *cogito*. The first is empiricist. The other two are rationalist.

1.1 An empiricist interpretation of the *cogito*

The empiricist interpretation starts with the claim that for phenomenal features, there is no appearance-reality distinction. To appear to be in pain and to be in pain, for example, is one and the same thing. Call this the *revelation thesis*. Assuming that it holds a priori, the revelation thesis seems to entail that when my experience has a certain phenomenal feature, I will normally be in a position to judge that my current experience has this feature and this judgment will be certain in the following sense: it will be inconceivable for me that this judgment is mistaken.⁴

It might be replied that the revelation thesis prevents one common kind of error, in which the appearance of a thing is mistakenly taken for what the thing really is, but that it does not prevent others. For even if my phenomenal features appear to me as they are, as the revelation thesis holds, I might fail to conceptualize those appearances properly. If I did not master the relevant concepts, I surely could take one experience type for another, say a tickling for a pain.⁵ Similarly I could misclassify an unusual experience, or one that lies at the limits of my introspective

⁴ Like for example Kripke (1980, 125, 143fn.72), I take inconceivable falsity and Cartesian certainty to be equivalent.

⁵ It might seem strange to say that an experience appears to me correctly even though I mischaracterize it. The problem comes from the fact that the appearance of an experience has two senses, one phenomenal, which figures in the revelation thesis and one epistemic, which is connected the way I characterize this experience. Those two sense of appearance are not always aligned. See Gertler (2010, 100-110) for more on that.

powers of discriminations. I could for example confuse the taste of an unusually tannic Bourgogne wine and that of a light Bordeaux. However, it seems reasonable to say that such gaps between the way a phenomenal feature appears to me and the way I conceptualize it will disappear when it comes to very simple phenomenal features, such as the property of having a phenomenal character. Accordingly, even if I can doubt *which* phenomenal features one of my mental states has, I cannot doubt *that* I have a state with phenomenal features when I have one.

Let thinking_e (the subscript *e* stands for empiricist) be “thinking” construed as a phenomenally conscious state. The point can be spelled out by saying that when I am thinking_e and I believe I am, this belief will be indubitable. The following proposition (I use single quotes for propositions and their constituents, double quotes for sentences and words) is thus indubitable:

- (C_e) ‘I am thinking_e’

It should be stressed that I am not only certain that there is a thought_e but also that this thought is mine. Phenomenal features of my mental states normally seem mine to me. We may say that they have a *subjective character*, or more simply that they are *subjective*.⁶ Just like having a *phenomenal character* is a phenomenal feature of my experiences in virtue of which they correctly appear as *phenomenally conscious* states to me, their *subjective character* is a phenomenal feature of my experiences in virtue of which they correctly appear as *being mine* to me. Finally, just like having a phenomenal character, being subjective is a simple phenomenal

⁶ It is arguable that this claim holds not just normally but universally and constitutively. Such a thesis, which I have dubbed elsewhere the *Cartesian Principle*, is suggested by Descartes himself (Descartes, 1985, II,19). It is endorsed, among others by James (1983), Jaspers (1997, 578) and more recently Levine (2001); Kriegel (2004a).

feature that we cannot normally mischaracterize or misconceptualize. The subjective character of my conscious states is accordingly indubitable as well. In virtue of the subjective and phenomenal character of my thoughts_e, then, I am certain that there are thoughts_e (phenomenal character) which are mine (subjective character), and that I exist, at least as a subject of thoughts_e.

This kind of empiricist interpretation of the cogito is quite pervasive in the phenomenological tradition (see Husserl (1999), Henry (1985, Ch. I-II)).⁷ It is also arguably endorsed by Russell (1912, Ch. II), as well as Chisholm (1982, 10) and Williams (1978, 64-70)).⁸

Although still programmatic, the above empiricist account of the cogito is developed enough to be contrasted with the rationalist accounts discussed below. Its most central features are:

- (Ei) Cogito-like thoughts (the thoughts we can be certain to have that give rise to the certainty that we think and exist) are thoughts_e.
- (Eii) The certainty of the cogito stems from their phenomenal and their subjective character.
- (Eiii) We just need to be aware of their phenomenal and subjective character to acknowledge this certainty.

⁷ Phenomenologists who, like Heidegger or Merleau-Ponty and unlike Husserl, do not claim the Cartesian heritage are notable exceptions. Merleau-Ponty (2007, 98) for example characterizes one of the rationalist interpretations as “the only solid meaning of the *cogito*” and rejects the empiricist interpretation as merely “psychological”. Heidegger (1997, 133) takes reflexivity to be essential to the *cogito* and criticizes it for that.

⁸ It is mentioned (and criticized) by Shoemaker (1996, 53) and Brandom (1997, 137).

We shall see that this interpretation differs from the rationalist ones in that it explicitly mentions *phenomenal consciousness* and that it does not invoke, as the latter do, the notion that some mental states refer to themselves (are “self-reflexive”).

1.2 A first rationalist interpretation of the cogito: rationality and ‘thought’

The proposition ‘I am saying something’ has an interesting property (Williams, 1978, 59). It will be true each time I say it but it will be false if I just write it down. ‘I am asserting something’ will be true whether I say it or write it. It will be false if I merely consider it or doubt it. ‘I am thinking’ can be seen as a degenerate case of such propositions that should come true each time I think it. This involves construing “thinking” as merely engaging in a proposition, whether by asserting it, doubting it, or just entertaining it. We think in that sense whenever we have a token propositional attitude. I will call such thinking “thinking_r” (the subscript *r* stands for “rationalist”). On the rationalist interpretation, the *cogito* expresses the following proposition:

- (C_r) ‘I am thinking_r’.

(C_r) is indeed *self-verifying* for me in that it is a priori that I cannot think_r it falsely:

1. If I am thinking_r (C_r) (either by believing it, doubting it, etc.) then I am thinking_r (by definition of thinking_r).
2. If I am thinking_r then (C_r) is true (by one conditional, T-in of the T-schema).

3. If I am thinking (C_r) then (C_r) is true (by (1) and (2)).

This self-verifying character stems from the fact that “think_r” is a self-reflexive term: In virtue of its meaning, thoughts_r about thought_r are (partly) about themselves and someone who thinks_r, that he is thinking_r will be thinking_r about his very act of thinking_r.

As it is a priori that no one can think_r without existing, (C_r)’s twin proposition

– (C'_r) ‘I exist’

will be self-verifying as well. Shoemaker (1996, 53) nicely summarizes the point:

“I think” is indubitable for a logical reason; it is a logically necessary condition of my being deceived about anything that I think, since being deceived is a matter of having false beliefs, which in turn is a special case of thinking - in the sense of “think” in question. “I think,” like “I exist,” is necessarily self-verifying, in the sense that it is a necessary condition of its being asserted, or even entertained in thought, that it be true.

Burge (1996, 93) likewise claims that (C_r) and (C'_r) are “self-evidently self-verifying” in that “it does seem that understanding [them] suffices for knowing [that they] are true. And the relevant understanding requires no great perspicacity” (see Ayer (1953, 29) for similar claims). This is basically the ground for the first rationalist interpretation of the cogito. The details, however, are much more complicated.

If a thought is self-verifying, doubting it makes it true. This is not, however, enough for certainty. For Descartes, the thought ‘Descartes thinks and exists’ is just as self-verifying as ‘I

think and exist', yet Descartes could and actually did doubt 'Descartes thinks and exists'. In the same way, when the inventor of the cogito argument thinks 'the inventor of the cogito argument thinks and exists' his thought is just as self-verifying as 'I think and exist,' and for the same reason. Yet for all he knows, someone in the past might have discovered the cogito argument first, someone long dead.

Unlike 'Descartes thinks,' 'I think_r' is in the first-person, so it is, so to speak, universally self-verifying: it is *a priori* that if *anyone* thinks it (and in particular if I think it) it is true. Despite common contentions to the contrary,⁹ however, this is still not enough to make 'I think_r' certain. C_r 's universal self-verification just entails that *it is a priori that if I doubt it, then it is true*. It is not obvious, however, that by itself, this should make it impossible for me to doubt it. I can after all doubt true things, so why couldn't I doubt things that are a priori made true by my doubting them? If we want to derive certainty from universal self-verification we must do two things: (i) we must construe the certainty that P as the impossibility to *rationaly believe* that P is false, rather than the impossibility to *conceive* that P is false (more on the difference below); (ii) more importantly, we must appeal to the fact that Moorean propositions of the form 'P but I do not believe that P' cannot be rationally believed. By (ii), if I believed that C_r is false, were rational and considered the matter, I would believe that (C_r is false and (I believe that C_r is false)). By *universal self-verification*, applied to the second conjunct of the proposition, I would

⁹ Ayer (1953, 29) and Shoemaker (1996, 53) are examples.

accordingly believe that (C_r is false and C_r is true) which contradicts the hypothesis that I am rational.¹⁰ This shows that I cannot rationally believe that C_r is false, and by (i), that C_r is certain.

On this interpretation, then,

(R¹i) cogito-like thoughts¹¹ constitute a subclass of thoughts_r.

(R¹ii) The cogito is certain because those thoughts_r contain the concept ‘thought_r’ and are to that extent self-reflexive.

(R¹iii) We acknowledge this certainty because

- we have a normal *procedural rationality*, and can accordingly follow normal patterns of deductive reasoning,
- we understand the conventional meaning of “thought_r” and of the first-person pronoun “I,”
- we can tell that Moorean propositions are not rationally believable, a capacity we might call “*Moorean sensitivity*”.

¹⁰ One might also derive the certainty of C_r from universal self-verification and the claim that I cannot doubt a proposition without being certain that I do (If I doubted C_r , I would then be certain that I do, and by universal self-verification I would be certain that C_r is true, which contradicts the initial assumption). However the claim that I cannot doubt a proposition without being certain that I do seems to be too close to the claim that C_r is certain to really illuminate it.

¹¹ Thoughts that we can be certain to have and that give rise to the certainty that we think and exist.

These conditions explicitly mention our capacity to use thoughts (including thoughts about thoughts) for reasoning. It is sometimes said that thoughts which are poised for free use in reasoning are conscious in the sense of being *cognitively accessible* (Block, 2004). It is interesting to note, however, that the above conditions do not explicitly mention consciousness in the sense associated with *phenomenal character*, i.e. *phenomenal consciousness*. At least *prima facie*, we can, for example, imagine a good thinking_r robot that would satisfy the above conditions and accordingly be certain that it thinks_r and exists even though it has no phenomenal consciousness.

1.3 A second rationalist interpretation of the cogito: rationality and the first-person

There are reasons to be dissatisfied with this interpretation of the cogito. First, while it explains why the cogito is immune to rational disbelief, the cogito also seems certain in a stronger sense. It is not only impossible to rationally *believe* that it is false; it is also impossible to *conceive* that it is false. Yet immunity to rational disbelief and inconceivability seem different: I can conceive that I believe falsely that it is raining, but I cannot rationally believe that it is the case. Moreover, since Moorean sensitivity forbids one to *believe* that C_r is false without realizing that one does, but not to *conceive* that C_r is false without realizing that one does, the first rationalist interpretation cannot explain why C_r is inconceivably false.

Second, this interpretation does not appeal only to the fact that ‘I think_r’ is self-reflexive and accordingly self-verifying, but also to the fact that it is in the first-person. It appeals, ultimately,

to Moore's paradox, another phenomenon related to the first-person. One might wonder if the appeal to the first-person does not make superfluous the considerations concerning the concept 'thought_r'. Do I really need to explicitly think about thinking_r in order to know for certain that I exist?

It has been suggested by many that the cogito might derive from the semantics of the first-person (cf. Williams (1978, 76-77), Anscombe (1975), Evans (1983, 251-3) as well as Kaplan (1989, 495,508-9), Lewis (1970, 186) and Zemach (1985)). The least controversial way to explain the indubitability of 'I think_r' in semantic terms appeals to the claim that 'I' being token-reflexive, it obeys *a priori* to the following reflexive rule:

(RR) 'I'-thoughts refer to their very thinker_r.

RR entails that uses of "I" in language refer to the thinker_r of the corresponding 'I'-thought. If, moreover, RR is *a priori*, then anyone who understands "I" should know this last claim. The *a priori* character of RR would render 'I exist' *a priori* equivalent to 'the thinker_r of that very thought_r exists' and 'I think_r' to 'the thinker_r of that very thought_r thinks_r', thoughts which are both (assuming that every thought_r has a thinker) *a priori* true and inconceivably false. An interpretation of the cogito along these lines is endorsed by those, like Kaplan (1989, 495,508-9) and Lewis (1970, 186), who take the cogito to derive its certainty merely from "the logic of

indexicals¹²” (we shall see in the appendix that Burge (2000) also espouses a variant of this interpretation). It implies that the certainty of the cogito is comparable to that of ‘I am here now’.

Even if it is stronger and more encompassing than the preceding interpretation, this one shares a few features with it. According to both interpretations, cogito-like thoughts_r are self-reflexive (either because they contain ‘thought_r’ or because they are ‘I’-thoughts). According to both interpretations, the cogito appeals to the subject’s procedural rationality and understanding of some self-reflexive terms. Indeed, on the second interpretation,

(R²i) cogito-like thoughts constitute a subclass of thoughts_r.

(R²ii) The cogito is certain because these thoughts_r contain the ‘I’-concept and thus are self-reflexive.

(R¹iii) We acknowledge this certainty because

- we have a normal *procedural rationality*, and can accordingly follow normal patterns of deductive reasoning,
- we understand the conventional meaning of first-person pronoun “I”.

Finally, like the preceding one, this interpretation appeals to the fact that some of our thoughts are cognitively accessible, but does *not* appeal, at least not explicitly, to *phenomenal* consciousness.

In order to condense both rationalist interpretations, let us say that

¹² Because Kaplan (1989)’s framework is for expression types rather than sentence uses, however, it allows for contexts with no sentence-use and no thought, and it does not count ‘I think_r’ (unlike ‘I exist’) as logically true (as true at every context in every “LD structure”).

Minimal Rationality. Someone is *minimally rational* if (i) she understands the meaning of “I”, of “thought_r”, and of the other words she uses, and (ii) her procedural rationality and Moorean sensitivity are not much worse than ours.

Such rationality is minimal because procedural rationality is just an aspect of rationality as a whole, along with epistemic rationality and practical rationality (Bermudez, 2001; Bortolotti, 2010). While procedural rationality is concerned with reasoning according to the laws of deductive logic (and probability theory), epistemic rationality and practical rationality are concerned respectively with maintaining an optimal relationship between beliefs and evidence and between beliefs and actions. According to both rationalist interpretations, someone who is minimally rational should be unable to doubt his existence.¹³

2 Two predictions

The empiricist interpretation and the two rationalist ones clearly seem different. The former appeals to phenomenal properties of cogito-like thoughts, the latter to cognitive-accessibility features. However, we should not exclude the possibility that these might actually be very close or even identical to each other. Some conceptions of phenomenal (and subjective) character might reconcile the two kinds of interpretation. I am thinking in particular of the so-called monitoring theories of phenomenal consciousness, and more specifically of the self-

¹³ Notice that even if rationalists, like Burge and Shoemaker, often appeal to a much more substantial notion of rationality, they seem to grant that the kind of rationality required for the cogito is really thin (Burge says for example that the cogito “requires *no great perspicacity*” and that it is “*self-evidently* self-verifying (Burge, 1996, 93)”, Shoemaker says that it is certain “for a logical reason (Shoemaker, 1996, 53)”).

representational approaches (Kriegel and Williford, 2006). According to these, a state is phenomenally conscious just when it is subjective, and it is subjective when it displays a certain form of self-reflexivity. They might thus claim that the certainty we have about our experiences according to the empiricist interpretation derives from the same form of self-reflexivity that yields the certainty that we think and exist according to one of the rationalist interpretations.¹⁴

Moreover, even if both interpretations were different, they might be equally correct. It might be, for example, that the rationalist interpretations explain why some people are certain that they think and exist while the empiricist interpretation explains why some other people are thus certain. Alternatively, we might all be, so to speak, doubly certain that we think and exist, that is, certain for both empiricist and rationalist reasons.

It would be wrong to think that there is no other way to assess those questions but through *a priori* reflections and introspection. Those interpretations make *empirical predictions* about the kind of people who might be *prima facie* uncertain whether they think and exist.¹⁵ The rationalist interpretations predict that

- (P_r) If someone is *prima facie* uncertain whether he thinks or exists, he is not minimally rational.¹⁶

¹⁴ Kriegel (2004b) for example appeals to his self-representationalist theory of consciousness in order to account for Moore's paradox.

¹⁵ Notice that someone might fail to be certain that he thinks and exists without being capable of having doubts about those matter and without, accordingly, being uncertain whether he thinks and exists (think about people in a comma who do not think at all). However neither the empiricist nor the rationalist interpretations make any prediction about such a person.

¹⁶ Burge (1988, 92) seems to endorse something like this prediction when, commenting on the first rationalist interpretation, he says that "One could be so far gone as to think to oneself: 'I do not know whether I am now thinking or not; maybe I am dead or unconscious; my mantra

The empiricist interpretation, by contrast, predicts that

- (P_e) If someone is *prima facie* uncertain whether he thinks or exists, the mental states he is reflexively aware of have an abnormal subjective or phenomenal character.

Suppose that the predictions of the empiricist interpretation are confirmed but those of the rationalist ones are not. We would then have *empirical* reasons to say that (i) the two kinds of interpretations differ, (ii) that they are not equally good, and (iii) that the empiricist one is better than the rationalist ones. In §§3-4, I will argue that this is precisely the case.

I know only of two kinds of people who could be uncertain whether they think and exist, and thus falsify one of the two predictions: pathological self-doubters, who suffer from either from acute forms of depersonalization or of the Cotard syndrome, and philosophers.¹⁷ Some philosophers, as we have seen, claim to be uncertain whether they think and exist. I take it however that they are *prima facie* certain about those things. What they put into question is rather that such a *prima facie* certainty can withstand ideal scrutiny. Pathological self-doubters constitute a more interesting case, as they are uncertain whether they exist for much less sophisticated reasons. They seem to be *prima facie* uncertain about it. I will argue that they vindicate the empiricist interpretation.

may have finally made me blissfully free of thought'. Such mistaken doubt would evince *cognitive pathology* (my emphasis)".

¹⁷ There are reasons to think that the Cotard syndrome was not totally unfamiliar to Descartes. He mentions some delusions in the first meditation (§4), like the belief that one's body is made of glass, that were attributed by his contemporary physicians to melancholy (Speak, 1990, the Cotard syndrome is still often construed as a form of delusional depression today) and that Cotard (1891) reported among his original patients. Interestingly enough, Descartes explicitly excludes the existence of such delusions as sensible grounds for doubting, invoking instead on the fallibility of the senses, the Dream hypothesis, and the Evil Genius hypotheses.

I must stress, however, that my argument is *strictly abductive*. I will not show that the data are *logically* or even *nomologically* inconsistent with the rationalist interpretations. My claim is only that the empiricist interpretation explains the data gathered so far *better*. This is what is meant by the claim that the data confirms the empiricist predictions but not the rationalist ones. It does not entail, again, that further data might never invalidate my conclusions.

3 Pathological self-doubt

As emphasized by Sierra (2009) in the most complete monograph on the topic, the term “depersonalization” is now used to denote a whole spectrum of abnormal experiences, ranging from short-lived impressions of emotional numbing and unreality (“I felt as if I was in a dream”) that can be induced by sleep deprivation or recreational drugs, to long lasting, deep and pervading modifications of the way things seem to the subject. In those acute forms of depersonalization, it can seem to the patients

- as if they missed bodily parts (“I can sit looking at my foot or my hand and not feel like they are mine” (Sierra, 2009, 27)),
- as if they missed the capacity
 - to have emotions or “feel things” (“I seem to be walking about in a world I recognize but don’t feel (Sierra, 2009, 32).” “Kissing my husband is like kissing a table, mister. The same thing. . . . Not the least thrill. Nothing on earth can thrill me. . . . My heart doesn’t beat. I cannot feel anything (Dugas and Moutier, 1911,

109)” “I can feel numb of feelings, almost empty inside. I hate the fact I can’t feel things as I used to (Sierra, 2009, 27).”)

- or even to think at all (“I have the feeling of not having any thoughts at all” (Sierra and Berrios, 2000, 161) (see also Sierra, 2009, 37))
- as if they were dead (“I don’t feel alive in any way whatsoever (Sierra, 2009, 8,29)” “a state of nothingness, no mood at all, as if I were dead (Simeon and Abugel, 2006, 30)”)
- or as if they themselves did not exist (“she feels like she just does not exist (Simeon and Abugel, 2006, 8)” “I have stopped being (Mayer-Gross, 2011, 106)” “I doubted of my own existence and at times even disbelieved in it (Krishaber, 1873)” “often I have to . . . enter a shop to talk, to ask for something, in order to get a new proof that I am really myself (Séglas and Meige, 1895, 131)”)

This last complaint, it should be noted, is not rare among patients suffering from severe depersonalization. It is characteristic of patients whose symptoms are deemed clinically significant according to the Cambridge Depersonalization Scale (the most detailed and valid measure of depersonalization experiences so far) (Simeon et al., 2008).

There is an important connection between acute forms of depersonalization and an uncommon psychiatric condition known as ‘Cotard syndrome.’ The Cotard syndrome is often said to be the delusional form of depersonalization, depersonalization being, conversely, the ‘as if’ form of the Cotard syndrome. The central feature of the Cotard syndrome is indeed a cluster of nihilistic delusions, which can include the patient’s denial of his own existence (see Enoch and Ball (2001); Young and Leafhead (1996); Debruyne et al. (2009) for recent review articles).

Patients suffering from the Cotard syndrome typically deny having certain bodily parts (“I used to have a heart. I have something which beats in its place . . . I have no stomach, I never feel hungry (Enoch and Ball, 2001, 167),”), deny having thoughts or emotions (“I have no feelings,” “I do not think”). In the most severe forms, they deny being alive (“I am dead” “I am a living dead” “I am a machine”), and can even negate their own existence (“No I do not exist”, “You know that we have drowned. It is not me who is talking to you. You are talking to yourself in me. I am not anything anymore. I do not exist,” say for example two patients from Camuset (Cotard et al. (1998, 162), here and elsewhere translations from French and Italian are mine). It should be noted that even if not all patients suffering from the Cotard syndrome deny that they exist, most of them do (69% according to Berrios and Luque (1995)’s survey of one hundred cases).

The classical distinction between the Cotard syndrome and depersonalization should not hide some important connections between the two. First, Cotard patients often display phases of (non-delusional) depersonalization. Given this and the similarity between the complaints of both kinds of patient, it is very common to suppose that they involve at least similar (if not identical) experiences.¹⁸ More importantly, *severe forms* of both conditions involve at least *doubts* concerning one’s existence; they involve, that is, *pathological self-doubt*. Depersonalization was often called “folie du doute (doubting madness)” or “practical Pyrrhonism¹⁹” when it was discovered (Ribot, 1896, 366-7) to reflect the patients’ ability to doubt everything, including their

¹⁸ The suggestion was already made before the term “depersonalization” was coined, by Cotard (1891) and Ségla and Meige (1895, 665). See also Janet (1903b, 353), Young and Leafhead (1996) and the references in Sierra (2009, 78).

¹⁹ Pyrrho and his heirs professed that we should doubt every claim and hold back from belief.

own existence: “He knows that he exists, that he lives and nevertheless it is as if he was dead. It is a perpetual doubt over . . . his own reality”. He “only doubts whether he is, . . . he knows it but he does not feel it (Dugas and Moutier, 1911, 11).”²⁰ There are thus two kinds of patients who can doubt that they think and exist: “depersonalized self-doubters”, who are not delusional and do not *believe* that (but only *feel as though*) they do not think and exist, and “Cotard self-doubters”, who are delusional and sometimes *believe* that they do not think and exist.

The rationalist interpretations predict that both kinds of self-doubter should fail to be minimally rational: they should either have an abnormal procedural rationality, fail to understand the meaning of the words they use or have an abnormal Moorean sensitivity. The empiricist interpretation predicts that the mental states of which they are reflexively aware should have an abnormal subjective or phenomenal character. In what follows, I will argue that the predictions of the empiricist interpretation are verified but those of the rationalist interpretations are not (§ 4). I will then argue that patients’ testimonies can help us see what is wrong with the rationalist interpretations (§ 5).

²⁰ Those doubts about one’s existence actually seem to follow from the feeling of non-existence: If it seems to someone as if not P, it is arguably at least conceivable for him that not P, that is, he can doubt that P.

4 Rationality and experience among self-doubters

4.1 Minimal rationality

Non-delusional self-doubters are minimally rational.

Even though the border between delusional and non-delusional self-doubt might not always be totally clear-cut (the confidence of patients suffering from the Cotard syndrome might fluctuate, especially during remission phases; patients diagnosed with depersonalization at times seem to express the *belief* that they do not think nor exist²¹), some self-doubters are clearly non-delusional. Confronted with such patients for the first time, psychologists of the turn of the nineteenth century were baffled to discover that their rationality seemed totally untouched. In the first monograph on depersonalization, Dugas and Moutier (1911) devoted a whole chapter and most of their introduction to the topic. Österreich was amazed by patients whose “critical rationality [*faculté de critique et de jugement*] is absolutely intact”. That faculty actually manifested itself in the way patients struggled not to take the way things seem to them at face value: “I had lost the notion of my own existence. . . . Even though this idea imposed itself upon me, I never forgot that it was illusory (Krishaber, 1873, 151)” . . . “it’s a constant struggle between my involuntary impressions and my judgment (Krishaber, 1873, 151)”. Non-delusional self-doubters not only seem minimally rational, they also seem to have a normal epistemic rationality (a normal capacity to correctly fit beliefs to evidence) and a normal practical

²¹ There is also a terminological problem here. Along with the definition of depersonalization given by the DSM-IV-TR, we tend to restrict the term depersonalization for non-delusional patients. This usage is not however universal and it was not for example that of Janet.

rationality. This probably explains why the rationality of non-delusional self-doubters was hardly ever assessed experimentally. I know only of three studies investigating the cognitive functioning of patients suffering from depersonalization, one of which could not differentiate between them and matched anxious or depressed patients (Sedman, 1972). The two others agree that patients suffering from depersonalization disorder and normal subjects show similar general intelligence, memory and attention (Guralnik et al., 2000, 2007). Patients' executive functioning is also perfectly intact (Simeon and Abugel, 2006, 99). These two studies only found a subtle difference in some very specific aspects of low-level attention and memory implied in tasks involving great perceptual overload – nothing that hints at a lack of minimal rationality (or even, for that matter, at a deficit of epistemic or practical rationality). We should accordingly assume that non-delusional self-doubters are minimally rational.

Delusional self-doubters—clinical impressions.

What about delusional self-doubters? Obviously, their belief is very awkward and it seems that they should not endorse it. Here again, however, a striking feature is that apart from that, the patients often seem rational and I will later argue that they are at least minimally rational. First, their delusions involve a highly determinate topic. Unlike, for example, persecutory delusions, the Cotard syndrome is *monothematic*. This means that outside the area of that specific theme, their cognitive functioning can be perfectly typical. Second, they are usually well aware that their delusional beliefs are in conflict with their background knowledge and can acknowledge their implausibility (“it is weird, I know it, I do not understand it myself (Janet, 1908, 515)”). Third, patients do not generally exhibit

aberrant epistemic practices. When asked, they can do their best to justify their contention in the face of contradictory evidence. Some even spontaneously engage in philosophical discussions about their non-existence (De Martis, 1956, 497-8). Those three points are nicely witnessed, for example, by Janet's patient Lætitia:²²

When I reflect on my situation I cannot understand it at all. Either I am alive or I am not alive, but I cannot be dead as my heart beats, but I am not alive as my person has disappeared . . . I feel like not being someone and nevertheless I talk, am I stupid? I am a body without a soul . . . It's more complicated than the Holy Trinity (Janet, 1928, 43).

Delusional self-doubters are procedurally rational.

The clinical impression of rationality among delusional self-doubters has been partly confirmed by experimental studies. Neuropsychological findings in Cotard syndrome point towards face processing impairments but like the many case reports gathered so far, they nowhere suggest an abnormal procedural rationality (ie. an abnormal capacity to reason according to the laws of logic and probability theory) (Swamy et al., 2007, 113). More generally the rationality of deluded patients has been empirically studied since the 1950's (Garety and Hemsley, 1997, Ch.3). The study of deluded patients in general (not specifically Cotard, usually patients suffering from schizophrenia) has revealed a tendency to "jump to conclusions" (they

²² Consider also the following patient of Angelo Hesnard, who was diagnosed with schizophrenia. He asked:

Am I thinking? Since there is nothing that can prove that I am thinking, I cannot know whether I exist (Parnas and Sass, 2001, 108).

make resolute judgments on the basis of little evidence). Some Cotard patients were moreover found to have an attributional bias (they tend to explain negative events in terms of internal rather than external causes²³ (McKay and Ciolotti, 2007; Young and Leafhead, 1996)). It is, however, unclear that such epistemic biases are *abnormally* severe (many studies actually suggest that they are no more severe than in non-delusional psychiatric patients). Moreover, they betray a form of epistemic irrationality (a suboptimal way to form and revise beliefs in the face of evidence), not a form of procedural irrationality. After more than 60 years, the search for abnormal patterns of procedural reasoning among patients has never led to convincing results (Kemp et al., 1997; Mirian et al., 2011). Deluded self-doubters might have biases that affect their epistemic rationality, and this might explain their difference with non-delusional self-doubters. They do not, however, exhibit the procedural irrationality that would prevent them from being minimally rational.

Delusional self-doubters understand the meaning of the words they use.

It is true that the patients' behavior is sometimes difficult to understand, as it sometimes seems not to fit very well with what they *say*. Campbell (2001, 91) has argued against the minimal rationality of Cotard patients on the ground their actions and epistemic practices seem to rule out their understanding of the conventional meaning of words. He notices for example that “a patient may claim to be dead despite being able to walk and talk” and “reali[zing] that no one would accept this claim.”

²³ Someone with such a bias will typically explain his car accident by his bad driving instead of blaming other drivers, the poor wether conditions, etc.

I believe that we can make sense of the patients' behavior without positing any deviant usage, or miscomprehension of the conventional meaning, of the words being used. We need to be careful, though, and distinguish different cases.

First, as noticed by Sass (2004) in response to Campbell, the words we use often do not have a *unique* and well-defined conventional meaning, and we might save the patients' rationality by appealing to some form of ambiguity. Patients' talk of "living dead" or "walking dead" (Young and Leafhead, 1996, 147) (Cotard, 1891, 311) (Vaxevanis and Vidalis, 2005) or even "zombie" (Sierra, 2009, 51) suggests that they might sometimes have a phenomenality-involving, rather than a merely biological, sense of "life" in mind. This sense is not deviant or uncommon, as witnessed by *our* usage of expressions such as "lived experience," "life after death," "zombie" etc.

That said, patients occasionally do deny that they are biologically alive as well. When their nihilistic delusions reach their climax and they deny existing, patients can deny having any trait whatsoever (Cotard 1891, 345). Moreover, their testimonies clearly show that they *sometimes* mean they are dead in the biological sense as well.²⁴ They can, for example, adduce the inactivity of their heart or of their brain (see among many, (Janet, 1928, 43) and Young and Leafhead

²⁴ It might be suggested that just like there is a phenomenality-involving sense of "life" and "thought", there is a phenomenality-involving sense of "exist", says 'exist_e', and that self-doubters do not really doubt that they exist, in the ordinary sense of the term but only that they exist_e. Just like in the case of life, that might be true of some patients, but many make it clear that they do not merely doubt to be phenomenally conscious when they doubt to exist. Moreover, one can acknowledge a close connection between existence and phenomenality without invoking a phenomenality-involving sense of phenomenality. I indeed believe that there is such a close connection but I will argue that it is one of justification—we know with certainty that we exist because we are phenomenally conscious— rather than one of partial synonymy. I thank XXX for pressing me on that point.

(1996, 158)). There is, however, evidence that when this is the case, their actions can be aligned with their beliefs and manifest their comprehension of what they say. For example, patients who claim to be dead often urge that they should be buried and refuse to eat or defecate (Cotard, 1891, 327,331,334). Lætitia often closed discussions with an abrupt “why do you want to talk, you do not exist and I do not exist, goodbye (Janet, 1928, 9)”. Others simply remain passive and apathetic (more on the interpretation of the patients’ speech in §5).

There is, more generally, a very simple argument that one cannot explain their delusion in terms of a lack of understanding of the conventional meaning of words. Patients can have some ‘symptom-free’ remission phases. If the delusion stemmed from a misunderstanding of the words used by the patients, during remission phases they should not only acknowledge that they live and exist, but also retract their past utterances and acknowledge that they have been living and existing all along. It is a common clinical observation, however, that during those phases, or even after recovery, some patients are not deluded about their present state but do not retract their past claims either. Young and Leafhead (1996, 147), for example, quote a patient who even said *after recovery* that she had been a living dead.

We can conclude that all self-doubters are at least minimally rational (and that non-delusional self-doubters have normal epistemic and practical rationality as well). The prediction of the rationalist interpretation is thus disconfirmed. Again, it should be emphasized that this does not imply that what they say is not very weird or faulty (biased). Merely weird or faulty reasonings are not enough, however, to confirm the prediction of the rationalist interpretation.

4.2 Abnormal experiences

There is, however, good reason to consider self-doubters' experiences abnormal in just the way predicted by the empiricist interpretation – and that this explains their doubts. When they are asked to justify their doubts, patients frequently invoke their experiences. More precisely, they adduce the absence of certain phenomenal features or “feelings” they expect their thoughts to have. As reported by Young and Leafhead (1996, 149), “what the [Cotard] patients often give as evidence of their non-existence or death is that they don't have proper feelings”. I will argue that the change in “feelings” invoked involves at least an attenuation of the subjective character of the mental states of which they are reflexively aware. Even though it is more controversial – and not, strictly speaking, necessary for the confirmation of the empiricist prediction (P_2) (‘the mental states of which the patients are reflexively aware have an abnormal subjective or phenomenal abnormal’) – I will also argue that the relevant change in feelings might involve, more globally, an attenuation of the whole phenomenal character of the mental states of which they are reflexively aware. In other words, not only the patients do not feel these mental states as theirs (attenuated subjective character), they might also not feel them at all (attenuated phenomenal character).

Attenuated subjective character.

Self-doubters typically complain that something is missing in their experience and that, as a result, their mental states do not appear to them to be theirs. This missing feature can be characterized as a “feeling of ‘I’,” “the feeling of myself (Janet and Raymond, 1898, 73),” “the awareness of myself (Krishaber, 1873, 171; Janet, 1903a, 324)” or even the “experience of ‘me’

(Simeon and Abugel, 2006, 143).” Given those characterizations of what they are missing, and given that the subjective character is the aspect of my conscious states in virtue of which they seem mine to me, the patients’ testimonies strongly suggest that the subjective character of the mental states of which they are reflexively aware has vanished. Interestingly, we could also account for self-doubters’ reports with a more modest hypothesis. In the milder forms of depersonalization, patients may say that their mental states do not *clearly* seem to them to be theirs rather than that they do not seem to be theirs.²⁵ Likewise, they may say that they have no “*clear* feeling of ‘I’” rather than no feeling of ‘I’ (Simeon and Abugel, 2006, 25). This suggests that the subjective character admits degrees, and that its intensity decreases with the severity of depersonalization. When it is slightly diminished it would appear as not clearly present rather than as absent. Moreover, if the subjective character were importantly diminished, it is arguable that it would sufficiently differ from the kind of phenomenal feature that normally classify as mineness or “feeling of ‘I’” not to be recognizable as such, and that the patients could be minimally rational judge that they have no feeling of ‘I’ at all and that their states do not seem to be theirs at all. We could accordingly minimize the differences between the patients and us by claiming that their subjective character is (substantially) diminished rather than totally absent. In what follows, I will talk of an attenuation of the subjective character, leaving open whether this attenuation is only partial or whether it totally extinguishes the subjective character.

The attenuation of subjective character of depersonalized and Cotard patients cannot only vary in intensity. Depending on the severity of their condition, it can also vary in extension. Self-

doubt arises when the attenuation is substantial enough and concerns all the states of which the patients are reflexively aware. The subjective character can recede, first of all, from the patients' *bodily sensations*, explaining why, as we have already seen, some parts of their body can seem alien to them. It can also recede from algedonic states of pleasure or pain. Patients suffering from severe depersonalization typically assent to the following description (taken from the Cambridge Depersonalization Scale) : "When a part of my body hurts, I feel so detached from the pain that it feels as if it were somebody else's pain (Sierra and Berrios, 2000)." After having been exposed to a noxious stimulus, a patient of Janet's withdrew her hand and explained:

It was painful and my arm felt like withdrawing, but it was not a genuine pain, it was a pain that did not reach the soul... It is a pain, if you want, but the surface of my skin is miles away from my brain, and I do not know whether I am suffering (Janet, 1928, 65).

An attenuation of the subjective character of intentions-in-action explains, likewise, why patients often lack a feeling of agency for what they do:

It is not me who acts, I see myself acting . . . I am a puppet . . . I am myself surprised by the precision of the automaton (Janet, 1908, 515)).

I would notice my hands and feet moving, but it is as if they did not belong to me and were moving automatically (Sierra, 2009, 29).

An attenuation of the subjective character of thoughts, more broadly, can explain why, like in thought-insertion, those feel alien:²⁶

²⁶ The overlap between the phenomenology of depersonalization and the symptoms of schizophrenia has already been noted (Sass et al., 2013). Notice however that while patients

Thoughts running through his brain again seemed somehow foreign. (Simeon and Abugel, 2006, 26).

I feel so detached from my thoughts that they seem to have a ‘life’ of their own. (Sierra and Berrios, 2000, 163).

In extreme cases of Cotard syndrome and depersonalization, the subjective character seems to recede from all mental states of which the patient is aware, leaving him with the feeling that he owns no thought at all and that he does not exist. Ramirez-Bermudez et al. (2010, 411) report a Cotard patient who “expressed the belief that she did not exist because she could not feel herself”. Debruyne et al. (2009, 197) report another who “had the constant experience of having no identity or ‘self’ and being only a body without content.” Likewise some depersonalized patients describe their experience as follows:

There was literally no more experience of ‘me’ at all. The experience of personal identity switched off and was never to appear again . . . The body, mind, speech, thoughts, and emotions were all empty; *they had no ownership*, no person behind them (Simeon and Abugel, 2006, 143-4, emphasis mine).

suffering from thought-insertion believe that some of their thoughts are alien in the strong sense of *belonging to someone else*, self-doubters usually feel as if their thoughts were alien in the weaker sense of *not belonging to themselves* (they can sometimes suspect that they are someone else’s thoughts, but this is not the rule). It is not clear however that this difference stems from a difference in experience rather than a difference in cognitive biases. Moreover, while in thought insertion the feeling of alienation can concern a few thoughts of the patients only, the feeling of alienation of self-doubters seems to affect all the mental states of which they are aware (see Billon (2013a) for an interpretation of thought insertion in terms of abnormal subjective character).

It is the mental sensibility²⁷ that is lacking, it is not me who feels. I have no interest in what I appear to be feeling. It is someone else who feels mechanically (Janet, 1908, 515).

The functions and acts of ordinary life, it is true, still remain to me; but in every one of them there is something lacking. That is, the sensation which is proper to them... Each of my senses, each part of my proper self is as if it were *separated from me* and can no longer afford me any sensation (Sierra, 2009, 8, emphasis mine).

The claim that self-doubters suffer from an attenuation of subjective character, it should be emphasized, is hardly revolutionary. An attenuation of subjective character is arguably the phenomenological common denominator of almost all accounts and descriptions of depersonalization (and accordingly of all accounts of the Cotard syndrome that take it to involve depersonalization-like experiences). What those accounts disagree on is mainly the *explanation* of this attenuation of subjective character.²⁸ They also disagree on whether depersonalization involves other symptoms that are irreducible to this attenuation of subjective character (see Sierra, 2009, 24-27). The attenuation itself is not in dispute. In his seminal paper, Dugas (1898) thus characterized depersonalization as an extinction of the feeling of self (*sentiment du moi*) that normally accompanies our mental states. He would later talk of a diminution of the “self-coefficient” of those mental states (Dugas and Moutier, 1911, 13,35). Jaspers (1997, 121) even *defined* “depersonalization” as a lack of subjective character:

²⁷ “Sensibilité morale”: in an empirical framework, the capacity responsible for thoughts and feelings in general as opposed to mere sensations.

²⁸ Is this attenuation caused by perceptive, interoceptive, emotional, agentic, memorial or metacognitive disorders? See Sierra (2009, Ch. I-II, X-XI) for an overview of the answers to those question (see also Dugas and Moutier (1911, 1-135) and Janet (1928, 35-88)).

Every psychic manifestation, whether perception, bodily sensation, memory, idea, thought or feeling carries *this particular aspect of 'being mine'* of having an 'I' quality, of 'personally belonging', of it being one's own doing. This has been termed *personalization*. If these psychic manifestations occur with the awareness of not being mine . . . we term them phenomena of depersonalization.

Other accounts or descriptions of depersonalization seem to at least imply an attenuation of subjective character. The DSM-IV-TR, for example, mentions "feelings of being detached from one's mental processes or body." The ICD-10 requires that the "the individual feels that his or her feelings and/or experiences are detached, distant, etc."²⁹

Attenuated phenomenal character.

As noted, the attenuation of subjective character described above is already enough to confirm the empiricist prediction. Yet I would like to argue that this is probably not the only modification in the phenomenal features of the relevant states. Many patients' testimonies suggest that their problem is not only that they do not feel the mental states of which they are aware *as theirs* (attenuated subjective character), but also that they do not feel them at all (attenuated phenomenal character) – that there is no phenomenality associated with them. Such attenuation of phenomenal character would provide a further confirmation of the empiricist interpretation. This is also independently interesting because many theories of consciousness claim that phenomenal character constitutively depends on subjective character and imply that an

²⁹ The DSM and the ICD are the most widely used classification systems for mental disorders.

attenuation of the latter should entail an attenuation of the former (see fn. 6 and Billon and Kriegel (in press)).

Some of the self-reports quoted above might already hint at a genuine extinction of phenomenal character (“my senses can no longer afford me any sensation,” “it is someone else who *feels mechanically*,” “it was not a genuine pain, it was a pain that did not reach the soul”). The hypothesis that many of their mental states have an extinct phenomenal character would also explain, more generally, why patients sometimes complain that they do not see, hear or feel things. Taken more or less at face value, those complaints led early clinicians to suppose that Cotard syndrome and depersonalization depend on a form of inner or peripheral anesthesia (an inhibition of interoception or perception more broadly).³⁰ Except for some anecdotal cases, no such anesthesia was ever experimentally confirmed. The patients who say that they do not see anything or that they are blind can spontaneously describe visual scenes and have normal discriminatory capacities. Likewise, objective measures proved that those who deny feeling their body have an intact bodily perception (see Cappon and Banks (1965), Cappon (1969) and Janet (1928, 40, 63-4)). There is something puzzling about the association of such a felt anesthesia with both an intact access to sensory information and an intact ability to use such information. The hypothesis of an extinct phenomenal character seems like a good way to solve this puzzle. The patients would feel anesthetized even though their sensory capacities are objectively intact because it seems to them that there is nothing it is like anymore to have such sensory capacities. Like in the case in the case of subjective character discussed above, the hypothesis of an

³⁰ Following Esquirol, Falret (1864, 287-8) for example thought that Cotard patients suffered from a general anesthesia. Krishaber (1873) argued that depersonalization depends on a numb “coenesthesia” (bodily perception).

extinction of phenomenal character might be slightly weakened. Patients suffering from milder forms of depersonalization can say that it is not clear to them that they feel things, rather than that they do not feel things. We might accordingly suppose that phenomenal character admits degrees and that an important decrease in its intensity might make it so different from what is usually classified as phenomenal character that self-doubters do not recognize it as such and complain of total lack of feelings.³¹ We could accordingly talk of attenuated rather than of extinct phenomenal character, leaving open whether this attenuation is partial or full.³²

Many testimonies of depersonalized and Cotard patients suggest that they can suffer from an attenuation of phenomenal character whose intensity and extension varies with the severity of their condition. The frequent reports of “emotional numbing” (“My heart does not feel anything anymore (Cotard, 1891, 320),” “The emotional part of my brain is dead (Mayer-Gross, 2011, 111)”) can for example be interpreted as expressing an attenuation in the phenomenal character of emotions. Other reports suggest that the phenomenal features of perception and imagery can

³¹ There are two ways to construe the attenuation of the phenomenal and subjective character a mental state. First, one might argue that the phenomenal or subjective character of each feature represented by a given mental state (the greenness, and the roundness of the apple I see, the bitterness of the beer I taste, etc.) admits degrees and that its decrease yields an attenuation of the subjective or phenomenal character of the mental state. Alternatively one might argue that the phenomenal or subjective character associated with a represented feature is an on-off property and that a mental state has an attenuated phenomenal or subjective character when some of the features it normally represents phenomenally and subjectively are represented non phenomenally or non subjectively.

³² Interestingly, the claim that self-doubt generally involves an attenuation rather than an extinction of phenomenal character allows one to argue that this attenuation is more intense among Cotard patients than among depersonalized patients. The difference between both conditions might accordingly not be, or not just be, a difference in epistemic rationality. This point is particularly relevant for the debate between the “one factor” and the “two factors” theorists of the Cotard syndrome (see for example Davies et al. (2001); Gerrans (2003); Young and De Pauw (2003)).

become blunted as well. Colors fade (Sierra, 2009, 35,43,159), volumes disappear (“at times it is like looking at a picture. It has no real depth (Sierra, 2009, 51)”), contrasts and brightness recede (patients describe the world as appearing through a “curtain”, a “blind”, a “fine wire netting”, a “fine mesh” (Shorvon, 1946, 784)). In the more severe forms or phases of the disorders, the phenomenal features of vision seem to all substantially recede:

Everything in vision is dead (Mayer-Gross, 2011, 111)

I can see, hear and smell but it is as if I didn't see, nor hear or smell (Dugas and Moutier, 1911, 10)

My eyes see and my spirit perceives, but the sensation of what I see is completely absent (Sierra, 2009, 8)

I see without seeing, I am a blind man who sees (Janet, 1928, 52).

When depersonalization and Cotard syndrome reach their climaxes, the attenuation gets more widespread and more intense, and patients can describe themselves as unconscious, indeed as zombies:

I suddenly wonder: is it really me here? Is it really me walking? Then I make enormous efforts in order to apply my consciousness to this unconsciousness (sic). So that at one point during this sort of crisis, I am conscious on the one hand that I am unconscious on the other (Séglas and Meige, 1895, 147)

I'm like a zombie unable to take in any information (Sierra (2009, 51), emphasis mine).

Everything in me is mechanical and happens unconsciously (Ball, 1882, 43)

It [a state he calls “the black”] is a state in which you feel nothing, in which you do not think, in which you do not mean what you do or think . . . It an eternal curtain in front of the eyes, in front of the mind. I am in emptiness, I am a body without a soul (Janet, 1928, 51-2).

I just sink into a kind of unconsciousness. I am just conscious enough to know that things are going on around me but nothing seems to register (Shorvon, 1946, 784).

At that point the attenuation would be important enough to give rise to self-doubts.

It might be objected that if the phenomenal character of the states of which self-doubters are reflexively aware were extinct or substantially attenuated, and if they consistently denied having a phenomenal character, they should deny having states that are defined in terms of phenomenal character, such as pains. They should also tend to confuse states that we typically distinguish by their phenomenal character such as pains and itches. I agree with the first prediction, but I take it to be confirmed. Some patients suffering from milder forms of depersonalization or Cotard syndrome can acknowledge feeling pains. This is not the case of patients whose condition is so severe that they doubt that they think and exist. Just like they deny feeling anything, self-doubters typically deny feeling pains. Some Cotard and depersonalized patients explicitly say that they do not have pains anymore³³ (Sierra 2009, 26). Others say that they do not have *genuine* pains, but only pseudo-pains that do not really affect them and that they arguably take to lack the proper phenomenal

³³ “Patients complain that that they are capable of experiencing neither pain or pleasure (Sierra, 2009, 26).”

character (Sierra, 2009, 49, 150; see also the patient of Janet (1928, 65)'s quoted above, who explicitly says that her pains are not genuine). They can in fact wound themselves in order to feel genuine pains again (Janet, 1928, 65). Moreover, we do not distinguish states that normally differ by their phenomenal character, such as pains and itches, by their phenomenal character only; we also distinguish them by their intentional content, their functional and their behavioral role. Given that those seem largely preserved among patients (they can for example normally distinguish noxious from benign stimuli, they withdraw their arm by reflex when it is pricked...), we should not expect them to confuse pseudo-pains for pseudo-itches.

5 What is wrong with the rationalist interpretations of the cogito?

Many philosophers will remain unconvinced by the claim that pathological self-doubters are minimally rational. Whatever the independent evidence of their rationality, the mere fact that the patients doubt their own existence might be considered an acceptable demonstration that they are actually *not* minimally rational. The rationalist interpretations of the cogito indeed seem intuitively correct and it seems that the simple reasonings to which they appeal should compel the patients if they are rational. No matter what they mean by “thought”, “life”, etc., they should be certain that they at least think, and accordingly exist. In a recent book, Graham (2009, 241) expresses this worry under the form of a challenge (notice that he seems to rely on something

like the first rationalist interpretation of the cogito, which appeals to Moore's paradox, and that he only targets Cotard (as opposed to depersonalized) self-doubters):

It is . . . notoriously difficult to deny one's own existence or, for that matter, to assert that one is dead without being aware of the gross inconsistency of such speech acts with one's living or existing as speaker . . . How can I say something that is utterly incompatible with understanding that I am saying it?

Bermudez (2001, 479) has the same worry in mind when he says that patients' beliefs are "pragmatically self-defeating" and that "it is far from clear that [they] can be consistently expressed". Ratcliffe (2004, 30) concurs that if we take the patients' reports literally, "then there is not merely a weakening of rationality but reckless abandonment."³⁴ If we are to maintain that self-doubters have a normal procedural rationality, are Moorean sensitive and use their words comprehendingly, we have to show what is wrong with the rationalist interpretations of the cogito.

5.1 Problems with the first rationalist interpretation

The patients' reports actually suggest an answer to Graham's challenge. They indicate that the patients' self-awareness is impaired in that they are not (clearly) aware of being entitled to use the 'I'-concept. We shall see, however, that the first rationalist interpretation cannot show that subjects with such an impaired self-awareness should endorse the cogito if they are minimally rational.

³⁴ Ratcliffe talks about death rather than non-existence. He moreover grants, in later works, that the claim that one is dead might not be absurd after all (Ratcliffe, 2010).

Self-doubters' self-awareness is impaired.

We are always clearly aware of being 'I's, that is, of *being entitled to refer to ourselves using the 'I'-concept*. Connectedly we are also certain that we are 'I's. These features of our normal self-awareness seem to be lacking among self-doubters. Even if they have *some* self-awareness, their self-awareness seems to be impaired.

Self-doubters ordinarily use the 'I'-concept to refer to themselves. They are not, however, clearly aware of being entitled to do so: they seem to be using it more or less blindly, that is, without feeling genuinely justified to do so.³⁵ In that sense, they are not clearly aware of being 'I's. Some, we have seen, say that they lack a “feeling of 'I',” “an experience of ‘me’” or “an awareness of myself.” Others explain that they have a merely third-personal or observational access to themselves, the observer not being an 'I' either:

[It is like] seeing life as if it were played like a film in a movie. But in that case where am I? Who is watching the film? (Simeon and Abugel, 2006, 15)

I find myself regarding existence as though from beyond the tomb, from another world; . . . I am as it were, outside my own body and individuality; I am depersonalized³⁶ (Simeon and Abugel, 2006, 133).

I felt that my brain was somewhere else and from there was just watching me... I was completely unable to tell whether I was still present or whether I was the part

³⁵ Compare with someone making a visual judgement that he believes to be too finely grained for his powers of discrimination. Or with blindsights, who can make accurate visual judgements without feeling entitled to make them (Weiskrantz, 2009).

³⁶ This last quote is actually from Amiel's journal, on which Dugas (arguably) drew to coin the name “depersonalization”.

that was gone. In short, there were two different beings, the one watching the other (Roberts, 1960, 481).

This suggests that they feel more entitled to refer to themselves in the third-person than in the first-person. Some can even spontaneously use the third-person to talk about themselves (see for example the self-reports in Shorvon (1946, 784), Mayer-Gross (2011, 127) and Janet (1903a, 323)).

Among Cotard self-doubters, the failure to be aware of oneself as an 'I' seems to lead to the outright *belief* that one is not an 'I', entitled to use the 'I'-concept (rather than to mere *doubts* about it). Indeed, many Cotard self-doubters boldly refuse to use the first-person in any of its guises ('I', 'my', etc.). Some refer to themselves using "he" or "she", or even by their name (see for example the classical case reports of Falret (1864, 287-8) and Levassor and Dromard (1908)). A patient of Enoch and Ball (2001, 167) called herself 'Madam Zero' in order to emphasize her non-existence. Some resort to the impersonal "it" (Ey, 1950) or "this" ("this is empty", said one patient referring to herself (Séglas and Meige, 1895, 475)). Others use complex periphrases.³⁷ Many only use the first-person in past tensed phrases, referring to phases that precede the onset of the disorder (Séglas and Meige, 1895; De Martis, 1956). This refusal to use the first-person is not a whim. It expresses the patients' belief that they are not I's. Asked why he used to avoid the first-person, a patient of Janet's (1928, 43) explained:

³⁷ Cotard (1891, 316) describes a patient who systematically says "the person of myself does..." instead of "I do..." Despite the residual "myself," and given that the patient conjugates verbs in the third-person, this is arguably a case of avoidance of the first-person as well.

It is because I am not Myself [the patient emphasizes] sick, I am not Myself sad .
.. I am not Myself at all. What is missing is myself, it is awful to elude oneself, to
live and not to be oneself.

It is true that some Cotard self-doubters still use the first-person pronoun “I”, which might suggest that they still believe that they are ‘I’s. However, it is not obvious that they take their uses of “I” in language to express a genuine use of the concept ‘I’ in thought. Many patients, for example, stress that their application of the word “I” compares to those of a machine, applications that cannot be taken seriously—think of the parrot using “I” or, to take an example from Williams (1978), or the machine that says “I speak your weight”. A patient of Sierra’s (2009, 29) explains: “It’s as if a machine was talking to you. Not a person at all, just a mechanical thing or object (Sierra, 2009, 29).” Asked how he can talk if he is dead, a patient of Séglas and Meige’s (1895, 667) likewise answers: “it is true that I talk, walk and work, but it is like an automaton.” A patient of Ackner’s says “The part of me that is there talking is like part of a machine.” JK says: “I am just a voice and if that goes I won’t be anything... if my voice goes I will be lost and I won’t know where I have gone (Young and Leafhead, 1996, 157)”. Dr... (sic) says: “I hear myself talking, it is someone else talking, a machine talking instead of me (Janet, 1908).”

Moreover, many only use the first-person in *negative* phrases, such as “I am not myself” or “I don’t exist” (the Cotard delusion was named “negation delirium” by Cotard (1891, 309)). But in such phrases, the negation can be charitably interpreted as a metalinguistic negation in Horn (1989, VI)’s sense, whereby “I don’t exist” means something like “‘I exist’ is not true / felicitously assertable”, and “I am not myself” means something like “‘I am myself’ is not true /

feliculously assertable”. Under those interpretations, patients are in fact *mentioning* rather than *using* the ‘I’-concept.

The first rationalist interpretation presupposes an intact self-awareness.

Suppose that your self-awareness is impaired in that you are not clearly aware of being an ‘I’. Suppose that as a result you *doubt* (depersonalized self-doubter) or *disbelieve* (Cotard self-doubter) that you are an ‘I’, i.e. that you are entitled to use the ‘I’-concept. In such a case, you might understand

- (i) that tokens of “I” express the ‘I’-thoughts of the one who uses them and refer to him
(Understanding “I”),
- (ii) that a first-personal thought like ‘I think’ is accordingly universally self-verifying
(Universal Self-verification),
- (iii) that no one can rationally believe a Moorean proposition such as ‘P but I do not believe P’
(Moorean Sensitivity),

and thus satisfy the conditions of minimal rationality involved in the first rationalist interpretation, but still doubt / disbelieve that you can say “I think” truly.³⁸ The reason why is that (i-iii) are first-person principles: they concern people who are ‘I’s, who are, that is, entitled to think about themselves using the ‘I’-concept. If you doubt / disbelieve that you are an ‘I’, you *ipso facto* doubt / disbelieve that you are concerned by those first-person principles.

³⁸ Notice that in conditions (i-iii) the first-person concept is mentioned rather than used. Accordingly, in order to satisfy them one must only have a normal capacity to think *about* ‘I’-thoughts, not a normal capacity to think *with* them and not, accordingly, the normal self-awareness that underlies this last capacity.

Nevertheless, you might very well acknowledge (i-iii) and agree that someone whose self-awareness is intact, and who is certain that he is an 'I', should assert "I think" with absolute certainty.

If self-doubters can doubt / disbelieve that they think and exist, then, it is because they are not clearly aware of being 'I's. We can accordingly answer Graham's challenge against the minimal rationality of Cotard self-doubters as follows. The claim that a minimally rational subject cannot say something that is incompatible with understanding that he is saying it only holds for subjects who think about themselves in the first-person. Cotard self-doubters however do not think about themselves that way.

5.2 Problems with the second rationalist interpretation

At this point, a partisan of the second rationalist interpretation could object as follows:

- (i) A minimally rational subject should at least be certain that there is a thought in some sense (a thought_r) when he thinks,
- (ii) he should understand the reflexive rule RR ('I'-thoughts refer to their very thinker_r) and know that 'I' is cognitively equivalent to 'the thinker_r this very thought_r',
- (iii) he should accordingly be certain that he can think about himself as an 'I' ('the thinker_r of this very thought_r'),³⁹

³⁹ Which implies, by the way, that his self-awareness should not be impaired in the way I have claimed it is.

(iv) and he should likewise be certain of the truth of ‘I think_r’ (‘the thinker_r of this very thought_r thinks_r’).

It is not obvious what to say about (i). Many self-doubters seem to doubt that there are thoughts—they can doubt that there is anything at all—and it is not clear why minimal rationality should forbid this. But even if the truth of (i) were granted, (iii) and (iv) are problematic. They both assume that a minimally rational subject cannot doubt that ‘the thinker_r of this thought_r’ is a referring concept. This, in turn, presupposes that a minimally rational subject should be absolutely certain that the thoughts_r of which she is aware have a unique thinker_r. Such a natural assumption is endorsed by many researchers who believe, like Burge, (2000, 247), that “the idea of mental states or events without an *individual* subject is *incoherent* [emphasis mine]”. The patients’ reports suggest that this assumption is nevertheless ungrounded. We saw that some patients feel like their thoughts have no thinker (“my thoughts seem to have a ‘life’ of their own (Sierra and Berrios, 2000, 163)”). Likewise some feel like their thoughts have many thinkers at once.⁴⁰

I can’t seem to find my actual self. I feel as though my deliberations are those of a public body or corporation rather than those of a person. I used to say ‘we’ rather than ‘I’. It is as though I had transcended personality, as if ‘myself’ had receded to an image which I regarded objectively, and which is not identified with the whole of me (Shorvon, 1946, 784)

⁴⁰ Many also say that they feel double, which might imply that they do not feel like their thoughts_r have a unique thinker.

Given that they can feel like it is not true that their thoughts_r have a unique thinker_r, it seems that patients can doubt / disbelieve it without violating the norms of minimally rationality.

The claim that one can be minimally rational without being certain that thoughts_r always have a unique thinker is also independently plausible. As noticed by Anscombe (1975, 57) and Nozick (1981, 72), whose objections against the (second rationalist interpretation of the) cogito surprisingly resemble the above patient's quote, we *can* imagine thoughts_r with many contributing thinkers.⁴¹ 'The thinker_r of this thought_r thinks_r and exists' is hence no more certain than 'the maker of this object (in front of me) makes objects and exists' (I can imagine that this object has no maker at all or that it has many). Of course, it might be that in some circumstances, we are certain that a given thought_r has one and only one thinker. I take it, however, that those are the circumstances in which this thought_r has a normal (non-attenuated) subjective and phenomenal character.

To sum up. Pathological self-doubters fail to be certain that they think and exist because they are not clearly aware of being 'I's, that is, of being entitled to use the 'I'-concept. They are not,

41

How, even, could one justify the assumption, if it is an assumption, that there is just one thinking which is this thinking of this thought that I am thinking, just one thinker? How do I know that 'I' is not ten thinkers thinking in unison? Or perhaps not quite succeeding (Anscombe, 1975, 57).

Can't we imagine distinct centers of consciousness and self-consciousness, though? If each of the parts of my body, including neurons, that cooperated in producing the word token had its own center of consciousness and each was referring to itself, as it did its share, then there would not be any one entity, capable of self-referring, which produced the token (Nozick, 1981, 72).

in particular, clearly aware that their thoughts have a unique thinker and they can refer to themselves as 'the thinker of this thought'. This impairment of self-awareness explains their doubts. Both rationalist interpretations are wrong because they presuppose, and they cannot show, that minimal rationality is sufficient for normal self-awareness. A normal self-awareness requires, however, a normal subjective character, which does not depend on minimal rationality.⁴² Once their impaired self-awareness is taken into account, the patients' claim that they do not exist thus appears to be consistent with their minimal rationality.

6 Conclusion

As self-doubters are minimally rational but have abnormal experiences, the predictions of the empiricist interpretation of the cogito are confirmed but those of the rationalist interpretations are not. We thus have reason to think that the two kinds of interpretation differ, that they are not equally good (for different kinds of people), and that the empiricist interpretation is *better*. Both rationalist interpretations fail, because they presuppose that the subject clearly appears to himself as an 'I'. It is this normal self-awareness condition, for which they cannot account, that explains the certainty that we think and exist. Normal self-awareness depends, moreover, on the subjective character (or "mineness") of our experiences rather than on minimal rationality alone.

It should be emphasized again that this argument, couched in terms of confirmed and disconfirmed predictions, is strictly abductive. Its conclusions might be defeated by further investigation. Perhaps we will find out that depersonalized patients are all lying about their

⁴² I study self-awareness (and its impairment among self-doubters) in much more details elsewhere (Billon, 2013b).

experiences, display a subtle form of irrationality or have a mind whose functioning is so different from ours that no lesson can be drawn from their case. Perhaps. My claim, however, is that in the present state of science, the data fit very well with the empiricist interpretation, and quite poorly with the rationalist interpretations. Finally, the case of pathological self-doubters might not only allow us to defend the empiricist interpretation of the cogito but also to refine it. The study of the neurocognitive underpinnings of the Cotard syndrome and of depersonalization is making important progresses these days (see for example Sierra and David (2011) and Charland-Verville et al. (2013)). It promises to illuminate not only the certainty that we exist but also the nature of phenomenality, subjectivity and the kind of self-awareness underlying it.⁴³

References

- Anscombe, G. (1975). The first person. In Samuel, G., editor, *Mind and Language*. Oxford University Press.
- Author (xxxxa). Xxx. xxx, x(x) :x-x.
- Author (xxxxb). Xxx. xxx, x(x) :x-x.
- Author (XXXX). Yyyyyyyyy. XX YYY.
- Author and Other, A. (xxxx). Xxx. xxx, x(x) :x-x.
- Ayer, A. J. (1953). "cogito, ergo sum". *Analysis*, 14(2) :pp. 27-31.
- Ball, B. (1882). La folie du doute. *La Revue scientifique de la France et de l'étranger*. II :42-47.
- Bermudez, J. (2001). Normativity and rationality in delusional psychiatric disorders. *Mind & language*, 16(5) :493-457.
- Berrios, G. and Luque, R. (1995). Cotard's syndrome : analysis of 100 cases. *Acta Psychiatrica Scandinavica*, 91(3) :185-188.
- Billon, A. and Kriegel, U. (in press). Jaspers' dilemma: The Psychopathological Challenge to Subjectivity Theories of Consciousness. In Gennaro R., editor, *Disturbed Consciousness : New Essays on Psychopathology and Theories of Consciousness*. MIT Press.
- Billon, A. (2013a). Does consciousness entail subjectivity: The puzzle of thought insertion. *Philosophical Psychology*, 26(2):291-314.
- Billon, A. (2013b). Basic self-awareness and pathological self-doubt. Manuscript submitted for publication.
- Billon, A. (2014). Making sense of Cotard syndrome: insights from the study of depersonalization. Manuscript submitted for publication.
- Block, N. (2004). Some concepts of consciousness. In Chalmers, D., editor, *Philosophy of mind, Classical and contemporary readings*, pages 206-218. Oxford University Press. abridged version of Block (1995).
- Bortolotti, L. (2010). *Delusions and other irrational beliefs*. Oxford University Press.
- Brandom, R. (1997). A study guide. In Rorty, R. and Brandom, R., editors, *W. Sellars, Empiricism and Philosophy of Mind*. Harvard UP.
- Burge, T. (1988). Individualism and self-knowledge. *Journal of Philosophy*, 89 :649-663.
- Burge, T. (1996). Our Entitlement to Self-Knowledge. In *Proceedings of the Aristotelian Society*, volume XCVI, pages 91-116. JSTOR.
- Burge, T. (2000). Reason and the first person. In Wright, C., Smith, B., and Macdonald, C., editors, *Knowing our own minds*, pages 243-270. Oxford : Oxford University Press.
- Campbell, J. (2001). Rationality, meaning and the analysis of delusion. *Philosophy, Psychiatry & Psychology*, 8(2/3) :89-100.
- Cappon, D. (1969). Orientational perception : III. Orientational percept distortions in depersonalization. *American Journal of Psychiatry*, 125(8) :1048-1056.
- Cappon, D. and Banks, R. (1965). Orientational perception : II. Body perception in depersonalization. *Archives of General Psychiatry*, 13(4) :375-9.
- Charland-Verville, V., Bruno, M.-A., Bahri, M. A., Demertzi, A., Desseilles, M., Chatelle, C., Vanhauzenhuysse, A., Hustinx, R., Bernard, C., Tshibanda, L., et al. (2013). Brain dead yet mind alive: A positron emission tomography case study of brain metabolism in cotard's syndrome. *Cortex*.
- Chisholm, R. (1982). *The foundations of knowing*. University of Minnesota Press, Minneapolis.
- Cotard, J. (1891). *Études sur les maladies cérébrales et mentales*. J-B. Baillière.
- Cotard, J., Camuset, M., and Séglas, J. (1998). *Du délire des négations aux idées d'énormité*. Editions L'Harmattan.
- Davies, M., Breen, N., Coltheart, M., and Langdon, R. (2001). Monothematic delusions : Towards a two-factor account. *Philosophy, Psychiatry, & Psychology*, 8(2) :133-158.
- De Martis, D. (1956). Un caso di sindrome di Cotard (Studio clinico e psicopatologico). *Rivista Sperimentale di Freniatria*, 80 :491-514.

⁴³ Former versions of this paper have benefited from useful comments by Olivier Massin, Frédérique de Vignement, Ivan Vukovic, Sam Wilkinson. I am also indebted to discussions with Philippe Vellozzo and Maxime Bubrovsky. I am especially grateful to Uriah Kriegel and to an anonymous referee of this journal for their precious help. I would also like to thank the audiences of the APIC seminar at the Institut Jean Nicod and of the philosophy seminar of the university of Hradec Králové.

- Debruynne, H., Portzky, M., Van Den Eynde, F., and Audenaert, K. (2009). Cotard's syndrome : A review. *Current Psychiatry Reports*, 11.
- Descartes, R. (1985). *The philosophical writings of Descartes*, edited by J. Cottingham and D. Murdoch. Cambridge University Press.
- Dugas, L. (1898). Un cas de dépersonnalisation. *Revue Philosophique de la France et de l'Étranger*, 45 :500-507.
- Dugas, L. and Moutier, F. (1911). *La dépersonnalisation*. F. Alcan.
- Enoch, M. and Ball, H. (2001). *Uncommon psychiatric syndromes*. Arnold Publishers. 4th edition.
- Evans, G. (1983). *Varieties of Reference*. Oxford University Press, Oxford.
- Ey, H. (1950). Délire des négations. *Études psychiatriques*, 2 :427-452.
- Falret, J. P. (1864). *Des maladies mentales et des asiles d'aliénés*. Baillière.
- Garety, P. and Hemsley, D. (1997). *Delusions : Investigations into the psychology of delusional reasoning*. Psychology Pr.
- Gerrans, P. (2003). A one-stage explanation of the Cotard delusion. *Philosophy, Psychiatry, & Psychology*, 9(1) :47-53.
- Gertler, B. (2010). *Self-knowledge*. Routledge.
- Graham, G. (2009). *The disordered mind : an introduction to philosophy of mind and mental illness*. Routledge.
- Guralnik, O., Giesbrecht, T., Knutelska, M., Sirroff, B., and Simeon, D. (2007). Cognitive functioning in depersonalization disorder. *The Journal of nervous and mental disease*, 195(12) :983-988.
- Guralnik, O., Schmeidler, J., and Simeon, D. (2000). Feeling unreal : Cognitive processes in depersonalization. *American Journal of Psychiatry*, 157(1) :103-109.
- Heidegger, M. (1997). *GA 6.2 (Nietzsche II)*. Vittorio Klosterman, Frankfurt am Main. edited by Brigitte Schillbach.
- Henry, M. (1985). *Généalogie de la Psychanalyse*. Épiphanée. PUF.
- Hohwy, J. and Rosenberg, R. (2005). Unusual experiences, reality testing, and delusions of alien control. *Mind and Language*, 20(2) :141-162.
- Horn, L. R. (1989). *A natural history of negation*, volume 960. University of Chicago Press Chicago.
- Husserl, E. (1999). *The idea of phenomenology*. Kluwer Academic Publishers. edited by L. Hardy.
- James, W. (1983). *The principles of psychology*. Harvard University Press.
- Janet, P. (1903a). *Les obsessions et la psychasthénie : Tome I Analyse des symptômes*, volume 1. Félix Alcan.
- Janet, P. (1903b). *Les obsessions et la psychasthénie : Tome II*, volume 2. Félix Alcan.
- Janet, P. (1908). Le sentiment de dépersonnalisation. *Journal de Psychologie normale et pathologique*, 5 :514-516.
- Janet, P. (1928). *De l'angoisse à l'extase, vol. 2 (Les sentiments fondamentaux)*. Paris, Alcan.
- Janet, P. and Raymond, F. (1898). Névroses et idées fixes, vol. 2. *Paris : Félix Alcan.(b)*.
- Jaspers, K. (1997). *General Psychopathology*. JHU Press, Baltimore, 2 edition. trans. J. Hoening and Marian W. Hamilton.
- Kaplan, D. (1989). Demonstratives. In Almog, J., John Perry, and Wettstein, H., editors, *Themes from Kaplan*, pages 481-563. Oxford : Oxford University Press.
- Kemp, R., Chua, S., McKenna, P., and David, A. (1997). Reasoning and delusions. *The British Journal of Psychiatry*, 170(5) :398.
- Kennedy, R. C. and Graham, G. (2006). Extreme self-denial. In *Cartographies of the Mind : Philosophy and Psychology in Intersection*. Dordrecht : Kluwer.
- Kriegel, U. (2004a). Consciousness and self-consciousness. *The Monist*, 87 :185-209.
- Kriegel, U. (2004b). Moore's paradox and the structure of conscious belief. *Erkenntnis*, 61(1) :99-121.
- Kriegel, U. and Williford, K., editors (2006). *The self-representational approaches to consciousness*. MIT Press / Bradford Books, Cambridge.
- Kripke, S. A. (1980). *Naming and necessity*. Basil Blackwell.
- Krishaber, M. (1873). *De la névropathie cérébro-cardiaque*. Masson
- Levassor, J. and Dromard, G. (1908). *Une forme atypique de dépersonnalisation chez une délirante chronique*. Bibliothèque Nationale de France.
- Levine, J. (2001). *Purple haze : The puzzle of consciousness*, volume 6. Oxford University Press Oxford.
- Lewis, D. (1970). Anselm and actuality. *Noûs*, 4(2) :pp. 175-188.
- Mayer-Gross, W. (2011). On depersonalization. *British Journal of Medical Psychology*, 15(2) :103-126.
- McKay, R. and Cipolotti, L. (2007). Attributional style in a case of cotard delusion. *Consciousness and Cognition*, 16(2) :349 - 359.
- Merleau-Ponty, M. (2007). *The Merleau-Ponty reader*, edited by T. Toadvine and L. Lawlor. Northwestern University Press.
- Mirian, D., Heinrichs, R. W., and Vaz, S. M. (2011). Exploring logical reasoning abilities in schizophrenia patients. *Schizophrenia Research*, 127(1-3) :178 - 180.
- Nozick, R. (1981). *Philosophical Explanations*. Cambridge. Harvard University Press.
- Pacherie, E. (2009). Perception, emotions and delusions : revisiting the capgras' delusion. In Bayne, T. and Fernandez, J., editors, *Delusions and Self-Deception : Affective Influences on Belief Formation*. Psychology Press.
- Parnas, J. and Sass, L. A. (2001). Self, solipsism, and schizophrenic delusion. *Philosophy, Psychiatry & Psychology*, 8(2-3) :101-120.
- Ramirez-Bermudez, J., Aguilar-Venegas, L., Crail-Melendez, D., Espinola-Nadurille, M., Nente, F., and Mendez, M. (2010). Cotard syndrome in neurological and psychiatric patients. *The Journal of neuropsychiatry and clinical neurosciences*, 22(4) :409-416.
- Ratcliffe, M. (2004). Interpreting delusions. *Phenomenology and the Cognitive Sciences*, 3(1) :25-48.
- Ratcliffe, M. (2010). Delusional atmosphere and delusional belief. In Schmicking, D. and Gallagher, S., editors, *Handbook of Phenomenology and Cognitive Science*, pages 575-590. Springer Netherlands.
- Revault d'Allonnes, G. (1907). *Les inclinations*. Felix Alcan.
- Ribot, T. (1896). *La Psychologie Des Sentiments*. Bibliothèque de philosophie contemporaine. F. Alcan.
- Roberts, W. W. (1960). Normal and abnormal depersonalization. *The British Journal of Psychiatry*, 106(443) :478-493.
- Russell, B. (1912). *The Problems of Philosophy*. Williams and Norgate, London.
- Sass, L. (2004). Some reflections on the (analytic) philosophical approach to delusion. *Philosophy, Psychiatry, & Psychology*, 11(1) :71-80.
- Sass, L., Pienkos, E., Nelson, B., and Medford, N. (2013). Anomalous self-experience in depersonalization and schizophrenia : A comparative investigation. *Consciousness and Cognition*, 22(2) :430 - 441.
- Sedman, G. (1972). An investigation of certain factors concerned in the aetiology of depersonalization. *Acta Psychiatr Scand*, 48(3) :191-219.
- Séglas, J. and Meige, H. (1895). *Leçons cliniques sur les maladies mentales et nerveuses : (Salpêtrière, 1887-1894)*. Asselin et HouzEAU.
- Shoemaker, S. (1996). *The First-Person Perspective, and Other Essays*. Oxford University Press.
- Shorvon, H. (1946). The depersonalization syndrome. *Proceedings of the Royal Society of Medicine*, 39(12) :779-791.
- Sierra, M. (2009). *Depersonalization : a new look at a neglected syndrome*. Cambridge Univ Press.
- Sierra, M. and Berrios, G. E. (2000). The cambridge depersonalisation scale : a new instrument for the measurement of depersonalisation. *Psychiatry Research*, 93(2) :153-164.
- Sierra, M. and David, A. S. (2011). Depersonalization : a selective impairment of self-awareness. *Consciousness and cognition*, 20(1) :99-108.
- Simeon, D. and Abugel, J. (2006). *Feeling unreal : Depersonalization disorder and the loss of the self*. Oxford University Press, USA.
- Simeon, D., Kozin, D. S., Segal, K., Lerch, B., Dujour, R., and Giesbrecht, T. (2008). De-constructing depersonalization : Further evidence for symptom clusters. *Psychiatry Research*, 157(1&e3) :303 - 306.
- Speak, G. (1990). An odd kind of melancholy: reflections on the glass delusion in Europe (1440-1680). *History of Psychiatry*, 1(2) :191.
- Swamy, N., Sanju, G., and Mathew Jaimon, M. (2007). An overview of the neurological correlates of cotard syndrome. *The European journal of psychiatry*, 21(2) :99-116.
- Unger, P. (1997). I do not exist. In Rea, M. C., editor, *Material constitution : a reader*, pages 175-190. Rowman & Littlefield Publishers Inc.
- Vaxevanis, A. and Vidalis, A. (2005). Cotard's syndrome. A three-case report. *Hippokratia*, 9(1) :41.
- Weiskrantz, L. (2009). Blindsight: a case study spanning 35 years and new developments.
- Williams, B. A. (1978). *Descartes : The Project of Pure Enquiry*. Harvester press.
- Young, A. and De Pauw, K. (2003). One stage is not enough. *Philosophy, Psychiatry, & Psychology*, 9(1) :55-59.
- Young, A. and Leafhead, K. (1996). Betwixt life and death : Case studies of the Cotard delusion. In Halligan, P. W. and Marshall, J. C., editors, *Method in madness : Case studies in cognitive neuropsychiatry*, pages 147-171. Psychology Press.
- Young, G. (2012). Delusions of death and immortality : A consequence of misplaced being in cotard patients. *Philosophy, Psychiatry, & Psychology*, 19(2) :127-140.
- Zemach, E. M. (1985). *De se and Descartes : a new semantics for indexicals*. *Noûs*, 19.

Appendix. Patching the rationalist interpretations with “substantial rationality”

It might be tempting for the rationalist to bite the bullet and grant that minimal rationality cannot explain the cogito, but to add that there is a more substantial notion of rationality that can. He or she could even claim that the two rationalist interpretations could be patched by adding a normal (unimpaired) self-awareness condition and by accounting for such a self-awareness in terms of this more substantial form of rationality. As I discuss the nature of self-awareness and its impairment among self-doubters in much more details elsewhere (Author, xxxxa) I will remain cursory here.

This ‘substantial rationality strategy’ is actually suggested by Burge (2000). In his 1996 paper, where he defended the first rationalist interpretation, Burge already noted that “there is more to the epistemic status of the cogito” than “its being self-evidently self-verifying” (Burge, 1996, 93). In his 2000 sequel, he puts forward what he takes to be another correct interpretation of the cogito, which can be considered as a patched version of the second rationalist interpretation.⁴⁴ According to Burge (2000) then, the mastery of the ‘I’-concept is enough, in itself, to account for the cogito:

For someone who has the I concept, there is no step from recognition of the occurrence of a thought to the conclusion that there must be a self. There is no step, inference, or

⁴⁴ He makes it clear, once again, there that he believes that there are many different and equally correct interpretations of the cogito, explaining that “[he] will neglect other answers to Lichtenberg [’s objections to the cogito] that [he] takes obvious and sufficient in themselves (Burge, 2000, 247).”

postulation at all Such applications [of 'I'] fall under the rule that the referent is the author of the thought. Given that the first-person concept is applied, there is no possibility of reference failure (Burge, 2000, 245).

This sounds very much like the second rationalist interpretation. However Burge ~~makes it clear~~ that minimal rationality is not sufficient for the mastery of the 'I'-concept: "only beings whose nature makes them capable of critical reason can have the full first-person concept (Burge, 2000, 111)". Burge and other rationalists might accordingly contend that what pathological self-doubters lack and which can explain their impaired self-awareness and their uncertainty is a more substantial form of rationality on which the full mastery of the 'I'-concept is based.

Appealing to a more substantial form rationality will however not do. First, nothing indicates that all pathological self-doubters lack the substantial rationality in question. We have seen, in particular than non-delusional patients seem to have an intact capacity for critical reasoning (§4.1). Second, self-doubters do have abnormal experiences and adduce them to justify their weird doubts. Explaining the cogito in terms of rationality, be it minimal or substantial, will simply leave those experiential disorders unexplained. Empiricists on the other hand directly explain the latter disorders. The empiricist interpretation fits better with the data than the substantial rationality interpretation.

