## Inserted Thoughts and the Higher-Order Thought Theory of Consciousness

## Rocco J. Gennaro

[in *Psychiatry and Neurosciences Update: Vol 4*, Pascual Angel Gargiulo & Humberto Luis Mesones-Arroyo (eds.), Springer, pp. 61-71, 2021.]

**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. 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” [1] [2]. In a previous publication, I argued that a HOT theorist can adequately respond to this concern with respect to somatoparaphrenia [3]. Somatoparaphrenia is a “depersonalization disorder” which is characterized by the sense of alienation from parts of one’s body. It is a bizarre body delusion where one denies ownership of a limb or an entire side of one's body. My focus in this chapter, however, is on “inserted thoughts” which is a common symptom of schizophrenia, although it will be useful to compare and contrast it with somatoparaphrenia. Schizophrenia is a mental disorder which often 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 somehow being inserted into one’s mind by someone else. Stephens and Graham, for example, have suggested that thought insertion should be understood as alienated self-consciousness or meta-representation [4]. I will argue that HOT theory also has nothing to fear from this phenomenon and can account for what happens in this admittedly unusual case.

**2. Somatoparaphrenia and HOT Theory**

Somatoparaphrenia is a pathology of self characterized by the sense of alienation from parts of one’s body. It is a very odd body delusion where one denies ownership of a limb or an entire side of one's body. It is thus sometimes called a “depersonalization disorder.” Relatedly, anosognosia is a condition in which a person who suffers from a disability seems unaware of the existence of the disability. A person whose limbs are paralyzed will insist that his limbs are moving and will become upset when caregivers say that they are not. Somatoparaphrenia is usually caused by extensive right-hemisphere lesions. Lesions in the temporoparietal junction are common in somatoparaphrenia but deep cortical regions (for example, the posterior insula) and subcortical regions (for example, the basal ganglia) are also sometimes implicated [5]. Anton’s syndrome is a form of anosognosia in which a person with partial or total blindness denies being visually impaired. The patient engages in rationalization in order to account for the inability to see. Patients with somatoparaphrenia utter some rather stunning statements, such as “parts of my body feel as if they didn’t belong to me” and “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” [6].

There has been some question as to whether or not the higher-order thought (HOT) theory of consciousness can account for the depersonalization psychopathology of somatoparaphrenia. Liang and Lane argue that it cannot [7]. The HOT theory of consciousness says that what makes a mental state a conscious mental state is that it is the target of a HOT to the effect that “I am in mental state M” (see figure 1). When the HOT is itself is unconscious, the conscious state is still outer-directed. When the HOT is conscious, we have *introspection*, and so the conscious thought is directed at the mental state [8].

[insert figure 1 here]

HOT theory has been critically examined in light of some psychopathologies because the theory says what makes a mental state conscious is a HOT of the form that “*I* am in mental state M.” The requirement of an I-reference leads some to think that HOT theory cannot explain some of these “depersonalization” pathologies. There would then seem to be cases where I can have a conscious state but attribute it to someone else. The “I” in the HOT is not only importantly *self-referential* but essential in tying the conscious state to *oneself* and thus to one’s *ownership* of M.

Rosenthal responds that one can be aware of bodily sensations in two ways that, normally at least, go together: (1) aware of a bodily sensation *as one’s own*, and (2) aware of a bodily sensation *as having some bodily location*, like a hand or foot [9]. Patients with somatoparaphrenia still experience the sensation as their own but also as having a mistaken bodily location. Such patients still do have the awareness in (1), which is the main issue at hand, but they also have the strange awareness in sense (2). So somatoparaphrenia leads some people to misidentify the bodily location of a sensation as some­one else’s even though the awareness of the sensation itself remains one’s own. But Lane and Liang are not satisfied and, among other things, counter that Rosenthal has still not explained why the identification of the *bearer* of the pain can­not also go astray [10].

I have replied that we can go further than Rosenthal in defending HOT theory [11]. First, we must first remember that many of these patients often deny feel­ing *anything* in the limb in question [12]. As even Liang and Lane point out, patient FB, while blindfolded, feels no tactile sensation when the examiner would in fact touch the dorsal surface of FB’s hand. In these cases, it is therefore difficult to see what the problem is for HOT theory at all. Second, when there really is a bodily sensation of some kind, a HOT theorist might also argue that there are really *two* conscious states that seem to be at odds. 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* often under the broader category of anosognosia [13] [14]. 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 [15] [16]. Beliefs, generally speaking, are themselves often taken to be intentional states integrated with other beliefs and mental states. They are typically understood as caused by perceptions or experiences which then lead to action or behavior. Thus, somatoparaphrenia is much closer to self-deception and involves frequent confabulation. If this is a reasonable interpretation of the data, then a HOT theorist can argue that the patient 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 conscious at any given time, is indeed strange and perhaps even self-contradictory in some sense, but the puzzlement has nothing to do with HOT theory.

A similar critique of self-representationalism [17] based on somatoparaphrenia might also be posed to what Billon and Kriegel call “subjectivity theories” which say that “there is something it is like for a subject to have mental state Monly if Mis characterized by a certain sense of “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” [18]. Patients with somatoparaphrenia seem to be cases where one has a conscious state without the “for-me-ness” aspect and thus not experienced as one’s own [19] [20]. However, Billon and Kriegel counter that “none of the patients that 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” [21]. Again, many disorders, including somatoparaphrenia, involve delusion or self-deception. A delusion is distinct from a belief based on incorrect or incomplete information, poor memory, illusion, or other effects of perception. Self-deception is a process of denying or rationalizing away the relevance, significance, or importance of opposing evidence and logical argument. Self-deception involves convincing oneself of a truth (or lack of truth) so that one does not reveal any self-knowledge of the deception. Delusions have certainly received extensive treatment from philosophers in recent years, sometimes in connection with self-deception [22] [23] [24].

Liang and Lane tell us that “what seems to be happening is that these tactile sensations are *represented as belonging to someone other than self*” [25] and that is problematic for HOT theory. But this is at best highly ambiguous because another way to represent sensations as belonging to someone else is via a propositional attitude such as a belief. And there would be no prob­lem for HOT theory, as such, as to whether or not these patients can have the conscious belief in S2. That is, a patient with somatoparaphrenia would still represent that *belief* and report it as her own. Still, Lane is not satisfied with these responses [26].

**3. Thought Insertion and Schizophrenia**

Another much discussed depersonalization disorder can be found in those with schizophrenia, which is a mental disorder characterized by disintegration of thought processes and emotional responsiveness. It most commonly manifests itself as auditory hallucinations, paranoid or bizarre delusions, or disorganized speech and thinking, and it is accompanied by significant social dysfunction. Thought insertion, a common symptom of schizophrenia, is the delusion that some thoughts are not one's own and are somehow being inserted into one's mind. In some particularly severe forms of schizophrenia, the patient seems to lose the ability to have an integrated or “unified” experience of her world and self. The person often speaks in an incoherent fashion and is unable to act on simple plans of action.

Stephens and Graham suggest that thought insertion should be understood as alienated self-consciousness or meta-representation[27]. They think that schizophrenics make introspective inferential mistakes about the source of inserted thoughts based on delusional background beliefs. Some bodily movements can of course be movements of *my* limbs without counting as *actions of mine* or as caused by me. Perhaps someone else is controlling my movements or perhaps they are entirely involuntary such as the physical tics and vocalizations in Tourette’s syndrome. Another relevant disorder is anarchic hand syndrome which is a neurological disorder in which individuals frequently perform seemingly voluntary movements that they do not consciously intend and cannot directly inhibit. But in these cases the bodily movements are still self-attributed to the person with the disorder, so something else must be going on to explain attributions to others in thought insertion. If a song spontaneously runs through my mind, I still think of it as an episode in my mind. But it may not count as *my* mental activity in the same way as when I am thinking through a math problem or trying to plan a trip. The latter, but not the former, involves intentional thought that expresses my agency. There seems to be something special going on when I consciously engage in some activity which involves mental effort and voluntariness. Stephens and Graham call the feeling of *having* a mental state the “sense of subjectivity” and the feeling of *causing* my mental state the “sense of agency.” They urge that these two can come apart in unusual cases so that thought insertion involves the sense of subjectivity without the sense of agency, which accounts for the curious “passivity experience” of schizophrenics. So attributing thoughts to someone else under these circumstances does make some sense since the mental state must be caused by something or someone (see also [28] [29] [30] for much more on the above line of thought).

Gallagher makes a similar distinction between a “sense of ownership” and a “sense of agency” [31] but, in contrast to Stephens and Graham’s “top-down” approach, argues instead that the primary deficit regarding thought insertion is more of a “bottom-up” problem with the first-person experience itself rather than a self-monitoring abnormality. What happens at the introspective level is not erroneous but rather a correct report of what the schizophrenic actually experiences, that is, thoughts that feel different and externally caused. Gallagher also points to some preliminary neurological evidence which indicates abnormalities in the right inferior parietal cortex for delusions of control and ownership.

It is worth mentioning again that in the background is the related issue of whether or not there is “pre-reflective,” or implicit, self-consciousness in all conscious states, which is often associated with a so-called “sense of ownership” or “sense of mineness” [32] [33] [34]. According to self-representationalism, for example, a fundamental aspect of all conscious experiences is that they seem to be *mine*. In being aware of any thought, action, perceptual experience, memory, or bodily experience, I am aware of it as being *my own*. Perhaps this is one reason why it seems difficult to understand how one could attribute a pain as located in another’s body or a thought caused by someone else or even located in another’s mind. Perhaps a sense of ownership explains why some mental states seem to be “immune to error through misidentification,” that is, impossible to be mistaken about with regard to mental state ownership (more on this below). Perhaps lacking this sense of ownership in abnormal cases can explain the odd and delusional thinking in such rare cases.

But even within the context of an implicit account of self-consciousness in experience (or a “sense of ownership”), there is significant disagreement between robust and deflationary accounts. For example, Zahavi and Kriegel defend a more robust understanding of the sense of ownership as a distinct aspect of the phenomenal character of all conscious states and as a necessary feature of all experience. By contrast, Bermúdez argues in favor of a deflationary account of the sense of ownership over one’s own body, according to which it consists in nothing more than the phenomenology of the spatial location of bodily sensations together with our tendency to judge the body in which they occur to be our own [35]. So perhaps there is no “sense” of ownership which accompanies all conscious states.

I have used something like this line of response against Ford and Smith’s argument in favor of the so-called “self-representationalist” theory of consciousness, whereby first-order conscious states are always accompa­nied by an inner-directed peripheral conscious awareness [36]. Ford and Smith contend that cases of depersonalization show that some­thing like Kriegel’s view is correct. But again, 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. There are many ways that normal consciousness can be disturbed or impaired (e.g. being unable to breathe) but surely we shouldn’t conclude that every such disturbance shows that the ability in question normally shows up regularly in our phenomenology [37]. (See [38] for another deflationary account from an Eastern philosophical viewpoint.)

Notice that this fits nicely with HOT the­ory, which 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 concept “I” is part of a *conscious* thought in the introspective case but 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 *dis*ownership (or thought disownership for that matter). But, like Bermudez, 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 always or even normally part of our consciousness.

**4. Thought Insertion and HOT Theory**

The puzzle of thought insertion is thus perhaps somewhat analogous to what we have seen with respect 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 and/or a sensation to someone else whereas some schizophrenics attribute thoughts (or at least their causal origin) to someone else. If this analogy is correct, then a HOT theorist might argue that the schizophrenic patient has (at least) the following *two* conscious states:

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

S2: a conscious *belief* that the thought has been inserted by someone else.

Having both S1 and S2, especially if both are conscious at any given time, is indeed strange and perhaps even irrational in some sense, but the puzzlement again has nothing to do with HOT theory. In S1, we can suppose that the patient has a HOT about her own thought, i.e. she is aware of the mental state as 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 it is that S2 is generated is of course interesting in its own right. For example, as Stephens and Graham believe, perhaps the best explanation is indeed that it is 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 these 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. Once again, we must also recognize the delusional element involved in S2. There are of course situations where those with thought insertion might even realize that they have come to believe contradictory things based on their abnormal experiences [39].

It would again be highly ambiguous if Liang and Lane were to say that inserted thoughts are “represented as belonging to someone else” as if that is problematic for HOT theory. This is partly because another way to represent such thoughts as belonging to someone else is via a propositional attitude such as a belief. And there would be no prob­lem for HOT theory, as such, as to whether or not these patients have the conscious belief in S2. That is, a schizophrenic patient with thought insertion would still represent *that belief* as her own while still having S1.

In some ways, then, I am somewhat sympathetic with Parrott’s account of thought insertion [40]. Parrott first notes that prevailing views in philosophy and cognitive science tend to characterize the experience of thought insertion as missing or lacking some element, such as a “sense of agency,” found in ordinary first-person awareness of one's own thoughts. However, in another sense, it might be that rather than lacking something, experiences of thought insertion have an *additional feature* not present in ordinary conscious experiences of one’s own thoughts. So thought insertion consists of two distinct elements: a state of ordinary first-person awareness of a thought and the thought that this state of awareness is highly unusual. The latter leads via delusional thinking and inferential reasoning to the belief noted above in S2. So it is really the additional S2 which creates most of the puzzle.

Smith makes a further distinction and explains that “we might…wish to make a three-way distinction between the sense of *agency* (the sense that one is the *author* of a mental state), the sense of *ownership* (the sense that one is the *owner* of a mental state), and what we might call the sense of *location* (the sense that a mental state is located *within* one’s own mind). The sense of location might be understood as being possessed if one is aware of a mental state in the ordinary way, i.e., introspectively. Crucial, it would seem, for evaluating the significance of thought insertion and related cases…will be determining which, if any, of the senses of agency, ownership, or location remain intact” [41]. Smith notes that it might be argued that what such patients retain is in fact the sense of location, rather than the sense of ownership. That is, it may be possible to take their descriptions at face value when they deny, in thought insertion for example, that the thoughts are their own (or were thought by them) while still accepting that the inserted thought occurs within the boundary of their own mind. The explicit mention of location is reminiscent of patients with somatoparaphrenia who arguably still experience the sensation as their own but also as having a mistaken bodily location, as Rosenthal had initially urged.

One difficulty of course is how literally to take reports of patients with schizophrenic inserted thoughts. Should we take them at face value or at least treat them with some skepticism given the delusional aspect in question? Consider, for example, some statements of those suffering from thought insertion. Frith quotes one patient as saying that “thoughts are put into my mind like “kill God”…They come from this chap, Chris” but then the patient says that “they are his [Chris’] thoughts” [42] which is a further stronger claim. So there is also some ambiguity in patient first-person reports. In a standard textbook on the topic, we are told that “in thought alienation [i.e. thought insertion] the patient has the experience that…others are participating in his thinking” [43]. It is difficult to know when to take such patients literally but it does also seem that they sometimes report that the inserted thoughts themselves are someone else’s.

 Thus, let us go further and consider the following three statements in this context:

(1) I am having (experiencing) a conscious thought T but it was inserted into me by someone else.

In (1), we might say that the location is in my mind, the ownership is mine, but the origin is elsewhere.

(2) I am having (experiencing) a conscious thought T but it is in another’s mind.

For (2), we might say that it is my ownership but not my location or origin.

(3) I am having (experiencing), or “aware of” in some sense, someone else’s conscious thought T.

In (3), it seems that T is actually someone else’s thought and so that even my ownership appears to be in doubt. The thought is “in my mind” in some sense (location) but I also think of it as someone else’s, perhaps analogous to having someone else’s furniture in my house. The notion of “ownership” might also sometimes be confused with location because there is a sense in which location would normally imply ownership. Is it “my chair” if I am renting a furnished house? Well, yes and no. Is it “my chair” if I have someone else’s chair temporarily in my house? Well, yes and no. Grammatically, there can also be some confusion or ambiguity. When I say that I am home sitting in “my chair” it would perhaps be more accurate to say that I am sitting in a chair that is in my house.

Still, with regard to (1)-(3), it would seem that there is still at least a grammatical reference to oneself, and thus at least some sense of ownership, in each case. For one thing, each starts with “I am having (experience)….” This is especially clear with respect to (1) and (2) and clearly reflects the typical content of a HOT, that is, something like “I am in M” or “I am aware of M” or “I am experiencing M,” as we have seen. Even in (2), the additional and puzzling belief that T is in another’s mind can be the result of delusional reasoning as described above with respect to S2 earlier. Still, there is nothing in (2) itself which automatically rules out my *ownership*. At most, perhaps the location and perceived origin is different or unusual. It may of course be false or delusional to think so, but that is a different matter and does not threaten HOT theory.

In addition, the more radical notion that “I am having a thought that is in your mind” is perhaps not impossible in some hypothetical or empirical instances. For example, if telepathy is real (or just possible), perhaps person A can cause person B to have certain thoughts or it is possible for person B to gain access to some of person A’s thoughts. Hirstein goes further and describes in some detail how what he calls “mindmelding” might be empirically possible [44]. He argues that it is indeed possible for one person to directly experience the conscious states of another. This would involve making just the right connections between two peoples’ brains. He then considers the consequences of the possibility that what appeared to be a wall of privacy can actually be breached. In line with the analogy above, I might also say that I am sitting in “my chair” but in “your house” in a case where I have lent you my chair temporarily.

Relatedly, Langland-Hassan discusses an actual case of craniopagus twins -- twins conjoined at the head and brain – where each twin seems to know what the other is seeing or feeling, and perhaps even thinking, in a way others cannot [45]. Interestingly, each twin (Krista and Tatiana Hogan) would seem to know these things through introspection. Perhaps the twins can *introspect* the same shared mental state, that is, have a kind of mental state *co-ownership*. If this is granted as a possibility, then mental state ownership need not be understand as an either-or matter. Perhaps the typical underlying assumption that each mental state only has one owner is false in some admittedly rare and bizarre circumstances. These considerations would also help to make sense of (3) above. Perhaps I can experience someone else’s thought T but T is *also* my thought because we can co-own T (and so T is still in my mind, in some sense). Alternatively, Langland-Hassan considers the more radical possibility that one person may be said to be introspectively aware of only another’s mental state.

There are also numerous other very strange and related delusions discussed in the literature, such as Cotard syndrome where people hold a delusional belief that they are dead (either figuratively or literally), do not exist, are putrefying, or have lost their blood or internal organs [46]. But notice that even here the statement that “I think I don’t exist” or “I think I’m dead” still at least contains an initial reference to oneself (as a grammatical fact). It seems to me that there must 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). But this is precisely what leads to a contradiction. Imagine: “Who thinks that you don’t exist?” Answer: “I do”? “Who is dead?” Answer: “I am”? Again, perhaps this is a misleading way of speaking but the same might be true of many patient reports.

In all the above cases, there is a conscious thought which can be explained by a HOT theorist in the usual way, but there is *another* belief or thought which contradicts or the first one or is delusional in some other way. It is also very important to recognize that HOT theory comes with a well-known noninferentiality condition, such that a HOT must become aware of its target mental state noninferentially, that is, in an unmediated way. As Rosenthal repeatedly emphasizes, the point of this condition is mainly to rule out certain 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 my 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 noninfer­entiality condition. Thus, HOT theory requires that the HOT arises in an unmediated manner. This is also important in this context because if there is any kind of inference to a belief, such as in S2, then the HOT would not arise in the requisite manner. Also, and perhaps more importantly, using 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. This, in turn, increases the likelyhood that delusional thinking will generate further false beliefs. Another way to think of this is that there is “introspective access” to a mental state M which involves a kind of “ownership” as well. Normally, at least, M is assumed to be mine in the sense that it is the object of my introspection. In cases of thought insertion, however, the patient sometimes seems to be calling that assumption into doubt.

**5. Immunity to Error through Misidentification (IEM)**

I mentioned above the much-discussed “immunity to error through misidentification” (IEM) principle [47]. According to Shoemaker, a cer­tain subset of thoughts about oneself is immune to error through misiden­tification [48]. As Shoemaker makes clear, one can think about oneself under any number of descriptions. But only some I-thoughts are immune to error through misidentification -- namely, those I-thoughts that are directed at one’s mind and mental life, as opposed to one’s body and corporeal life. Wittgenstein observed that I can see in the mirror a tangle of arms and mistakenly take the nicest one to be mine [49]. I may think to myself, “I have a nice arm.” In that case, I may not only be wrong about whether my arm is nice, but also about who it is that has a nice arm. Such an I-thought about my body (or body part) is not immune to error through misidentification. In extreme abnormal cases, such as in mirror self-misidentification, one might even believe that *one’s own* reflection in a mirror is some other person. However, one cannot be mistaken that it is I who is having the experience in question. I cannot be mistaken that I am perceiving the reflection in the mirror. In a sense, then, I can be wrong about what my mental state is directed at (the object) but not about who is having the mental state in the first place (the subject).

Let us return to the above statements (1)-(3):

(1) I am having (experiencing) a conscious thought T but it was inserted into me by someone else.

(2) I am having (experiencing) a conscious thought T but it is in another’s mind.

(3) I am having (experiencing), or “aware of” in some sense, someone else’s conscious thought T.

Recall that HOT theory has been critically examined in light of some psychopathologies because, according to HOT theory, what makes a mental state conscious is a HOT of the form that “I am in mental state M.” The requirement of an I-reference leads some to think that HOT theory cannot explain these “depersonalization” pathologies. There would seem to be cases where I can have a conscious state but attribute it to someone else or possibly even introspect another’s mental state. Thought insertion seems to threaten HOT theory because it contradicts the notion that the accompanying HOT that “I am in mental state M.” Recall that Lane and Liang protest that Rosenthal has still not explained why the identification of the bearer of the pain cannot also go astray in somatoparaphenia, especially since Rosenthal clearly holds that misrepresentation can occur between a HOT and its target. But whatever one thinks of standard cases of misrep­resentation between the first-order and second-order level on HOT theory, it is not clearly relevant here because those abnormal cases involve differences in the *contents* of the two respective states. Although Lane and Liang claim that there should be equally the possibility of a mismatch between the “I” in the HOT and the “I” in the first-order mental state, it is unclear how this could be so.

Notice again that (1), (2), and (3) still all start with what Wittgenstein called the “I-as-subject” and thus there seems to be some kind of minimal sense of “ownership” which remains even in the most extreme cases of thought insertion. This minimal sense can be the extent to which a HOT theorist allows for a HOT that “I am in mental state M” with regard to S1 even if the patient also has another delusional conscious state which runs counter to S2. Wittgenstein use­fully distinguished between the “I-as-subject” (e.g., “I have a pain”) and the “I-as-object” (“I have a broken arm”). Crucially, however, there is never an I-as-object in the *content* of the *first-order* conscious state, but 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 a first-order state whereas the content of the state refers to the outer world. This would be a kind of “raw bearer” of the state, as Rosenthal calls it [50]. After all, if we assume that any mental state must have a bearer, then even first-order states should involve some primitive concept of I. The same is true for the unconscious HOT that accompanies a first-order conscious state, but here there is *also* an I-as-object referenced in the *content* of the HOT (i.e., “I think that *I am in M*”). Still, these I-concepts are normally parts of unconscious thoughts, and so there is little reason to suppose that there is any phenomenological sense of “myness” in these cases. 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 HOT that may account for any subjective sense of myness. Further, there would be no 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. (Caruso also presents an interesting discussion of thought insertion and IEM in the context of HOT theory. I do agree with him that the overall sense of mental state “ownership,” to the extent that there is one, comes more from our ability to *introspect* our mental states [51]).

Seeger also discusses IEM and several psychopatholgies including somatoparaphrenia and thought insertion [52]. He is very careful to clarify and define some of the claims at issue. For example, he describes the “Self-Awareness Thesis” as the claim that “if one is introspectively aware of a mental state, then one is necessarily aware of that state as one’s own state” but then convincingly argues that the self-awareness thesis is not quite the same as what he calls “The Immunity Thesis” which says that “if one self-ascribes a state of which one is introspectively aware, then that state is one’s own state.” So it may be that thought insertion is a counterexample to the self-awareness thesis but not to the immunity thesis since, as we saw in the previous section, the schizophrenic patient may indeed assert that an introspected thought is someone else’s. Seeger explains that the pathologies show that it is possible to be introspectively aware of a mental state without self-ascribing it. However, when a patient does self-ascribe an introspected thought, the thought must be his own. Self-awareness and self-ascription can come apart in unusual circumstances. In any case, my main concern here is that thought insertion is not a threat to HOT theory as opposed to arguing for a particular preferred definition of IEM.

In any case, I conclude overall that HOT theory can withstand the alleged threat from the phenomenon of thought insertion in a way somewhat analogous to somatoparaphrenia. Indeed, I think that HOT theory can even help to explain what happens in these cases, especially when one is clear about the nature of delusions and is careful about the concepts in question, such as the complex nature of self-awareness and the more subtle details of HOT theories. It is important to recognize the difference between the content of first-order mental sates and the content of HOTs (including introspective states).

**References**:

[1] Rosenthal D. Consciousness and mind. New York: Oxford University Press; 2005.

[2] Gennaro R. The consciousness paradox: Consciousness, concepts, and higher-order thoughts. Cambridge, MA: The MIT Press; 2012.

[3] Gennaro R. Somatoparaphrenia, anosognosia, and higher-order thoughts. In: Gennaro R. editor. Disturbed consciousness: New essays on psychopathology and theories of consciousness. Cambridge, MA: The MIT Press; 2015.

[4] Stephens G, Graham, G. When self-consciousness breaks: alien voices and inserted thoughts. Cambridge, MA: The MIT Press; 2000.

[5] Vallar G, Ronchi R. Somatoparaphrenia: a body delusion: A review of the neuropsychological literature. Experimental Brain Research 2009; 192: 533-551.

[6] Sierra M, Berrios G. The Cambridge depersonalisation scale: a new instrument for the measurement of depersonalisation. Psychiatry Research 2000; 93: 153-164, p. 160, 163.

[7] Liang L, Lane T. Higher-Order thought and pathological self: The case of somatoparaphrenia. Analysis 2009; 69: 661-668.

[8] Rosenthal D. A theory of consciousness. In: Block N, Flanagan O, Güzeldere, G editors. The nature of consciousness. Cambridge, MA: MIT Press; 1997.

[9] Rosenthal D. Consciousness, the self and bodily location. Analysis 2010; 70: 270-276.

[10] Lane T, Liang C. Mental ownership and higher-order thought. Analysis 2010; 70: 496-501.

[11] Gennaro R. Somatoparaphrenia, anosognosia, and higher-order thoughts. In: Gennaro R. editor. Disturbed consciousness: New essays on psychopathology and theories of consciousness. Cambridge, MA: The MIT Press; 2015, p. 57-58.

[12] Bottini G, Bisiach E, Sterzi R, Vallar G. Feeling touches in someone else’s hand. NeuroReport 2002; 13: 249-252.

[13] Breen N, Caine D, Coltheart M, Hendy J, Roberts C. Towards an understanding of delusions of misidentification: four case studies. Mind and Language 2000; 15: 74-110.

[14] Feinberg T. Neuropathologies of the self: clinical and anatomical features. Consciousness and Cognition 2011; 20: 75-81.

[15] Bortolotti L. Delusions and other irrational beliefs. New York: Oxford University Press; 2009.

[16] Radden J. On delusion. Abingdon and New York: Routledge; 2010.

[17] Kriegel U. Subjective consciousness. New York: Oxford University Press; 2009.

[18] Billon A, Kriegel U. Jaspers’ dilemma: the psychopathological challenge to subjectivity theories of consciousness. In: Gennaro R. editor. Disturbed consciousness: New essays on psychopathology and theories of consciousness. Cambridge, MA: The MIT Press; 2015, p. 29.

[19] Liang L, Lane T. Higher-Order thought and pathological self: The case of somatoparaphrenia. Analysis 2009; 69: 661-668.

[20] Lane T. Self, belonging, and conscious experience: A critique of subjectivity theories of consciousness. In: Gennaro R. editor. Disturbed consciousness: New essays on psychopathology and theories of consciousness. Cambridge, MA: The MIT Press; 2015.

[21] Billon A, Kriegel U. Jaspers’ dilemma: the psychopathological challenge to subjectivity theories of consciousness. In: Gennaro R. editor. Disturbed consciousness: New essays on psychopathology and theories of consciousness. Cambridge, MA: The MIT Press; 2015, p. 37.

[22] Hirstein W. Brain fiction: Self-deception and the riddle of confabulation. Cambridge, MA: MIT Press; 2005.

[23] Bayne T, Fernandez J. eds. 2009. Delusion and self-deception. Psychology Press; 2009.

[24] Pliushch I, Metzinger T. Self-deception and the dolphin model of cognition. In: Gennaro R. editor. Disturbed consciousness: New essays on psychopathology and theories of consciousness. Cambridge, MA: The MIT Press; 2015.

[25] Liang L, Lane T. Higher-Order thought and pathological self: The case of somatoparaphrenia. Analysis 2009; 69: 661-668, p. 664.

[26] Lane T. Self, belonging, and conscious experience: A critique of subjectivity theories of consciousness. In: Gennaro R. editor. Disturbed consciousness: New essays on psychopathology and theories of consciousness. Cambridge, MA: The MIT Press; 2015.

[27] Stephens G, Graham, G. When self-consciousness breaks: alien voices and inserted thoughts. Cambridge, MA: The MIT Press; 2000.

[28] Frith C. The cognitive neuropsychology of schizophrenia. East Sussex, UK: Psychology Press; 1992.

[29] Bortolotti L, Broome M. A role for ownership and authorship in the analysis of thought insertion. Phenomenology and the Cognitive Sciences 2009; 8: 205-224.

[30] Parnas J, Sass L. Bodily awareness and self-consciousness. In: Gallagher S, editor. The Oxford Hand­book of the Self. New York: Oxford University Press; 2011.

[31] Gallagher S. Agency, ownership, and alien control in schizophrenia. In Zahavi D, Grunbaum T, Parnas J. editors. The structure and development of self-consciousness: Interdisciplinary perspectives. Amsterdam and Philadelphia: John Benjamins; 2004.

[32] de Vignemont F. The mark of bodily ownership. Analysis 2013; 73: 643-651.

[33] Billon A. Does consciousness entail subjectivity? The puzzle of thought insertion. Philosophical Psychology 2013; 26: 291-314.

[34] Zahavi D, Kriegel U. For-Me-Ness: what it is and what it is not. In: Dahlstrom D, Elpidorou A, Hopp, W editors. Philosophy of mind and phenomenology. Rowman and Littlefield: Routledge Press; 2015.

[35] Bermudez J. Bodily awareness and self-consciousness. In: Gallagher S, editor. Oxford handbook of the self. New York: Oxford University Press; 2011, pp. 161-166.

[36] Ford J, Smith DW. Consciousness, self, and attention. In: Kriegel U, Williford, K. editors. Self-represen­tational approaches to consciousness. Cambridge, MA: MIT Press; 2006.

[37] Gennaro R. The consciousness paradox: Consciousness, concepts, and higher-order thoughts. Cambridge, MA: The MIT Press; 2012, pp. 127-129.

[38] Chadha M. No-Self and the phenomenology of ownership. Australasian Journal of Philosophy 2018; 96: 14-27.

[39] de Haan S, de Bruin L. Reconstructing the minimal self, or how to make dense of agency and ownership. Phenomenology and the Cognitive Sciences 2010; 9: 373-396, p. 385.

[40] Parrott M. Subjective misidentification and thought insertion. Mind and Language 2017; 32: 39-64.

[41] Smith, Joel, "Self-Consciousness", *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2017/entries/self-consciousness/>, p. 12.

[42] Frith C. The cognitive neuropsychology of schizophrenia. East Sussex, UK: Psychology Press; 1992, p. 66.

[43] Fish F, Hamilton M. Fish’s clinical psychopathology. Bristol, UK: John Wright; 1985, p. 49.

[44] Hirstein W. Mindmelding: consciousness, neuroscience, and the mind’s privacy. New York: Oxford University Press; 2012.

[45] Langland-Hassan P. Introspective misidentification. Philosophical Studies 2015; 172: 1737 -1758.

[46] Billon A. Making sense of the Cotard Syndrome: Insights from the study of depersonalisation. Mind and Language 2016; 31: 356-391.

[47] Prosser S, Recanati F. editors. Immunity to error through misidentification: new essays. Cambridge, MA: Cambridge University Press; 2012.

[48] Shoemaker S. Self-reference and self-awareness. Journal of Philosophy1968;65: 555-567.

[49] Wittgenstein L. The blue and brown books, 2nd ed. Oxford: Blackwell; 1958/1969.

[50] Rosenthal D. Awareness and identification of self. In: Liu J, Perry J. editors. Consciousness and the self: new essays. Cambridge, MA: Cambridge University Press; 2012.

[51] Caruso G. Free Will and Consciousness. Rowman and Littlefield: Lexington Press; 2012, ch. 7.

[52] Seeger M. Immunity and self-awareness. Philosophers’ Imprint 2015; 15 (17): 1-19, p. 10, 13-14.