Cotard syndrome, self-awareness, and I-concepts

Rocco J. Gennaro (rjgennaro@usi.edu)

Abstract
Various psychopathologies of self-awareness, such as somatoparaphrenia and thought insertion in schizophrenia, might seem to threaten the viability of the higher-order thought (HOT) theory of consciousness since it requires a HOT about one’s own mental state to accompany every conscious state. The HOT theory of consciousness says that what makes a mental state a conscious mental state is that there is a HOT to the effect that “I am in mental state M.” I have argued in previous work that a HOT theorist can adequately respond to this concern with respect to somatoparaphrenia and thought insertion. There is also Cotard syndrome which is a rare neuropsychiatric disorder in which people hold the delusional belief that they are dead, do not exist, or have lost their blood or internal organs. In this paper, I argue that HOT theory has nothing to fear from it either and can consistently account for what happens in such unusual cases. I analyze Cotard syndrome in light of my previous discussion of somatoparaphrenia and thought insertion, and argue that HOT theory can provide a somewhat analogous account without the worry of inconsistency. It is crucial to recognize that there are multiple “self-concepts” and levels of HOTs which can help to provide a more nuanced explanation. With regard to the connection between consciousness and self-consciousness, it is proposed that Cotard patients are indeed capable of having some “I-thoughts” about their bodies and mental states.

Keywords
Consciousness ∙ Cotard syndrome ∙ Higher-order thoughts ∙ I-concepts ∙ Self-awareness ∙ Somatoparaphrenia ∙ Thought insertion

1 Introduction

Various psychopathologies of self-awareness, such as somatoparaphrenia and thought insertion in schizophrenia, might seem to threaten the viability of...
the higher-order thought (HOT) theory of consciousness since it requires a HOT about one's own mental state to accompany every conscious state (which might also be viewed as some form of self-consciousness). The HOT theory of consciousness says that what makes a mental state a conscious mental state is that there is a HOT to the effect that “I am in mental state M” (Rosenthal, 2005). I have argued in previous work that a HOT theorist can adequately respond to this concern with respect to somatoparaphrenia and thought insertion (Gennaro, 2015, 2020). Somatoparaphrenia is a monothematic delusion and a disorder of self-awareness characterized by the sense of alienation from parts of one’s body (Vallar & Ronchi, 2009). It is a bizarre body delusion where one denies ownership of a limb or an entire side of one’s body. Schizophrenia is a mental disorder which most commonly manifests itself through auditory hallucinations, paranoid or bizarre delusions, or disorganized speech and thinking. Thought insertion is the delusion that some thoughts are not “one’s own” in some sense and are inserted into one’s mind by someone else (Frith, 1992). Thus, it may seem that a HOT with a self-referential “I” in its content could not be present.

There is also Cotard syndrome which is a rare neuropsychiatric disorder in which people hold the delusional belief that they are dead (either figuratively or literally), do not exist, are putrefying, or have lost their blood or internal organs (Billon, 2016). Once again, there seems to be a stunning abnormality of self-awareness in these patients which may also shed light on the very connection between conscious states and self-consciousness. Although it will be useful to compare and contrast Cotard syndrome with somatoparaphrenia and thought insertion, I will argue that HOT theory also has nothing to fear from it and can consistently account for what happens in such unusual cases. More specifically, in section 2, I explain HOT theory emphasizing the features which play a prominent role in later sections. In section 3, I present a relatively brief discussion of somatoparaphrenia and thought insertion in order to set the stage for addressing Cotard syndrome. In section 4, I analyze Cotard syndrome in light of the previous section and argue that HOT theory can provide a somewhat analogous account without fear of inconsistency. It will also be crucial to recognize that there are multiple “self-concepts” (that is, “I-concepts”) and levels of HOTs which can help to provide a more nuanced explanation of the relationship between conscious mental states and self-consciousness. In section 5, I shift to a related and important ambiguity in self-awareness with respect to the problem of personal identity. In section 6, I offer some concluding thoughts.

2 HOT theory

One question that should be answered by any theory of consciousness is: What makes a mental state a conscious mental state? There is a long tradition that has attempted to understand consciousness in terms of some kind of higher-order awareness. For example, John Locke (1689) once said that “consciousness is the percep-
tion of what passes in a man’s own mind.” This intuition has been revived by a number of contemporary philosophers (Lycan, 1996; Rosenthal, 1997, 2005). The general idea is that what makes a mental state conscious is that it is the object of some kind of higher-order representation (HOR). A mental state M becomes conscious when there is a HOR of M. So, for example, my desire to drink some water becomes conscious when I am (non-inferentially) “aware” of the desire. Intuitively, it seems that conscious states, as opposed to unconscious ones, are mental states that I am “aware of” in some sense. This is sometimes referred to as the Transitivity Principle (TP):

**(TP)** A conscious state is a state whose subject is, in some way, aware of being in.

Conversely, the idea that I could be having a conscious state while totally unaware of being in that state seems odd or perhaps even contradictory. A mental state of which the subject is completely unaware is clearly an unconscious state. For example, I would not be aware of having a subliminal perception and thus it is an unconscious perception. Any theory which attempts to explain consciousness in terms of higher-order states is known as a higher-order representational theory of consciousness. The most common division of HOR theories is between higher-order thought (HOT) theories and higher-order perception (HOP) theories. HOT theorists, such as David Rosenthal, think it is better to understand the HOR as thoughts involving some kind of conceptual component. HOP theorists urge that the HOR is a perceptual or experiential state of some kind which does not require the kind of conceptual content invoked by HOT theorists (Lycan, 1996). Although HOT and HOP theorists agree on the need for a HOR theory of consciousness, they disagree on the superiority of their respective positions (such as in Rosenthal, 2004). I focus on HOT theory here because I have defended the theory in previous publications, but also because it is the HOT concepts that generate the apparent objections and HOT theory lends itself to a far more interesting discussion regarding varieties of self-awareness.

It should be noted that higher-order thought theorists agree that one must become aware of the lower-order (LO) mental state noninferentially. We might suppose with Rosenthal, for example, that the HOT must be caused noninferentially by the LO state in order to make it conscious. The subject is directly aware of the LO state. The point of this condition is mainly to rule out some alleged counterexamples to HOT theory, such as cases where I become aware of my unconscious desire to kill my boss because I have consciously inferred it from a session with a psychiatrist, or where my envy becomes conscious after making inferences based on my own behavior. The characteristic feel of such a conscious desire or envy may be absent in these cases, but since awareness of them arose via conscious inference, the HOT theorist accounts for them by adding this noninferential condition.

It might seem that HOT theory results in circularity by defining consciousness in terms of HOTs. It also might seem that an infinite regress results because a conscious mental state must be accompanied by a HOT, which, in turn, must be...
accompanied by another HOT ad infinitum. However, the widely accepted reply is that when a conscious mental state is a first-order world-directed state the higher-order thought (HOT) is not itself conscious. When the HOT is itself conscious, there is a yet higher-order (or third-order) thought directed at the second-order state. In this case, we have introspection which involves a conscious HOT directed at an inner mental state. When one introspects, one’s attention is directed back into one’s mind. As we shall see, this distinction is crucial to the discussion which follows.

It is worth mentioning that most versions of representationalism are reductionistic regarding consciousness which, for many, is part of the attraction. One motivation for representationalism is that a naturalistic account of intentionality or representation can arguably be more easily attained, such as via causal theories whereby mental states are understood as representing outer objects by virtue of some reliable causal connection. The idea, then, is that if consciousness can be explained in representational terms and representation can be understood in purely physical terms, then there is the promise of a naturalistic theory of consciousness. Most representationalists think that there is then room for a “second-step” reduction to be filled in later by neuroscience. HOT theorists also tend to endorse this line of argument (Gennaro, 2012, chapter two).

3 Somatoparaphrenia and thought insertion

Somatoparaphrenia is a bizarre delusion characterized by the sense of alienation from parts of one’s body (Vallar & Ronchi, 2009). Patients with it often deny ownership of a limb or an entire side of their body. Somatoparaphrenia might seem to threaten the viability of the higher-order thought (HOT) theory of consciousness since it requires a HOT about one’s own mental state to accompany every conscious state. Liang and Lane (2009), for example, initially argued that at least one specific case of somatoparaphrenia is problematic for HOT theory because it contradicts the notion that, when I am in a conscious state, I must have the accompanying HOT that “I am in mental state M.” The “I” is not only importantly self-referential but essential in tying the conscious state to oneself and thus to one’s ownership of M.

Rosenthal (2010) responds that one can be aware of bodily sensations in two ways which, normally at least, go together: (a) awareness of a bodily sensation as one’s own, and (b) awareness of a bodily sensation as having some bodily location, like a hand or foot. Patients with somatoparaphrenia still experience the sensation as their own but also as having a mistaken bodily location. These patients still do have the awareness in (a), which is the main issue at hand, but they have a very strange awareness in sense (b) which leads them to misidentify the bodily location of the sensation in someone else. Lane and Liang (2010) protest that Rosenthal has still not explained why the identification of the bearer of the sensation cannot also go astray in somatoparaphrenia, especially since Rosenthal clearly holds that
misrepresentation can otherwise occur between a HOT and its target mental state.

A HOT theorist might also argue that there are really two conscious states that seem to be at odds (Gennaro, 2015). There is a conscious feeling in a limb but also the (conscious) attribution of the limb to someone else. It is crucial to emphasize that somatoparaphrenia is often characterized as a delusion of belief under the broader category of anosognosia (Prigatano, 2010). A delusion is often defined as a false belief that is held based on an incorrect (and probably unconscious) inference about external reality or oneself that is firmly sustained despite what almost everyone else believes and despite what constitutes incontrovertible and obvious proof or evidence to the contrary (Bortolotti, 2009). Beliefs are themselves often taken to be intentional states integrated with other beliefs and mental states. They are typically understood to be caused by perceptions or experiences which, in turn, lead to action or behavior. Thus, somatoparaphrenia seems closer to self-deception and involves frequent confabulation. If this is a reasonable interpretation, then a HOT theorist can argue that the patient actually has the following two conscious states:

S1: a conscious feeling (i.e., a tactile sensation) in the limb in question, and

S2: a conscious belief that the limb (and thus sensation) belongs to someone else.

Having both S1 and S2, especially if both conscious at any given time, is indeed strange and perhaps even contradictory in some sense but the puzzlement has nothing to do with HOT theory itself (but see also Lane, 2015, for much more on this debate). The patient would simply have a different unconscious HOT directed at S1 and S2.

Schizophrenia is a mental disorder which most commonly manifests itself through auditory hallucinations, paranoid or bizarre delusions, or disorganized speech and thinking. Thought insertion is the delusion that some thoughts are not “one’s own” in some sense or are being inserted into one’s mind by someone else (Frith, 1992). The puzzle of thought insertion is thus perhaps somewhat analogous to somatoparaphrenia. For example, just as we might distinguish between experiencing a sensation as one’s own as opposed to its bodily location in patients with somatoparaphrenia, so we might distinguish between experiencing a thought as one’s own as opposed to its causal origin in patients with schizophrenia. Those with somatoparaphrenia attribute a limb or a sensation to someone else whereas some schizophrenics attribute thoughts (or at least their causal origin) to someone else. If this rough similarity is correct, then a HOT theorist might argue that the schizophrenic patient has the following two conscious states:

S1: a conscious thought in one’s own mind, and

S2: a conscious belief that the thought in S1 has been inserted by someone else.
Having both S1 and S2, especially if both are conscious at any given time, is again very strange and perhaps even irrational, but the puzzlement is not a problem for HOT theory. In S1, we can suppose that the patient has a HOT about her own first-order conscious thought, i.e. she is aware of a mental state occurring in her mind. In S2, the patient also has a conscious belief that S1 has been inserted into her mind by someone else. Once again, and analogous to somatoparaphrenia, there is an essentially delusional element involved in schizophrenia and inserted thoughts. Why or how S2 is generated is of course an interesting and difficult question in its own right. For example, as Stephens and Graham (2000) suppose, perhaps the best explanation is that S2 is consciously or unconsciously inferred from a lack of sense of agency (or “passivity” experience) or perhaps it is inferred from a perceived lack of causal origin. But none of the alternative explanations cause trouble specifically for HOT theory’s ability to explain S1. Just as HOT theory says about all conscious states, when a patient is aware of having a (first-order) conscious thought, there is a HOT about that thought (Gennaro, 2020).

In both of the above disorders, there is a conscious thought (S1) which can be explained by a HOT theorist in the usual way, but there is a second belief or thought which apparently contradicts the first one (S2). Recall also that HOT theory comes with a well-known noninferentiality condition, such that a HOT must involve awareness of its target mental state noninferentially, that is, in an unmediated way. This is important because if there is any kind of conscious inference to a belief, such as in S2, then the HOT would not arise in the requisite manner for its target mental state to be conscious. More importantly, however, using conscious inference and reasoning to become aware of a mental state is much more likely to occur via introspection, that is, by consciously thinking about one’s own mental states. Notice that S2 is more of an introspective state with a belief directed at a mental state. This, in turn, increases the likelihood that delusional thinking will generate further false beliefs.

Notice that this also fits nicely with HOT theory which can explain why there would seem to be a phenomenological sense of “myness” or “ownership” when one introspects, namely, that the HOT is itself conscious. By contrast, it is doubtful that such a sense is present when one has an unconscious HOT. The concept “I” is part of a conscious HOT in the introspective case but it is part of an unconscious thought in the first-order case. Nonetheless, it is certainly true that when there is a disturbance or abnormality in one’s I-concept, such as one’s bodily representation, one’s consciousness will be altered and result in some odd feelings of body or thought disownership.

With regard to any potential misrepresentation between a HOT and its target, recall the worry that there could equally be the possibility of a mismatch between the “I” in the HOT and the “I” in the first-order mental state. But it is unclear how this could be so. Consider Wittgenstein’s (1958) distinction between the “I-as-subject” (e.g., “I have a pain”) and the “I-as-object” (“I have a broken arm”). Note that what he called the “I-as-subject” seems to involve some kind of minimal re-
Cotard syndrome is a rare neuropsychiatric disorder in which people hold the delusional belief that they are dead, do not exist, are putrefying, or have lost their blood or internal organs (Billon, 2016). Importantly, most patients also have severe depression and anxiety problems, not to mention the way that they describe their self-related and very odd experiences. Some bizarre patient accounts are as follows:

She had the constant experience of having no identity or “self” and being only a body without content....she was convinced that her brain had vanished, her intestines had disappeared, and her whole body was translucent.
(Debruyne, Portzky, Van den Eynde, & Audenaert, 2009, p. 197)

He said "I speak, breathe and eat, but I am dead".
(Nejad & Toofani, 2005, p. 250)

It was as if it was not me walking, it was not me talking, as if it was not me living [...] I can look at me, I am somehow bothered by my body, as if it wasn’t me, as if I lived on the side of my body, on the side of myself if you like. I don’t know how to explain.
(Janet & Raymond, 1898, p. 70)

4 Cotard syndrome

Cotard syndrome is a rare neuropsychiatric disorder in which people hold the delusional belief that they are dead, do not exist, are putrefying, or have lost their blood or internal organs (Billon, 2016). Importantly, most patients also have severe depression and anxiety problems, not to mention the way that they describe their self-related and very odd experiences. Some bizarre patient accounts are as follows:
One general difficulty is just how literally to take the reports of Cotard patients. Should we take them at face value or treat them with some skepticism given the delusional aspect in question? Young and Leafhead (1996, p. 154), for example, note that there is no symptom or symptom complex present in all the patients described by Cotard and that this “gives grounds for arguing that the term ‘Cotard’s syndrome’ should be used cautiously; at best it represents an idealized pattern which in practice is not found even in pure cases.” Further, Cotard patients sometimes speak in more qualified terms, such as it is “as if” I do not exist or have a body (as above in the third quotation).

Like somatoparaphrenia and thought insertion, one might again suppose that Cotard syndrome threatens the HOT theory of consciousness. Recall that HOT theory requires that a HOT to the effect that “I am in mental state M” accompanies each conscious state. But if I believe that I don’t exist, for example, then it would seem that I do not really have a self-referencing HOT. Depending upon the specific Cotard patient, the “I-concept” would not seem to refer to her “self” or her “body” or her “mind.” So any analogy to thought insertion or somatoparaphrenia might depend to some extent on which quotation we focus on. When some aspect of one’s body is denied or feels alienated, perhaps it is more like somatoparaphrenia. When a mental state is denied or feels alienated, it is somewhat closer to thought insertion. It is difficult to grasp what it must be like to experience these abnormal conditions but we must also keep in mind the delusional element found in all three cases.

I will mainly focus here on cases where a patient apparently denies her own existence or “self,” sometimes called the “nihilistic delusion” (Radovic, 2017) which goes back originally to Jules Cotard (1882); see also (Metzinger, 2003). This is one difference between someone with Cotard syndrome and those with somatoparaphrenia and thought insertion. Billon (2016) argues that Cotard syndrome is the delusional counterpart of depersonalization disorders which are dissociative disorders marked by periods of feeling disconnected or detached from one’s body and thoughts. A Cotard patient also has an impaired self-awareness and the feeling that there is no “I” in some sense. Billon says that “…depersonalised patients are not delusional: they refuse to endorse the belief that they do not think and exist” (2016, p. 370) but he also points out that the “border between mere depersonalisation and the Cotard syndrome might not be always clear-cut” (2016, fn. 22). He then goes on to explain that “[g]iven the similarity between the experience of depersonalised patients and that of Cotard patients, what differentiates [them] is plausibly some kind of rationality deficit […]” (Billon, 2016, p. 380).

In any case, consider the following statements:

At times she would say, “I am no longer alive”; “I am dead”. (Enoch & Ball, 2001, p. 159)

He often said, “I am a ghost” or “I am dead”, and sometimes denied the existence of certain parts of his body saying, for example, “I have no blood”. (Enoch & Ball, 2001, p. 165)
[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 & Abigel, 2006, p. 15)

But notice first that even the statement that “I am dead” still contains an initial reference to oneself (at least as a grammatical fact). It seems to me that there will always be some reference to oneself when expressing any kind of thought or belief (even if delusional, irrational, or self-contradictory in some other way). Of course this is precisely what leads to an apparent contradiction. Imagine: “Who thinks that you don’t exist?” Answer: “I do”? “Who is dead?” Answer: “I am”? Again, perhaps this is merely a misleading way of speaking but the same might be true of other patient reports.

Still, we can see how the pattern of reply above to somatoparaphrenia and thought insertion can be applied here as well. That is, we may suppose that at least some Cotard patients have the following two conscious states:

**S1:** a conscious feeling or experience of unfamiliarity with myself (or body or mental states); and

**S2:** a conscious belief or thought that I am dead or I do not exist.

Here, once again, having both S1 and S2 is indeed strange and perhaps even self-contradictory in some sense, but it still wouldn’t tell against HOT theory. In S1, we can suppose that the patient has a HOT about her own feeling or experience, that is, she is aware of her feeling or mental state. In S2, the patient also has a conscious belief or thought that she does not exist.

How or why S2 is generated is of course interesting in its own right. For example, perhaps the best explanation is that S2 is simply generated by the unfamiliar feeling or experience described in S1. This would be a so-called “endorsement” account. But there are at least two possibilities here, namely, the endorsement and the explanationist theories of delusion formation. The endorsement account holds that delusions are formed by endorsing as veridical the content of the unusual experience (Bayne & Pacherie, 2004) whereas the explanationist account holds that delusions are formed as a way to explain (or rationalize) an unusual experience (Maher, 1999). One way to frame these two theories of delusion formation is by asking the following question: Is the delusional belief formed sometime prior to the person having the abnormal conscious experience or sometime afterward? Alternatively, we might ask: Are delusions bizarre convictions that alter one’s way of seeing the world, or are they hypotheses formulated to account for some unusual experiences and then endorsed as beliefs? In the previous discussion I have perhaps shown more sympathy with the explanationist account but, as Langdon and Bayne (2010) point out, most delusions are likely hybrids of both views especially when we consider them over periods of time. They propose a continuum from “received” to “reflective” delusion, whereby the former largely come to pass initially
via the endorsement process and the latter are formed mainly by an explanationist processes. Langdon and Bayne (2010) also importantly distinguish between “spontaneous” and “provoked” confabulation where the latter arises only in response to direct questioning. After all, what makes these patient dialogues so compelling and bizarre are their rather stunning confabulations in response to questions.

Thus, in some ways, I am rather sympathetic with Young’s two-stage “interactionist account” of the Cotard delusion (Young, 2012). He tells us that the interactionist explanation incorporates, for example, Ratcliffe’s (2008) emphasis on existential change and disturbed phenomenology, but at the same time acknowledges the necessary role played by some form of second-stage cognitive disruption [...] Furthermore, the interactionist approach holds that the formation of the delusional belief alters the nature of the phenomenal experience, such that what is perceived [...] is actually oneself dead or immortal (in the case of the Cotard delusion). (Young, 2012, p. 136)

Additionally, we might even suppose that there are unconscious inferences which we, by definition, wouldn’t be aware of making. It may even be that S2 is initially generated, at least in part, via unconscious reasoning designed to “make sense of” or rationalize the experience involved in S1. Surely delusional rationalizing need not always be a conscious process especially during spontaneous confabulations. Of course, once S2 is generated, then it may be used thereafter to affect or interpret first-order feelings or experiences. Notice also that the noninferential condition on HOT theory does not rule out unconscious inferences playing some role in generating an introspective state. Rather, it only rules out that S1 itself is or feels conscious in the same way when it arises via conscious inference.

In any case, regardless of the alternative explanations for the genesis of S2, Cotard syndrome as such does not seem to cause trouble specifically for HOT theory’s ability to explain S1 or S2. Just as HOT theory says about all conscious states, when a patient is aware of having a (first-order) conscious feeling or experience, there is a HOT about that state. This is what we have in the case of S1. But we should also recognize the delusional element involved in the formation of S2. There are of course situations where those with Cotard syndrome might even realize that they have come to think or believe contradictory things based on their abnormal experiences.

To be clear: I do not mean to suggest that the discussion above on endorsement and explanationist theories of delusion results either in some novel view of mine in relation to Cotard syndrome or that my view depends essentially on arguments contained in one of those accounts. The main point is rather that no matter how or when S2 is generated with respect to S1, HOT theory is not threatened and is indeed consistent with both S1 and S2. Still, HOT theory does have a natural way to account for S2 as more of an introspective state. Again, if there is an additional
and particularly unique issue posed by Cotard, it is with respect to cases of self-denial (”nihilistic delusions”), not with respect to HOT theory’s ability to account for S1 and S2.

With regard to any potential misrepresentation between a HOT and its own target, it is again important to note that some kind of minimal “I-as-subject” seems to remain even in the most extreme cases of Cotard syndrome. This minimal I-concept can merely be the extent to which a HOT theorist allows for a HOT that “I am in mental state M” with regard to S1. But there is never an I-as-object in the content of the first-order conscious state; rather, there is an implicit (and unconscious) I-as-subject at the second-order level as well as an I-as-object in a typical HOT. According to HOT theory, there would only be an I-as-subject concept in the first-order state. However, when one introspects and has a conscious HOT directed at a mental state, there is not only a conscious I-as-subject concept but also a conscious I-as-object concept in the content of the conscious HOT that may account for any further subjective sense of myness or ownership. Again, there would be no potentially problematic mismatch between an I-as-object in the content of a mental state M and its HOT because there isn’t an I-as-object concept at all in the content of M itself.

In terms of the general topic of this special issue, “can there be states of consciousness that lack any kind of self-consciousness?”, then, the answer partly depends upon how one defines “self-consciousness”, and there are clearly various notions of self-consciousness or self-awareness ranging from very minimal to very sophisticated. I have previously argued at length that conscious states and conscious creatures (including most animals) are self-conscious in at least one sense; namely, that each conscious state is accompanied by a HOT and so any conscious creature has HOTs (Gennaro, 1996). For various reasons I won’t repeat here, I also urged that a plausible minimal form of self-consciousness is simply having any kind of higher-order or meta-psychological thought. This can include any HOT, conscious or not, and even those containing such minimal I-concepts as “bodily self-awareness”; that is, distinguishing one’s own body from other things. So we are continuously “aware of” our mental states but this awareness is generally not a conscious awareness unless we are also attending to our mental states, in which case we have an unconscious third-order awareness of the second-order introspective awareness of the mental state. The answer to the above question would thus be “no” according to this way of understanding “self-consciousness” and according to HOT theory. However, my rather rudimentary notion of self-consciousness will likely not be robust enough to satisfy some philosophers. This is understandable and it is also partly why we must distinguish various levels of self-consciousness including the more sophisticated notions discussed in this paper, such as “introspection” as found in HOT theory. In this sense, of course, each conscious state does not involve self-consciousness according to HOT theory because we can have conscious states without introspecting them (because conscious states are most often outer-directed). So if one identifies self-consciousness with introspection, then
the answer to the above question would be “yes.” Still, Cotard patients do seem capable of some level of self-awareness and even having some introspective states, albeit often including some wildly bizarre beliefs or thoughts about themselves (as in S2). (Actually, it is sometimes also useful to distinguish further between two kinds of introspection: momentary focused and deliberate (Gennaro, 1996, p. 19). The former is a brief conscious focus on a mental state such as a pain or desire, whereas the latter involves consciously reasoning about our first-order mental states for a period of time.)

It is worth mentioning here that HOT theory, like many contemporary theories of consciousness in analytic philosophy of mind, tends to focus on explaining “state” consciousness, that is, “what makes a mental state conscious?” We sometimes speak of an individual mental state, such as a pain or perception, as conscious. This is in contrast to “creature” consciousness since we also often speak of organisms or creatures as conscious. Creature consciousness is simply meant to refer to the fact that an organism is awake or aware of its surroundings, as opposed to sleeping or in a coma. However, state consciousness is often implied by creature consciousness and vice versa. For example, a conscious creature will typically have conscious states and having conscious states are likely only to occur in generally conscious creatures.

One might wonder how the above discussion relates to what some authors characterize as “consciousness of oneself” or an “I-concept” specifically involving the capacity to form I*-sentences or I*-thoughts. The asterisk used here was introduced by Castañeda (1966) and can be found in the literature on de se propositional attitudes where an intentional state is directed at oneself (Perry, 1979). It is not just that I can have thoughts or beliefs about myself but that I myself (or I*) realize that the thought or belief is about me. I may come to realize that the messy shopper accidentally dropping food in the aisles is me or suddenly realize that the person in the store mirror is me. So this is clearly a more sophisticated notion of self-consciousness where one is thinking about oneself as an object. As Baker (2013) explains, “I* is typically embedded in the”that-clause” of a complex first-person sentence with a psychological or linguistic main verb” (p. 33), such as “I believe that I am short.” “I am not only the thinker of the thought but I am also part of the object of my thought” (pp. 32-33).

With regard to Cotard’s syndrome, on my view, I*-concepts enter the picture at the point I have called the “I-as-object” at the level of introspection. However, as we have seen, the delusional element in Cotard patients leads them to have radically false beliefs about themselves. Further, such beliefs and thoughts are sometimes directed at their bodies (or body parts) and sometimes directed at their mental states. As we have seen, this only complicates the matter (more on this in section 5). But, for example, in extreme forms of Cotard syndrome, as Metzinger (2003, p. 456) puts it, “we are faced with a delusional belief that can be expressed as follows: [I am certain that I* do not exist].” We might similarly suppose that there is the thought “I think that I* do not exist” which echoes the kind of intentional state reflected in S2.

©The author(s). https://philosophymindscience.org ISSN: 2699-0369
My main problem with Baker and others, however, is that she holds that the “capacity to use I*-language (i.e. a robust first-person perspective) is constitutive of self-consciousness” (Baker, 2013, p. 39) where “the robust first-person perspective is to conceive of oneself as oneself* in the first person” (Baker, 2013, p. 35). It is unclear to me why she doesn’t think that it’s preferable to suppose that there are degrees of I-concepts and self-consciousness ranging from more basic to rather sophisticated. Although Baker does allow for a “rudimentary first-person perspective,” she doesn’t consider anything less than having the more robust complete mastery of an I-concept as being a genuine self-concept at all. Why suppose that complete mastery is required for self-consciousness or for having an I-concept? And why should all self-concept possession be so closely tied to a linguistic ability? Surely there is room for degrees of concept possession not only for self-concepts but for most concepts (see Gennaro, 2012, chapters six through eight). This would also, in turn, open the door to the view that many animals and human infants are not only conscious but also possess concepts and self-consciousness to some degree.

It is worth mentioning in this context the way that Guillot explores our sense of self-awareness, that is, by making a threefold distinction within the more general notion of “subjective character” as follows:

1. “For-me-ness” is when the “subjective character is a subject’s characteristic awareness of her experience” (Guillot, 2017, p. 31); that is, the object of self-awareness is the experience itself.

2. “Me-ness” is when “subjective character is a subject’s awareness of herself as part of having the experience” (Guillot, 2017, pp. 31–32); that is, what makes an experience special for its subject is the fact that the subject is somehow aware of herself.

3. “Mineness” is when subjective character is a subject’s “awareness of herself as having the experience” (Guillot, 2017, p. 32); that is, subjective character presents it as a phenomenal awareness that my experiences are mine. Guillot holds that “the experience of prototypical Cotard patients exhibits for-me-ness, but neither me-ness nor mineness. She also suggests”that schizophrenic inserted thoughts might exhibit for-me-ness and me-ness, but not mineness” (2017, p. 44).

This analysis seems plausible to some extent. For example, Cotard patients do seem to be aware of a (rather unusual) experience (as reflected in S1) and thus retain some very minimal sense of self-awareness, but they do not seem to relate that experience to themselves in any coherent or significant way. As a matter of fact, they even often explicitly deny that there is a self to which that experience can be ascribed.
However, in relation to HOT theory, I have argued elsewhere for independent reasons that there isn’t ubiquitous self-awareness (i.e. for-me-ness) that accompanies each normal conscious first-order state. For example, I argue specifically against Kriegel’s self-representationalist view at length (Gennaro, 2008, 2012, chapter five), that is, against the notion that “for-me-ness” is a form of peripheral (i.e. non-attentive but conscious) inner awareness. In doing so, I address the differences between us with respect to reductionist motivations, as was briefly mentioned early in this paper. I also take issue with the phenomenological argument he adduces in favor of his view. It seems to me that our conscious attention is often so focused at the world and its objects that it seems unlikely that we are continuously consciously self-aware. It is also clear that such a view lacks plausibility especially as compared to the widely accepted and fairly obvious fact of peripheral outer-directed conscious awareness, such as the awareness in one’s peripheral visual field while watching a concert or working on one’s computer. Much the same goes for me-ness and mineness. Again, the HOT is typically unconscious when directed at a first-order conscious state. Me-ness and mineness, on my view, enter the picture only upon introspection, that is, when a HOT is consciously directed at one’s own mental state.

A similar line of response might also be raised against Ford and Smith’s (2006) argument in favor of the self-representationalist theory of consciousness. The claim here again is that there is an element of self-awareness – Guillot’s “for-me-ness” seemingly – internal to every conscious state. Ford and Smith contend that, for example, cases of depersonalization show that self-representationalism is correct. But just because the removal of something – for example, normal proprioception – causes deficits in one’s conscious mental states, it surely does not follow that the awareness of that thing is part of normal conscious experience. The relation could be causal instead of constitutive. That is, the typical abilities and awareness in question might merely, in the normal case, causally contribute to the phenomenology of one’s conscious mental states without being part of the conscious state itself (see also Howell & Thompson, 2017).

With regard to “bodily self-awareness,” Bermúdez (2011, pp. 161–166), for example, distinguishes between a “sense” and a “judgment” of ownership. He argues, in the end, that the so-called “sense” of ownership is best viewed as a “judgment” rather than as a feeling itself. In contrast to de Vignemont (2007), Bermúdez rejects the “inflationary” conception of the sense of ownership according to which there is a distinctive positive phenomenology of bodily ownership: “There are facts about the phenomenology of bodily awareness [...] and there are judgments of ownership, but there is no additional feeling of ownership” (Bermúdez, 2011, p. 166). There are still of course conscious bodily sensations and proprioceptive states. What one actually feels is the first-order conscious state accompanied by the unconscious thought about the mental state, as opposed to any phenomenology of so-called “myness”. Recall that HOT theory can explain why there is a phenomenological sense of myness when one introspects, namely, that the HOT is
itself conscious whereas no such sense is present when one has an unconscious HOT. The “I-concept” is part of a conscious thought in the introspective case but part of an unconscious thought in the outer-directed first-order case. Nonetheless, it is again certainly true that when there is a disturbance in one’s I-concept (including bodily self-awareness), one’s consciousness will likely be altered and result in some very odd beliefs and feelings of disownership. But, like Bermúdez, I do not find it compelling to argue that if a deficit of bodily awareness is manifested in consciousness, then that aspect of bodily awareness is normally part of our conscious states, albeit even in some peripheral way.

5 Self-awareness and personal identity: An ambiguity

With respect to Cotard patients specifically, Radovic (2017) hints at an interesting and related aspect of self-awareness which might well explain at least some of their odd statements:

In order to grasp in what sense Cotard patients believe themselves to be dead, it might be helpful to take a closer look at different conceptions of death and, in particular, the received notion of biological death. Biological death may be viewed from two different perspectives: a first-person point of view and a third-person point of view. From a naturalistic first-person point of view, death is commonly understood as the permanent absence of consciousness and the disintegration of personhood. Death is the end of our selves as conscious persons. By contrast, from a naturalistic third-person point of view, animals such as humans are dead when the heart does not beat, when there is no respiration, when the body is permanently still, or when the brain exhibits no neural activity. (Radovic, 2017, p. 688)

This line of analysis brings to mind a related distinction between two conceptions of “personal identity,” namely, psychological continuity and bodily continuity. The question raised is often framed as “What makes a person the same person over time?” Any such discussion in western philosophy tends to begin with the work of John Locke whose account of personal identity famously appealed to psychology and, more specifically, to consciousness and memory (Locke, 1689). On his view, a later person (P2) is identical to an earlier person (P1) just in case P2’s consciousness “can be extended backwards” to P1. This is taken to mean that P2 remembers P1’s thoughts and experiences, “from the inside” so to speak. My personhood goes with my first-person perspective and especially with my consciousness and memory, not necessarily with my body. The basic idea is familiar enough: I have memories from my first campus visit to the University of Southern Indiana, to experiences back in graduate school at Syracuse University, and so on. You do not
have those same memories and so my having them makes me the same person
as those previous persons (as opposed to you). They are episodic memories, that
is, I remember them as experienced from the first-person point of view. Episodic
memory is different than what is sometimes called “procedural memory,” that is,
memory for skills and abilities like playing the piano or riding a bike. Of course,
Locke is not saying that a fifty-year-old person must (episodically) remember ev-
er-ything from one’s earlier life since there is some normal forgetting of one’s own
past.

Although Locke’s view has been dominant in western philosophy, other
philosophers disagree and have put forth what has come to be known as “ani-
malism” which holds that being the same “biological organism” or having the
“same body” is essential to one’s personal identity (Olson, 2007). On this view,
one continues to exist “just in case one’s purely animal functions – metabolism,
the capacity to breathe and circulate one’s blood, and the like – continue” (Olson,
1997, p. 16). To elaborate, Olson explains that “an animal’s life is understood as
a self-organizing biological event that maintains the organism’s complex internal
structure. […] An organism must constantly take in new particles, reconfigure
and assimilate them into its living fabric, and expel those that are no longer useful
to it. An organism’s life enables it to persist and retain its characteristic structure
despite constant material turnover” (Olson, 2007, p. 28).

Interestingly, Locke himself recognized the ambiguity and thus distinguished
between “same person” and “same body” or “same biological organism.” For one
thing, Locke was clearly open to the possibility of something like what is now
called Dissociative Identity Disorder (DID) where “two or more persons” can be
associated with “one body.” Locke used the term “same man” to denote to latter.
Radovic (2017) correctly points out that Cotard patients often do seem to in-
dicate that they are dead in the biological sense as opposed to the psychological
sense. Consider:

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 depersonalized. (Simeon & Abugel, 2006, p. 133)

I felt that my brain was somewhere else and from there was just watch-
ing me […] I was completely unable to tell whether I was still present
or whether I was the part that was gone. In short, there were two
different beings, the one watching the other. (Roberts, 1960, p. 481)

Much of the above, therefore, highlights the ambiguity of self-reference talk and
the use of “I”. “I am dead” might be taken to mean that “my body is no longer alive
or functioning” or “I am no longer in my body” as opposed to “I am no longer
conscious” or “my conscious existence has ceased.” This could also partly explain
the apparent contradiction in some statements made by Cotard patients. That is,
if such a patient is saying “I am dead” and means it in the biological sense, then
there could still be a sense in which “I am alive” and detached from my body in

©The author(s). https://philosophymindscience.org ISSN: 2699-0369
Cotard syndrome, self-awareness, and I-concepts

some way, such as immortal or more like a “ghost.” Thus, it seems that something like “bodily self-awareness” still lingers in the background (perhaps more like somatoparaphrenia):

At one point he had stated that he was a ghost and that no one could see him. He explained that his physical body had been transformed into the immaterial body of a ghost. Nevertheless, he indicated that he visualized his body without any difficulty. About 2 months prior to admission he had concluded that eating was unnecessary since he was already dead. (Silva, Leong, Weinstock, & Gonzales, 2000, p. 188)

He often said, “I am a ghost” or “I am dead”, and sometimes denied the existence of certain parts of his body saying, for example, “I have no blood.” (Enoch & Ball, 2001, p. 165)

This suggests that Cotard patients often refer to themselves in third-person terms rather than in first-person terms.

Recall our two earlier statements:

S1: a conscious feeling or experience of unfamiliarity with myself (or body parts or mental states); and

S2: a conscious belief or thought that I am dead or I do not exist.

In light of this discussion, it may even be the case that S1 and S2 are not really in direct conflict at all. That is, Cotard patients may sometimes be referring to their bodies (or body parts) in S2 whereas S1 reflects a more first-person perspective on their own mental states. This is somewhat similar to Wittgenstein’s distinction between the “I-as-subject” and the “I-as-object” whereby the former may indicate a first-person point of view and the latter a third-person point of view.

Thus, with respect to HOT theory, the I-concept referent in a HOT might vary depending upon several factors, including which disorder or delusion we are addressing. It might refer to one’s own body as a sort of bodily self-awareness, perhaps more like somatoparaphrenia. It might refer to one’s own mental state, perhaps closer to thought insertion. But it might even refer to one’s own existence. It seems rather uncontroversial to say that are different self-concepts and some are more relevant than others under certain conditions. Billon makes a related point when he suggests that at least sometimes, Cotard patients “feel more entitled to refer to themselves in the third-person than in the first-person” (2016, p. 376).

Along somewhat similar lines, it is worth mentioning that some Cotard patient reports seem to resemble descriptions of out-of-body experiences (OBEs). An OBE occurs when one experiences one’s own body (or even the world) from above or outside one’s body. You feel like you left your body and are “floating” above it. This seems to indicate that your “mind” or “soul” is separating out from your body. A materialist may of course explain how very similar experiences can be induced

©The author(s). https://philosophymindscience.org ISSN: 2699-0369
by stimulating neurons in the temporo-parietal junction (Blanke & Arzy, 2005). Although Cotard patients can have a wide variety of neurological and psychiatric pathologies, at least some have been known to have abnormalities in the temporo-parietal area (Kudlur, George, & Jaimon, 2007).

6 Conclusion

In closing, then, we have seen that various psychopathologies of self-awareness might seem to threaten the viability of the higher-order thought (HOT) theory of consciousness since it requires a HOT about one’s own mental state to accompany every conscious state. The same might be said of Cotard syndrome but, by analogy with somatoparaphrenia and thought insertion, I have argued that HOT theory can answer this objection and can consistently account for what happens in this admittedly unusual case. There are possibly irrational combinations of mental states but HOT theory has the resources to explain why they do not cause trouble for the theory. There are also multiple “I-concepts” and levels of HOTs which can help to provide a more nuanced explanation. For example, there would not really be any “self-concept” mismatch between a conscious state and its HOT. In addition, Wittgenstein’s distinction between “I-as-subject” and “I-as-object” as well as different approaches to the problem of personal identity can aid us in understanding how apparently contradictory claims can be explained away. Finally, with regard to the connection between consciousness and self-consciousness, it seems clear from the above discussion that Cotard patients are capable of having some “I-thoughts” about their bodies and minds which seem to indicate various degrees of self-awareness. As we have seen, they do think about themselves in various ways. So I do not suppose that Cotard syndrome is characterized by a total lack of self-consciousness. However, perhaps not surprisingly given their delusional nature, many such thoughts about themselves are often rather bizarre and clearly false.

Acknowledgments

I would like to thank Tom Metzinger, Raphaël Millière, two anonymous reviewers, and the editors-in-chief of this journal for numerous comments on previous drafts of this paper. I would also like to thank Andrea Wright and Laura Bernhardt for their assistance in preparing the manuscript for publication in accordance with the specific submission format guidelines.

References


©The author(s). https://philosophymindscience.org ISSN: 2699-0369


**Open Access**

This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.