Jaspers' Dilemma:

The Psychopathological Challenge to Subjectivity Theories of Consciousness

Alexandre Billon and Uriah Kriegel

Forthcoming in R. Gennaro (ed.), *Disturbed Consciousness*. MIT Press.

Introduction/Abstract

According to what we will call subjectivity theories of consciousness, there is a constitutive connection between phenomenal consciousness and subjectivity: there is something it is like for a subject to have mental state M *only if* M is characterized by a certain mine-ness or for-me-ness. Such theories appear to face certain psychopathological counterexamples: patients appear to report conscious experiences that lack this subjective element. A subsidiary goal of this chapter is to articulate with greater precision both subjectivity theories and the psychopathological challenge they face. The chapter's central goal is to present two new approaches to defending subjectivity theories in the face of this challenge. What distinguishes these two approaches is that they go to great lengths to interpret patients' reports at face value – greater length, at any rate, than more widespread approaches in the extant literature.

1. Consciousness and Subjectivity

Compare your experiences of drinking apple juice and drinking a banana smoothie. These experiences are very different in many respects: there is a gustatory apple-ish way it is like for you to have the former and a gustatory banana-ish way it is like for you to have the latter; there is a tactile juice-ish way it is like for you to have the former and a tactile

smooth-ish way it is like for you to have the latter; and so on. But there is also one respect in which the two experiences are exactly the same: in both cases it is *for you* that it is like something to have them. By this we mean not only that both experiences *are* yours, but more strongly that both are *experienced* as yours. We call this the *subjectivity* of experience. Your apple-juice and banana-smoothie experiences are different in gustatory and tactile respects, but are the same in respect of subjectivity.

It is an open question what the relationship is between subjectivity and phenomenal consciousness. Call the following the *subjectivity principle*:

(SP) Necessarily, a mental state M exhibits phenomenal consciousness only if M exhibits subjectivity.

According to SP, there is a necessary, constitutive connection between phenomenal consciousness and subjectivity. Some theories of consciousness in the extant literature are committed to SP, some to ~SP, and some to neither. Call those that are committed to SP *subjectivity theories* of consciousness. According to subjectivity theories, a phenomenally conscious state that lacks this dimension of for-me-ness or subjectivity is metaphysically impossible.

There are three main kinds of subjectivity theory currently being discussed. One is 'higher-order representationalism.' According to Rosenthal (1990), every conscious state is a state the subject is aware of, and moreover aware of as her own. This awareness is implemented by a higher-order representation of the subject's experience (see also Gennaro 1996, 2012). Crucially, proponents of higher-order representationalism typically hold that the phenomenal character of a conscious state is determined by the manner in which it is higher-order represented.[1] To that extent, they appear committed to SP.[2]

A second kind of subjectivity theory is self-representationalism. According to Kriegel (2009), conscious states are states the subject is aware of (as hers), not because they are targeted by higher-order representations however, but because they are targeted *by themselves* (see also Williford 2006). Every conscious state represents itself (and moreover, represents itself as belonging to the subject), and it is in virtue of this self-representation that the subject is aware of it (as hers). The fact that the conscious state is

represented by itself means that it is represented by a conscious state, which in turn means that the subject's awareness of it (as hers) does show up in the subject's overall phenomenology. This is a clear case of a subjectivity theory, then.

A third kind of subjectivity theory is what we call 'acquaintance theory.' According to Levine (2001), the subject's awareness of her own conscious states is not implemented in a representational state at all. Rather, it involves a *sui generis* awareness relation – 'acquaintance' – that is unlike representational relations in some crucial respects, most notably by being 'factive' (see also Billon 2005). This *sui generis* acquaintance relation is intended to capture the elusively immediate character of our awareness of our conscious states, as it shows up in our phenomenology. It is thus intended as built into the phenomenology in the manner required by subjectivity theories.

These are the main *contemporary* kinds of subjectivity theory, but the latter has a long history. Already Locke writes that "Whilst [the soul] thinks and perceives… it must necessarily be conscious of its own Perceptions" (ECHU 2.1.12), and of course that "It [is] impossible for any one to perceive, without perceiving, that he does perceive" (ECHU 2.27.9). Earlier yet, Aristotle writes in the *Metaphysics* 12.9 that "[conscious] knowing, perceiving, believing, and thinking are always of something else, but of themselves on the side (*en parergo*)" (1074b35-6).

Subjectivity theories were later pursued in the Brentano School and the phenomenological movement. Brentano himself was a self-representationalist, holding that every conscious state is intentionally directed primarily at some external object but secondarily at itself:

> [Every conscious act] includes within it a consciousness of itself. Therefore, every [conscious] act, no matter how simple, has a double object, a primary and a secondary object. The simplest act, for example the act of hearing, has as its primary object the sound, and for its secondary object, itself, the mental phenomenon in which the sound is heard. (Brentano 1874: 153-5)

Husserl, meanwhile, is an early acquaintance theorist, holding that our awareness of our own lived experience is not a standard kind of object-positing intentionality but a special, *sui generis*, non-object-positing intentionality:

> Every act is consciousness of something, but there is also consciousness of every act. Every act is 'sensed,' is immanently 'perceived' (internal consciousness), although naturally not posited, meant (to perceive here does not mean to grasp something and to be turned towards it in an act of meaning). (Husserl 1928: 130)

Although disagreement on the nature of subjectivity persisted among Brentano's and Husserl's students, the notion that some kind of subjectivity theory must be right became orthodox in Austro-German philosophy, including outside the Brentano School and the phenomenological movement (for example in the Heidelberg School).

An interesting case concerns the German philosopher Karl Jaspers, whose thought combines neo-Kantian, phenomenological, and existentialist elements. Jaspers started out as a psychiatrist (his doctorate was in medicine and his *Habilitation* in psychology), but converted to philosophy circa 1920. Jaspers' commitment to a subjectivist theory of consciousness is unquestionable:

> Self-awareness is present in every psychic event… 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. We have termed this '*personalization*.' (Jaspers 1913: 121; italics original)

However, Jaspers is unique among Austro-German philosophers of the time in realizing that certain psychopathological phenomena presented a prima facie threat to subjectivity theories. This is the topic of the next section.

2. Alienation Symptoms and Jaspers' Dilemma

One of the most striking features of Jaspers' (1913) *General Psychopathology* is the treatment of certain patients, mostly schizophrenic, who expressly disown some of their mental states. One of Jaspers' patients describes some of his thoughts as follows:

> I have never read nor heard them; they come unasked; I do not dare to think I am the source but I am happy to know of them without thinking them. They come at any moment like a gift and I do not dare to impart them as if they were my own. (Jaspers 1913: 123)

Such a symptom, to which he referred as "implanted thoughts," is now known as *thought*

insertion. Patients suffering from thought insertion complain of having thoughts 'in them' that are not theirs and seem to be merely 'inserted in them.' Other patients described by Jaspers suffer from what we call today 'delusions of alien control'. Immediately after shouting, one of them explains: "I never shouted, it was the vocal chord that shouted out of me" (Jaspers 1913: 124). Another says: "The 'shouting miracle' is an extraordinary occurrence… my muscles are subject to some influences that can only be ascribed to some external force" (Ibid.)[3]

It is natural to interpret such statements as reporting conscious states lacking subjectivity: the patient does not experience the relevant conscious state *as his*. Importantly, patients suffering from thought insertion and alien control seem to mean what they say: they reject watered-down or metaphorical interpretations.[4] However, to take their reports at face value is to reject SP. Jaspers was thus confronted with the following dilemma: either (a) we can make sense of the patients' reports, but subjectivity theories should be rejected, or (b) subjectivity theories need not be rejected, but the patients' reports must be deemed unintelligible or incomprehensible. In some passages, Jaspers seems to lean toward (a): "If these psychic manifestations occur with the awareness of their not being mine… we term them depersonalization" (1913: 121). In other places, however, he unequivocally goes for (b), claiming that the patients' reports are "in principle psychologically inaccessible to us" and incomprehensible (in the sense that we cannot understand what experience would warrant them), because "we are not able to have any clear sight of a [conscious] psychic event without our self-awareness being involved" (1913: 578). Jaspers concluded that phenomenological psychopathology cannot make sense of thought insertion and alien control delusions; they can be addressed only neuroscientifically.

A century later, Jaspers' dilemma remains unresolved. It is fair to say, however, that intelligibility-denial is a minority position today. One reason is that there is no independent evidence for the relevant patients' irrationality. They are not generally committed to inconsistencies, at least no more than healthy controls (whose speech we would not deem unintelligible).[5] After more than sixty years, the search for abnormal patterns of deductive reasoning in deluded patients has never led to convincing results

(Kemp et al. 1997; Mirian et al. 2011).[6]

Furthermore, such "alienation symptoms" (suggesting consciousness without subjectivity) can be found in patients suffering from somatoparaphrenia and depersonalization, two conditions that involve no schizophrenia or dementia. Intelligibility-denial is particularly unappealing for these patients. Patients suffering from somatoparaphrenia typically complain that one of their limbs belongs to someone else. This strange condition is often associated with hemineglect: patients neglect (are unaware of) items on one side of visual space, including the disowned limbs. Importantly for our present purposes, although most somatoparaphrenics are unable to feel sensations in the disowned limb, some (we know of four cases) do feel sensations in those limbs but feel the limbs are not theirs:

> Immediately after the experiment, we asked F.B. how she could report touches on someone else's hand. Her response was initially elusive; however she eventually explained the phenomenon as follows (though adding 'Yes, I know, it is strange'): her absent-minded niece would always forget her hand on the patient's bed while leaving the hospital, so that F.B. used to take care of it until the niece came back to visit her again (Bottini et al. 2002: 251).

Specifically asked about how it was possible to perceive stimuli delivered to another's hand, another patient (A.F.) reported that "many strange things can happen in life" (Moro et al. 2004: 440). These reports suggest that patients have sensations which do not feel theirs, and which to that extent lack subjectivity. Given the very circumscribed character of these delusions (somatoparaphrenia is a so-called monothematic delusion), intelligibility-denial is implausible. Somatoparaphrenia thus provides another instance of Jaspers' dilemma that seems to threaten subjectivity theories.

Intelligibility-denial is even less plausible with depersonalization (Sierra 2009). Jaspers, we saw, used the term 'personalization' to refer to what we call subjectivity. Even though we shall see that it is interesting and motivated, Jaspers' use of the term is somehow idiosyncratic. The term 'depersonalization' was coined by the French philosopher and psychologist Ludovic Dugas at the turn of the century, to characterize deep and pervading modifications of the way things appeared to patients who reported a whole spectrum of abnormal experiences, ranging from the feeling of lacking bodily parts

to the feeling of being unreal or of not being oneself.[7] Consider the following range of reports:

> Parts of my body feel as if they didn't belong to me. (Sierra and Berrios 2000: 160)

> Often I have to... enter a shop to talk, to ask for something, in order to get a new proof that I am myself. (Séglas and Meige 1895: 141)

> I must be someone, I am someone, everybody else feels someone, but I am not myself. I suddenly felt I am in half, there is two of me. (Shorvon 1946: 782)

> I have stopped being. (Mayer-Gross 2011: 106)

Such patients seem to lack a sense of self, or as Simeon and Abugel (2006: 25) put it, "a clear feeling of 'I'." If they have no feeling of 'I', then arguably their conscious states cannot be experienced as belonging to such an I. Without a felt me, there can be no felt mine-ness – no subjectivity. If so, these patients' conscious states must lack subjectivity. This includes conscious thoughts:

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

As well as conscious intentions-in-action:

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

And even algedonic sensations of pain and pleasure:[8]

> 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: 163)

> 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 three kilometers away from my brain, and I do not know whether I am suffering. (Janet 1928: 65; our emphasis)

When depersonalization reaches its climax, subjectivity seems to withdraw from all conscious states, leaving the subject with a bold feeling of inexistence:

> 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... My eyes see and my spirit perceives, but the sensation of what I

see is completely absent. (Sierra 2009: 8)

> 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; our emphasis)

What makes depersonalized patients particularly problematic for SP is that they still *believe* the states that feel alien to be theirs. Patients say that they feel *as if* those states did not belong to them. There is a feeling of alien-ness, but patients do not *endorse* the feeling. That is, they do not take the appearance of alien-ness at face value, and thus ultimately self-attribute those states. In other words, they are not delusional.[9] Now, given that they are non-delusional, it is hard to maintain that depersonalized patients' reports are sensible. Jaspers' dilemma, when extended to depersonalization, thus strongly suggests an argument against subjectivity theories of consciousness.

To summarize, Jaspers' dilemma is the forced choice between two independently unappealing options: rejecting SP, and hence subjectivity theories of consciousness, or denying the intelligibility of certain patients' reports. Although originally concerned with certain symptoms of schizophrenia, Jaspers' dilemma can be extended to somatoparaphrenia and depersonalization. In all three cases, there is an 'alienation symptom,' whereby patients report that some of their mental states – call them 'alien states' – are not theirs (schizophrenia and most somatoparaphrenia) or at least do not *feel* theirs (depersonalization and some somatoparaphrenia). In all three cases, it seems that if we want to make sense of patients' reports, we must suppose that even though they are conscious, the alien states lack subjectivity. They thus appear to constitute counterexamples to SP.

3. Consciousness without Subjectivity?

It is generally preferable to make sense of patients' reports than to dismiss them as unintelligible. This is particularly so when the patients appear otherwise rational. Now, the most natural way of making sense of reports of the sort cited above is to suppose that

they involve conscious experiences that lack subjectivity. This suggests a straightforward argument from alienation against subjectivity theories. The argument proceeds in two steps.

First, if we want to make sense of patients' reports about alien states, we must suppose that these states (a) are conscious but (b) lack subjectivity. Call this the *Interpretive Constraint*. The rationale for this constraint is that (i) to make sense of these reports, we must treat them as the reports of someone sensible, and (ii) someone sensible would make the above-cited reports only if (a) their alien states were conscious (b) but lacked subjectivity. Plausibly, (i) is definitional. As for (ii), we must suppose (a) because the patients manage to *report* their alien states, and reportability is still our best third-person operational index of consciousness; and we must suppose (b) because the content of the reports is that the relevant states do not feel as though they belong to one, and what we call subjectivity is precisely this feeling.

With the Interpretive Constraint in place, it is but a short step to conclude that alien states are conscious but lack subjectivity. One only needs to assume that *we should make sense of the patients' reports*. We may call this the *Rationality Constraint*, since as noted, there is good evidence that some of the relevant patients are sufficiently rational to be deemed sensible. As we have seen, the Rationality Constraint is particularly plausible for somatoparaphrenia, depersonalization, and other cases of non-delusional alienation symptoms (where the patients do not endorse their alienation feelings).

We may now formulate the *Alienation Argument* against subjectivity theories as follows:

1. **Alienation symptom.** Some patients report that their alien states feel as though they are not theirs (that is, as though they lack subjectivity).

2. **Interpretative constraint.** To make sense of such reports, we must suppose that alien states are (a) conscious but (b) lack subjectivity.

3. **Rationality constraint.** We should make sense of these patients' reports.

4. **Conclusions.** We must suppose that alien states are conscious but lack subjectivity. Accordingly, SP and the subjectivity theories are false.

As it stands, this argument applies to all subjectivity theories (higher-order, self-representational, and acquaintance theories).

A fortiori, it applies to specific subjectivity theories. Thus, Liang and Lane (2009) have put forward a clear instance of this general schema against Rosenthal's higher-order representationalism. They argue that patient FB (the somatoparahrenic who reported sensing touch in her niece's hand) represents her tactile sensations as belonging to someone other than self (Liang and Lane 2009: 664-5). From this they conclude that FB has conscious sensations lacking subjectivity, and that therefore Rosenthal's higher-order representationalism must be rejected. They see clearly, however, that the point generalizes to other subjectivity theories of consciousness: "If the conclusions reached here are correct, Kriegel's views and the views of others who posit a necessary connection between [consciousness and subjectivity] are wrong" (Ibid.: 667).

4. Defending Subjectivity Theories

We consider the Alienation argument a genuine and important challenge to subjectivity theories of consciousness. Both of us, however, remain committed to such theories. In the remainder of this paper, we discuss the various options available to the subjectivity theorist in responding to the challenge, providing a menu of options for the proponent of SP.

The first option is to deny Premise 1 of the Alienation Argument, that is, deny that there really are alienation symptoms. On this view, we mischaracterize patients' reports when we say that they feel as if their alien states are not theirs. It could be argued, for example, that in truth we do not fully understand their complaints (the complaints that motivate ascribing to them experiences that are not or do not feel as their own). Alternatively, one might want to dispute the significance or the reliability of the case

reports we have relied on. The standards of methodological rigor in one domain of research may be laxer than in others, after all.

Such moves have some plausibility when it comes to somatoparaphrenia (see Rosenthal 2010). For none of the patients we know of claim feeling *sensations that are not theirs*. Rather, they say that they feel touch in someone else's limb. This does not yet imply that they feel sensations that are not their own – unless it is analytic that one cannot feel one's sensations but in one's own body, which we have phenomenological and empirical reasons to deny (de Vignemont 2007). Thus, it seems possible to experience tactile sensations at the tip of tools When touching the ground with a cane, we can, arguably, feel the touch at the end of the cane rather than at the hand holding it (O'Shaughnessy 2003).[10] So denying Premise 1 may be plausible for somatoparaphrenics. However, it seems less so when it comes to depersonalized patients and schizophrenics suffering from thought insertion. For to have alienation symptoms *just is* to report that some mental state is not, or does not feel to be, one's own.[11] Given that the relevant patients say precisely this, and given that the huge number of converging case reports makes this claim reliable, there is no question but that they suffer from alienation symptoms (see Mullins and Spence 2003). Thus depersonalization and schizophrenia call for another response to the Alienation Argument.

One might venture to deny the Rationality Constraint, Premise 3 of the Alienation Argument). For example, Coliva (2002) argues that schizophrenics would not be classified as mentally ill if they were sensible. She asks rhetorically:

> Why should we consider her report as an expression of some kind of cognitive illusion, which we take as a symptom of mental illness, as opposed to, at most, a possible mistake in identifying the producer of the thought? And, connectedly, why should we try to cure her, rather than just, at most, correct her? (Coliva 2002: 42)

However, there are many criteria for mental illness that do not involve unintelligibility, and that this objection is not cogent. For example, even though reports by depressed and bipolar subjects are eminently sensible, depression and bipolar disorder are usually considered mental illnesses. This classification has nothing to do with rationality or intelligibility. It stems from the distress and the disability they cause

Still, we acknowledge that many will find attractive the claim that schizophrenics' reports do not make sense. Regardless of whether this response is plausible for schizophrenia, however, it is entirely implausible for depersonalization. Depersonalized patients, as we have seen, are not delusional. It is true that there is a syndrome which is connected to depersonalization and which involves delusions: the Cotard syndrome, often characterized as the delusional form of depersonalization (depersonalization being, conversely the "as if" form of the Cotard syndrome).[12] There are, however, clear cases of non-delusional patients reporting alien states. Those patients, to whom we have reserved the term "depersonalization" in the first place, seem perfectly rational. Their cognitive functioning has been well studied. They were found not to differ from matched anxious or depressed patients (Sedman 1972), and to differ from normal subjects only in some very specific aspects of low-level perceptual memory and attention (Guralnik et al. 2000, 2007). Their general intelligence, their executive functioning and other aspects of their memory and attention are perfectly normal. They should accordingly be no less intelligible than normal subjects.

The last option is to deny Premise 2 of the argument, the Interpretative Constraint. This means either denying that to make sense of the reports, we must suppose that alien states are conscious (2a), or denying that to do so we must suppose that alien states lack subjectivity (2b). We think that these are the most plausible strategies available to the subjectivity theorist and accordingly discuss them more fully in the remainder of the chapter. Indeed, the most popular subjectivist response to the Alienation Argument, which we may call the *agency response*, takes this form. However, we think the agency response faces tremendous difficulties, which moreover apply to other responses structurally similar to it. Instead, we offer two alternative responses to the Alienation Argument that we find much more plausible.

5. Subjectivity and Subjectivity*

In this section we want to discuss responses to the argument that have the following general structure. Subjectivity, it is claimed, is a slippery and potentially ambiguous

notion. The fact that someone says "X is not mine" or "X does not feel mine" might therefore be interpreted in different ways, depending on the different senses of the notion. This suggests a response that claims:

- **Ambiguity**: that there is a sense of 'subjectivity', call it subjectivity*, different from that used in SP and subjectivity theories;
- **Rewriting**: that alien states (i) do not differ from non-alien states with respect to subjectivity, but only (ii) with respect to subjectivity*.

Call responses to the Alienation Argument that take this form *subjectivity* responses*. There are various subjectivity* responses, depending on the notion of subjectivity* appealed to. They all face the same twofold challenge. First of all, such responses must show that lack of subjectivity* can account for the phenomenological difference between alien states and non-alien states. They also need to show, however, that patients' behavior suggests that their alien states are indeed subjective; this task is often neglected in the literature.

When they do attend to this challenge, subjectivity* responses tend to argue as follows:

1. The patient acknowledges, or would acknowledge, that her alien state feels *in her* or *in her mind*;
2. Being in her (or her mind) entails being subjective in the sense relevant to SP; so,
3. The patient effectively acknowledges, or would acknowledge, that her alien state is subjective.

Campbell (1998, 1999), Gallagher (2000), and Graham and Stephens (2000) appear to use this reasoning. Campbell writes:

> The thought inserted into the subject's mind is indeed in some sense his, just because it has been successfully inserted into his mind; it has some special relation to him. He has, for example, some especially direct knowledge of it. (Campbell 1999: 610)

In the same vein, Gallagher writes:

> For that reason the schizophrenic should provide a positive answer to what he might rightly regard as a nonsensical question: Are you sure that *you* are the one who is experiencing these thoughts? After all, this is precisely his complaint. *He* is experiencing thoughts that seem to be generated by others. (Gallagher 2000: 231; italics original)

In other words, the patient would not complain if it were not for the fact that some sense of mine-ness is still there.

This line of response has an obvious attraction to it. Nonetheless, we think the way it has been pursued in the extant literature faces significant difficulties and bears improvement. For starters, the above argument's conclusion (that patients effectively acknowledge that their alien states are subjective) is somewhat uncharitable to the patients. It means that while patients say something like "X does not feel mine" or "X is not mine (even though it is in me/my mind)", what they really should have said is rather "X does feel mine" or "X is mine" (Billon 2013). More importantly, it is unclear how proponents of the subjectivity* response propose to account for the phenomenal difference between an alien state X and a non-alien state Y. The task is to (i) spell out what distinguishes the phenomenology of X from that of Y without appealing to a difference in subjectivity and (ii) argue that it amounts to a difference in subjectivity*. But the standard construals of subjectivity* seem to lack the resources to manage this with plausibility.

Consider the most popular subjectivity* response in the case of thought-insertion, which identifies subjectivity* with the *sense of agency*.[13] On this view, whereas healthy subjects experience themselves as *doing the thinking*, schizophrenic patients suffering from thought insertion do not. Accordingly, inserted thoughts are not experienced by subjects as *done* or *performed* or *authored* by them. Subjects feel themselves patients rather than agents, so to speak, of the thought processes taking place in their mind. The thinking is something that happens to them, not something that they do. Nonetheless, they experience the thinking as taking place in *their* mind. The main problem with this response is that healthy subjects have many thoughts that do not seem to come with a sense of agency. Many of our daily thoughts come unbidden. Some – obsessive thoughts, for example – are even *intrusive*, occurring *against* our will. It is far from clear that in

such cases we experience ourselves as *doing the thinking*. Yet such thoughts do not feel alien the way the schizophrenic patient's inserted thoughts do.[14] Indeed, schizophrenic patients have some intrusive thoughts as well but they too manage to distinguish them from inserted thoughts (Eisen et al. 1997).[15]

Partly in reaction against the agency accounts of thought insertion, some have recently described the alien feeling associated with thought insertion in terms of a "sense of endorsement."[16] Inserted thoughts, on this account, are thoughts to which the patient does not feel *committed*, independently of whether she feels like the agent of those thoughts (Fernàndez 2010: 67; Bortolotti and Broome 2009). This lack of sense of endorsement might manifest itself in various ways: inability to provide reasons for endorsing the thought content, failure to act consistently with the thought being true, and so on (Bortolotti and Broome 2009: 210).[17] Here too, however, it is dubious that a sense-of-endorsement response can plausibly account for the phenomenal difference between alien and non-alien states. Merely intrusive thoughts often go unendorsed by their subjects, and patients suffering from obsessive thoughts typically endorse thoughts that directly contradict their intrusive thoughts (Purdon 2004). A classic example involves a caring and loving mother who is obsessed with thoughts like "I should kill my child" or "it would be better if he were dead." This mother will not only fail to behave as if she took these thoughts to be true, she will also be totally unable to find reasons for endorsing the thought. The very occurrence of the thoughts, it is true, might frighten her, with the idea that she might act upon them, or have unconscious reasons for believing them. This kind of fear, however, does not amount to anything like endorsing the thought.[18] The thought remains unendorsed, yet clearly does not qualify as an inserted thought. This suggests that the phenomenal difference between inserted thoughts and intrusive thoughts is not a matter of sense of endorsement. Non-inserted thoughts can lack the phenomenology of endorsement just as much as inserted thoughts.

An assumption shared by the agency and endorsement versions of the subjectivity* response is that the phenomenal difference between alien and non-alien states must amount to some feature missing from the pathological case that is present in the "normal" case. One version identifies the sense of agency as the missing feature, the

other identifies it as the feeling of endorsement, but in both cases the unargued-for assumption is that patients are *missing* something. Importantly, this is a *substantive* assumption: the phenomenal difference between alien and non-alien states could just as well consist in new features being *added* to the alien states (Zahavi and Kriegel ms: 11). On this view, an inserted thought, qua inserted, instantiates all the phenomenal properties of a normal thought; but in addition, it instantiates an extra phenomenal property: *it feels inserted*. The patient complains because this extra phenomenal property is foreign, absent from normal experience.

Generalizing the strategy, one may hold that alien states exhibit a phenomenology of alienation absent from non-alien states. It may be that the phenomenology of alienation is entirely different in schizophrenia, somatoparaphrenia, and depersonalization, so that each condition involves its own distinctive extra phenomenal property; or it may be that there is an underlying phenomenal commonality among some such conditions, so it is the selfsame extra phenomenal feature exhibited in all of them. Either way, the phenomenal difference between alien and non-alien states is explained in terms of presence of an extra feature in the alien states, not in the non-alien states.

It might be objected that the "something extra" approach is methodologically problematic, compared to the "something missing" approach, insofar as it requires new posits. The "something missing" approach comports better with the so-called null hypothesis, as it posits no new features, instead explaining the data in terms of absence of already acknowledged features. The obvious response to this objection, however, is that as we have just seen the explanations provided by the "something missing" approach are inadequate. To be sure, some other explanation, citing a feature other than agency and endorsement, may fare better. But pending such an as yet non-existent explanation, there is clear evidence for the existence of a phenomenology of alienation, namely, the relevant verbal reports of schizophrenia, somatoparaphrenia, and depersonalization patients. These reports constitute evidence for the existence of the phenomenology of alienation, we are suggesting, because the latter may be the best explanation of them.

Importantly, to posit a phenomenology of alienation present in alien states but absent in non-alien states is not to commit to a categorically new, sui generis type of

phenomenology. For all we have said here, the phenomenology of alienation may result from an unusual combination of ordinary phenomenal elements. Compare debates on cognitive phenomenology. Some philosophers maintain that cognitive states have no phenomenology, others that they exhibit a sui generis non-sensory phenomenology. But an intermediate position is that cognitive states have a distinctive phenomenology, though one that results from a distinctive combination of sensory elements (e.g., Robinson 2006). Regardless of how plausible this is for cognitive phenomenology, we maintain that this is a highly plausible view of the phenomenology of alienation experienced by the kinds of patients discussed here.

One may speculate about a broadly "ideological" preconception of psychopathology operating in the background of the "something missing" approach. Pathology is portrayed as a kind of *imperfection*, involving a diminution of the normal state, which in comparison represents a kind of perfection or fullness. This "ideology" runs deep in professional psychopathology, but suffers from a distinct dearth of evidence. Intuitively, in the above patients something atypical *occurs*, takes place; it is not as though something *fails to occur* that typically does. Patients complain about something upsetting being *present* in their phenomenology; they do not bemoan the loss of something. As far as they are concerned, the problem is not that their pathological phenomenology is *impoverished* in comparison to the phenomenology they experienced prior to pathology. On the contrary, their phenomenology has been *augmented* by a new and foreign element.[19] To that extent, the "something extra" approach is much more charitable to patients, taking their reports much more at face value.

The "something extra" approach is motivated by interpretive charity toward inserted-thought subjects. However, it seems to fit less naturally depersonalization patients. The latter often explicitly report something *missing* in their phenomenology. In one of the earliest descriptions of depersonalization, Sierra mentions five patients, all of whom

> complained almost in the same terms of a lack of sensations... To them it was a total lack of feelings, as if they were dead... They claimed they could think clearly and properly about everything, but the essential was lacking even in their thoughts. (Sierra 2009: 8)

If we put a premium on interpretive charity, then, we might want to seek another approach that could accommodate the relevant reports within the framework of subjectivity theories.

6. Consciousness and Consciousness*

A subjectivist tack rather neglected in the literature is what we might call the *consciousness* response*. This attempts to show:

- **Ambiguity**: that there is a sense of 'consciousness', call it consciousness*, that differs from the one used by SP and subjectivity theories;
- **Rewriting**: alien states resemble non-alien states (i) only with respect to consciousness* and (ii) not with respect to consciousness in the sense relevant to SP (the phenomenal sense).

It has often been noted that the notion of consciousness is multiply ambiguous. As Block (1995: 227) puts it, "there are a number of very different 'consciousnesses'... These concepts are often partly or totally conflated, with bad results." It may thus be that alien states are conscious in one sense of the term but not another. We might thus invoke a form of ambiguity here as well, denying that alien states are conscious in the sense relevant to subjectivity theories (*phenomenal* consciousness), even if conscious in some other (non-phenomenal) sense – the 'conscious*' sense.

Consider that patients' reports indicate they are aware of their alien states in a way similar to the way we are aware of our phenomenally conscious states. In particular, they do not have to make any effort to access their alien states: the latter somehow impose themselves on them (otherwise they would probably not *bother* subjects). We might put this by saying that patients' access to their alien states seems "immediate." Arguably, there is a sense of 'consciousness' whereby any mental state to which one has a seemingly immediate access is definitionally conscious. We might call this the *reflective* sense of consciousness.[20] It might then be argued that reflective consciousness (=consciousness*) does not a priori entail phenomenal consciousness (=consciousness),

and that even though alien states are conscious in the reflective sense, they are not conscious in the phenomenal sense.[21] The idea here is not that the subject has no phenomenology at all when in the alien state. Rather, the state itself does not *contribute* to the subject's overall phenomenology at the time. The alien state has no phenomenal character of its own, but its subject has simultaneous mental states which do. Among those is a mental state that represents her alien state as "in her". In other words, the subject has (i) a first-order state M_1 that is alien but not phenomenal and (ii) a second-order state M_2 that is non-alien but phenomenal and represents M_1. The fact that M_2 occurs and is phenomenal accounts for the subject's phenomenal awareness of M_1, rendering the latter reflectively conscious; however, it does *not* render M_1 *phenomenally* conscious.

To make the hypothesis more vivid, imagine the following situation. You wake up one morning and start hearing a tinnitus ringing inside your head. At first, this is just a meaningless buzz, but with time something strange happens: it begins to sound like an articulated voice. Even stranger, the voice seems to express your repressed, unconscious states. When you stand in front of someone you have reasons to detest, it says "Oh! I hate him!" It turns out, you discover, that a mad neuroscientist has implanted in your brain a small monitoring device that scans your unconscious states and expresses them through inner voices. In this case, then, you are aware of some of your phenomenally unconscious states, and moreover the awareness is phenomenally conscious (having a distinctive auditory phenomenology). Accordingly, even though the reflected-on states have no phenomenology of their own, the reflecting states do and thus give rise to a very rich phenomenology (one you might describe as an "inner voice"). Plausibly, even if you know that the reflected-on states are *yours*, you do not *experience* them that way. On the contrary, you experience the voice as alien.

This reflective awareness response is defended by Billon (2013) for thought insertion. A partisan of this response must not only show that it is consistent with the data, but also put forward independent evidence for the claim that inserted thoughts lack phenomenality. Billon appeals to the fact that stock descriptions of thought insertions readily compare the inserted thought to representations lacking intrinsic phenomenality.

Some patients talk of pictures being flashed in their mind, others of mere pieces of information, or more frequently, of voices. All of those are representations of which we can be aware thanks to "second-order" states, but which have no phenomenality in and of themselves. Relatedly, Billon appeals to the fact that this hypothesis would neatly explain certain experimental data, such as the well-documented phenomenological and neurobiological continuity between thought insertion and alien voices (Miller 1996; Moritz & Larøi 2008; Copolov et al. 2004; see Billon 2013: 309). Admittedly, the patients we know of do not explicitly say that their inserted thoughts are unconscious in some sense. In order to say such a thing (without exposing themselves to the risk of contradiction), however, patients would need to be clear on the different senses of 'consciousness'. And we cannot reasonably expect them to spontaneously master some conceptual distinctions that philosophers have just started to draw rigorously.

The reflective-consciousness response can be extended to depersonalization. Indeed, it seems to work even better here, given that some depersonalized patients explicitly affirm that their alien states are unconscious even though they are conscious *of* them:

> 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… in order to realize that I am making the walking movements. So at some point during this kind of crisis, before the absolute certainty [of being myself, before the crisis] I am conscious on one side that I am unconscious (sic) on the other side. (Séglas and Meige 1895: 147)[22]

Some suggest that there is nothing it is like for them to see anymore: "Everything in vision is dead," says a patient from Mayer-Gross (2011: 111). Or even that there is nothing it is like to *perceive*: "I can see, hear and smell but it is as if I didn't see, or hear or smell" (Dugas and Moutier 1911: 10).[23]

Even more commonly, depersonalized patients complain that their *emotional* phenomenology is blunted or absent:

> The emotional part of my brain is dead.... All my emotions are blunted. (Shorvon 1946: 784)

> Kissing my husband is like kissing a table, sir. The same thing... Not the least thrill. Nothing on earth can thrill me. Neither my husband nor my child… My heart doesn't beat. I cannot feel anything. (Dugas and Moutier 1911: 109)

When depersonalization reaches its climax, subjects may describe themselves as totally unconscious, indeed as zombies:

> 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)
>
> *I'm like a zombie* unable to take in any information. (Sierra 2009: 51; our emphasis)
>
> It 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…. I am in emptiness, I am a body without a soul (...) I see without seeing, I am a blind man who sees (Janet 1928: 51-2).
>
> It's the mental sensibility 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)[24]

Thus depersonalization is particularly amenable to a consciousness* response. To be sure, here as elsewhere we may choose to simply refuse to accept patients' reports, rather than try to accommodate them "as is" within a subjectivist framework. Our present point is that this is not *mandatory*: there are ways to retain SP while making sense of patients' reports more or less at face value.

The consciousness* response, in this reflective-awareness version, protects SP from its putative psychopathological counterexamples, then. Admittedly, it does so at the cost of denying that states of which we are reflectively aware must be phenomenally conscious. This is a heavy cost for most subjectivity theories, but importantly, it is not strictly inconsistent with such theories. The response fits particularly well with *acquaintance* theories: it can be claimed that the patient is reflectively aware of her alien state but that, since reflective awareness involves a representational relation rather than the requisite acquaintance relation, the patient's alien state is not phenomenally conscious. It may also suit the self-representational approach: it might be claimed that reflective awareness involves a numerically distinct higher-order representation rather than the kind of self-representation required. As a result, the subject is reflectively aware of her alien state without the latter being phenomenally conscious. The reflective-awareness response may even be consistent with the higher-order representationalism. It may be held that not all higher-order representations yield phenomenal consciousness, only ones exhibiting certain specific features (e.g., being non-inferential), and that the

higher-order representation involved in patients' reflective awareness of her alien states lacks (some of) the relevant features.[25]

Conclusion

Jaspers' dilemma poses a psychopathological challenge to subjectivity theories of consciousness. In this face of this challenge, it is possible of course to simply renounce subjectivity theories. Philosophers who have chosen to hold on to such theories have tended to respond to the challenge by dismissing patients' reports as unintelligible, or else by radically reinterpreting them. In this paper, we have attempted to articulate two new ways of defending subjectivity theories, characterized by a stronger desire to make sense of patients' reports while taking them more or less at face value. Our conviction is that it is *possible* to respect patients' rationality, sensibility, and intelligibility while holding on to subjectivity theories of consciousness in the face of what we have called the Alienation Argument. The "something extra" version of the subjectivity* response and the reflective-consciousness version of the consciousness* response offer, we claim, plausible ways of doing so. Importantly, the subjectivity theorist need not be wedded to a single approach for relevant psychopathological phenomena. A divide-and-conquer strategy that handles different phenomena in different ways is entirely coherent, perhaps even antecedently plausible. Our goal here has been to present a more comprehensive menu of options from which the subjectivity theorist might choose to address each challenging psychopathology.[26]

References

Amiel, H.-F. (1881). *Amiel's Journal: The Journal Intime of Henri-Frédéric Amiel*. London: MacMillan, 1913.

Aristotle. *Metaphysics*, trans. W.D. Ross. Oxford: Clarendon Press, 1908.

Billon, A. (2005). *En Personne, la réalité subjective de la conscience phénoménale*.

PhD Dissertation, École Polytechnique.

Billon, A. (2013). Does consciousness entail subjectivity? the puzzle of thought insertion. *Philosophical Psychology* 26: 291-314.

Bortolotti, L. (2010). *Delusions and other irrational beliefs*. Oxford University Press.

Bortolotti, L. and Broome, M. (2009). A Role for Ownership and Authorship in the Analysis of Thought Insertion. *Phenomenology and the Cognitive Sciences* 8: 205-224.

Bottini, G., Bisiach, E., Sterzi, R., and Vallar, G. (2002). Feeling touches in someone else's hand. *Neuroreport* 13: 249-252.

Brentano, F.C. (1874). *Psychology from Empirical Standpoint*, trans. A.C. Rancurello, D.B. Terrell, and L.L. McAlister. London: Routledge and Kegan Paul, 1973.

Campbell, J. (1998). Le modèle de la schizophrénie de Christopher Frith. In *Subjectivité et Conscience d'Agir : Approches Cognitive et Clinique de la Psychose*. Paris: PUF.

Campbell, J. (1999). Schizophrenia, the space of reasons and thinking as a motor process. *The Monist* 82: 609-625.

Campbell, J. (2004). The ownership of thoughts. *Philosophy, Psychology, Psychiatry*, 9: 35-39.

Coliva, A. (2002). Thought insertion and immunity to error though misidentification. *Philosophy, Psychiatry & Psychology* 9: 27-34.

Copolov, D., Trauer, T., and Mackinnon, A. (2004). On the non-significance of internal versusexternal auditory hallucinations. Schizophrenia Research 69: 1-6.

Dugas, L. and Moutier, F. (1911). *La dépersonnalisation*. Paris: F. Alcan. available at http://www.biusante.parisdescartes.fr/histmed/medica/cote?79749.

Eisen JL, Beer DA, Pato MT, Venditto TA, Rasmussen SA. Obsessive-compulsive disorder in patients with schizophrenia or schizoa-ective disorder. American Journal of Psychiatry 1997;154:271-3.

Fernández, J. (2010). Thought Insertion and Self-Knowledge. *Mind and Language* 25: 66-88.

Frith, C. D. (1992). *The Cognitive Neuropsychology of Schizophrenia*. Lawrence Erlbaum Associates.

Gallagher, S. (2000). Self-reference and schizophrenia: A cognitive model of immunity to error through misidentification. In Zahavi, D., editor, *Exploring the Self: Philosophical and Psychopathological Perspectives on Self-Experience*. Amsterdam: John Benjamins.

Gennaro, R.J. 1996. *Consciousness and Self-Consciousness*. Philadelphia and Amsterdam: John Benjamins.

Gennaro, R.J. 2012. *The Consciousness Paradox*. Cambridge MA: MIT Press.

Graham, G. and Stephens, G. L. (2000). *When self consciousness breaks, alien voices and inserted thoughts*. Cambridge MA: MIT Press.

Hoffman, R. E. (1986). Verbal hallucinations and language production processes in schizophrenia. Behavioral and Brain Sciences 9: 503-548.

Janet, P. (1908). Le sentiment de dèpersonnalisation. *Journal de Psychologie normale et pathologique* 5: 514-516.

Janet, P. (1928). *De l'angoisse à l'extase, vol. 2 (Les sentiments fondamentaux)*. Paris: F. Alcan. available at http://classiques.uqac.ca/classiques/janet_pierre/angoisse_extase_2/Janet_angoisse_2_1.pdf.

Jaspers, K. (1913). *General Psychopathology*, trans. J. Hoenig and M.W. Hamilton (seventh edition). Manchester: Manchester University Press, 1963.

Husserl, E. (1928). *On the Phenomenology of the Consciousness of Internal Time*, trans. J. Brough. Dordrecht: Kluwer Academic Publishers, 1991.

Kemp, R., Chua, S., McKenna, P., and David, A. (1997). Reasoning and delusions. *The British Journal of Psychiatry* 170: 398-405.

Krishaber, M. (1873). *De la névropathie cérébro-cardiaque*. Paris: Masson. available at http://gallica.bnf.fr/ark:/12148/bpt6k767014.

Liang, C. and Lane, T. (2009). Higher-order thought and pathological self: The case of somatoparaphrenia. *Analysis* 69:661-668.

Locke, J., *An Essay Concerning Human Understanding*, ed. P. H. Nidditch, Oxford: Oxford University Press, 1975.

Mayer-Gross, W. (2011). On depersonalization. *British Journal of Medical Psychology* 15:103-126.

Miller, L. J. (1996). Qualitative changes in hallucinations. The American Journal of Psychiatry 153: 265-267.

Mirian, D., Heinrichs, R. W., and Vaz, S. M. (2011). Exploring logical reasoning abilities in schizophrenia patients. *Schizophrenia Research* 127:178-180.

Moritz, S., & Larøi, F. (2008). Differences and similarities in the sensory and cognitive signatures of voice-hearing, intrusions and thoughts. Schizophrenia Research 102: 96-107.

Moro, V., Zampini, M., and Aglioti, S. (2004). Changes in spatial position of hands

modify tactile extinction but not disownership of contralesional hand in two right brain-damaged patients. *Neurocase* 10:437-443.

Mullins, S., & Spence, S. A. (2003). Re-examining thought insertion semi-structured literature review and conceptual analysis. The British Journal of Psychiatry, 182, 293–298.

de Nerval, G. (1855). *Aurelia and other Writings*. Boston: Exact Change, 1996.

O'Shaughnessy, B. (2003). The Epistemology of Physical Action. In Roessler, J. and Eilan, N. (eds.), *Agency and Self-Awareness*. Oxford: Clarendon.

Purdon, C. (2004). Empirical investigations of thought suppression in OCD. *Journal of Behavior Therapy and Experimental Psychiatry* 35: 121-136.

Robinson, S. W. (2006). Thoughts without Distinctive Non-Imagistic Phenomenology. *Philosophy and Phenomenological Research* 70: 534-561.

Rosenthal, D. M. (1990). 'A theory of consciousness.' ZiF Technical Report 40, Bielfield, Germany. Reprinted in N.J. Block, O. Flanagan, and G. Güzeldere (eds.), *The Nature of Consciousness: Philosophical Debates*. Cambridge MA: MIT Press, 1997.

Rosenthal, D. M. (2005). Consciousness and Mind. New York: Oxford University Press.

Rosenthal, D. M. (2010). Consciousness, the self and bodily location. *Analysis*, 70:270-276.

Salkovskis, P. (1989). Cognitive-behavioural factors and the persistence of intrusive thoughts in obsessional problems. Behaviour Research and Therapy 27: 677–682.

Sass, L., Pienkos, E., Nelson, B., and Medford, N. (2013). Anomalous self-experience in depersonalization and schizophrenia: A comparative investigation. *Consciousness and Cognition* 22:430-441.

Séglas, J. and Meige, H. (1895). *Leçons cliniques sur les maladies mentales et nerveuses:(Salpêtrière, 1887-1894)*. Asselin et Houzeau. available at http://gallica.bnf.fr/ark:/12148/bpt6k76631s.

Shorvon, H. (1946). The depersonalization syndrome. *Proceedings of the Royal Society of Medicine* 39:779–791.

Sierra, M. (2009). *Depersonalization: a new look at a neglected syndrome*. Cambridge: Cambridge University Press.

Sierra, M. and Berrios, G. E. (1998). Depersonalization: neurobiological perspectives. *Biological Psychiatry* 44:898-908.

Sierra, M. and Berrios, G. E. (2000). The Cambridge depersonalisation scale: a new instrument for the measurement of depersonalisation. *Psychiatry Research* 93:153-164.

Simeon, D. and Abugel, J. (2006). *Feeling unreal: Depersonalization disorder and the loss of the self*. Oxford and New York: Oxford University Press.

Sousa, P., & Swiney, L. (2013). Thought insertion: Abnormal sense of thought agency or thought endorsement? *Phenomenology and the Cognitive Sciences* 12: 637-654.

de Vignemont, F. (2007). Habeas corpus: The sense of ownership of one's own body. *Mind and Language* 22: 427-449.

Williford, K.W. (2006). The self-representational structure of consciousness. In U. Kriegel and K.W. Williford (eds.), *Self-Representational Approaches to Consciousness*. Cambridge, MA: MIT Press.

Zahavi, D. and Kriegel, U. (ms). For-me-ness: What It Is and What It Is Not.

[1] Thus, two subjects who differ only in that one higher-order represents his state as phenomenally F whereas the other higher-order represents his as phenomenally G will experience different overall phenomenologies.

[2] Rosenthal writes: "I cannot represent my conscious pain as belonging to someone distinct from me" (Rosenthal 2005: 357); "being conscious of a mental state as belonging to someone other than oneself would plainly not make it conscious" (2005: 342).

[3] Delusions of alien control and thought insertion are diagnostic symptoms of schizophrenia. Other symptoms include alien voices (patients report hearing voices), thought broadcasting (reports of thoughts being publicly accessible), thought control (reports of thoughts being controlled by an alien agency), and thought withdrawal (the patient reports that his thought suddenly vanishes, as if an external agency pulled it out of his mind).

[4] See for example Hoffman (1986: 508).

[5] Jaspers could grant the consistency of the patients. But he took it that making sense of someone required the capacity to project imaginatively being in that person's shoes, and to simulate the experiences of that person. He thought that such an imaginative projection was impossible with schizophrenics suffering from the relevant symptoms. Given the independent reasons to believe that patients' reports do make sense, however, it seems that the imagination criterion might be too strong. See Eilan (2000) for more on this.

[6] Some patients do exhibit other irrational biases, such as jumping to the conclusions, but that alone does not threaten their intelligibility (see Bortolotti 2010 for a book-length defense of this claim).

[7] Dugas borrowed the term 'depersonalization' from the Swiss writer Amiel, who seemed to suffer from the condition and described himself in his diaries as follows: "now I find myself regarding existence as though from beyond the tomb, from another world; all is strange to me; I am, as it were, outside my own body and individuality; I am *depersonalized*, detached, cut adrift." (Amiel 1881). He would later construe depersonalization as a withdrawal of the subjectivity from all mental states (Dugas and Moutier, 1911:13-4), which explains Jaspers' usage.

[8] Indeed, the aforementioned impression of lacking bodily parts may be explained in terms of bodily sensations that lack subjectivity.

[9] See Dugas and Moutier (1911: 10-11). So-called Cotard Syndrome is often said to be the delusional counterpart of depersonalization, where patients do endorse the same feelings and take the relevant appearances at face value. We follow modern classifications (ICD-10, DSN-IV) in keeping the term "depersonalization" for non-delusional patients. It should be noted, however, that some researchers, such as Janet, used the term to refer to delusional (Cotard) patients as well. There are borderline cases between delusional (Cotard) and non-delusional (depersonalized) versions, but there are also clear-cut cases of non-delusional patients.

[10] Patients suffering from phantom limbs also seem to, in some sense, experience pain outside their body (Rosenthal 2010). More generally, a broad set of empirical data suggests that a region of space in which someone feels bodily sensations will not always be represented as belonging to one's body: it will only be so represented if it falls within the subject's "body schema" (de Vignemont 2007).

[11] Importantly, Premise 1 of the Alienation Argument does not claim that the relevant patients *have* alien states, only that they *report* having that. To object to this premise is therefore to deny that they so report. But this is patently implausible.

[12] It is also true that some patients, whose degree of confidence in their non-existence often remains low, seem to be neither clearly delusional nor clearly non-delusional.

[13] Sousa and Swiney (2013: 637) refer to it, justifiably we think, as "the standard approach to the core phenomenology of thought insertion".

[14] Similarly, patients suffering from schizophrenia can complain of "controlled thoughts", which they take both to be the their own and to be under the control of an external agency (Jaspers 1913: 122-3, Mullins and Stephens 2001). Like intrusive thoughts, controlled thoughts do not come with a sense of agency, and it is hard to see how the agency view of thought insertion could distinguish them from thought insertion (Billon 2013).

[15] To distinguish the phenomenology of inserted thoughts from that of intrusive thoughts, many appeal to a form of psychological discontinuity. Intrusive thoughts, unlike inserted thoughts, feel somehow psychologically continuous with the patient's other thoughts. This psychological discontinuity can be construed as a discrepancy between the content of the thought and the subject's self-view (Graham and Stephens 2000: 173), or "the subject's long-standing beliefs and desires" (Campbell 1999: 621), or her "implicit expectancies" (Gallagher 2000). However, studies of obsessive phenomena, including intrusive thoughts, reveal that the latter's content can perfectly mirror that of inserted thoughts and feel discontinuous with the subject's self-view, long-standing beliefs and desires, and implicit expectancies (see Billon 2013: 296-8). Conversely, Hoerl (2011: 189-190) provides examples of inserted thoughts which the patients seem to acknowledge as psychologically continuous ("it feels pretty normal or fits what I suspect," "my own thoughts might say the same thing").

[16] Some use the term "sense of authorship" for "sense of endorsement" but the former has also been used as a synonym for "sense of agency," so we avoid it here.

[17] Proponents of the endorsement approach may not have a subjectivist agenda. Bortolotti and Broome (2009) certainly claim that inserted thoughts are not experienced by the patients as their own, though Fernandez (2010) seems to imply that they are subjective in our sense. In any case, our present interest is in whether the endorsement approach could be harnessed to the subjectivist agenda.

[18] This kind of fear does seem to play an important role, however, in causing the recurrence of the thought (Salkovskis 1989).

[19] Note, in this connection, that schizophrenics' memoirs of crisis and institutionalization episodes often convey a sense of new abundance and freedom in their inner life. The mid-nineteenth-century French poet Nerval declares this at the beginning of his autobiography, which focuses on his inner life during an eight-month institutionalization: "I will try… to transcribe the impressions of a long illness which took place entirely in my mind's mysteries; – and I am unsure why I use this term illness, since never, as far as I am concerned, did I feel in better health. Sometimes I felt my power and activity has doubled; it seemed to know everything, understand everything; imagination brought me infinite delights." (Nerval 1855: 3) Nerval's autobiography proceeds to describe many episodes of what we would now conceptualize as thought insertion.

[20] Using Block (1995) distinctions, we might say that a state is conscious in the reflective awareness sense when its subject is access-conscious *of* it, or when it is reflectively-conscious.

[21] Notice that even higher-order representationalists can accommodate the claim that reflective awareness does not entail phenomenal consciousness. They can claim that, to make M phenomenally conscious, my awareness of M need not only to (i) seem immediate in the sense that we have specified, but also to (ii) display other specific features.

[22] Notice the "sic" added by the psychiatrists, who are not sure how to understand this "consciousness of unconsciousness".

[23] Interestingly, some patients say that things appear to them as through a "curtain", a "blind", a "fine wire netting", a "fine mesh" or a "glass wall" (Shorvon 1946, 784). This suggests that the phenomenality of their visual states is in some sense *attenuated*, as if only some of its standard phénoménal features are present. It is an open question how to interpret such reports of attenuated phenomenality, and how they might affect the dialectic. Here we bracket such questions.

[24] Note well: by "mental sensibility" ("la sensibilité morale" in the original), people at the time meant roughly the faculty responsible for conscious thoughts in general.

[25] This does require the higher-order representationalist to identify a feature that can be plausibly shown to be missing in the relevant psychopathological cases. This kind of additional burden is absent, however, in self-representational and acquaintance theories.

[26] For comments on a previous draft, we would like to thank Rocco Gennaro and Tim Lane. Work for this chapter was supported by grants ANR-10-IDEX-0001-02 PSL* and ANR-10-LABX-0087 IEC.