

Spence, S., D. Brooks, S. Hirsch, P. Liddle, J. Meehan, and P. Grasby. 1997. A PET study of voluntary movement in schizophrenic patients experiencing passivity phenomena (delusions of alien control). *Brain* 120:1997–2011.

Stephens, G. L., and G. Graham. 2000. *When Self-Consciousness Creaks: Alien Voices and Inserted Thoughts*. Cambridge, MA: MIT Press.

Synofzik, M., G. Vosgerau, and A. Newen. 2008. Beyond the comparator model: A multifactorial two-step account of agency. *Consciousness and Cognition* 17:219–239.

Wegner, D. 2004. Precis of *The Illusion of Conscious Will*. *Behavioral and Brain Sciences* 27:649–659, discussion 659–692.

Wolpert, D., Z. Ghahramani, and M. Jordan. 1995. An internal model for sensorimotor integration. *Science* 269:1880–1882.

Wolpert, D., and R. Miall. 1996. Forward models for physiological motor control. *Neural Networks* 9:1265–1279.

5 Self, Belonging, and Conscious Experience: A Critique of Subjectivity Theories of Consciousness

Timothy Lane

Nothing alien happens to us, but only what has long been our own.

—Rilke (1954, 50)

1 Introduction

Subjectivity theories of consciousness take self-reference, somehow construed, as essential to having conscious experience. These theories differ with respect to how many levels they posit and to whether self-reference is conscious or not. But all treat self-referencing as a process that transpires at the personal level, rather than at the subpersonal level, the level of mechanism.

Working with conceptual resources afforded by preexisting theories of consciousness that take self-reference to be essential, several attempts have been made to explain seemingly anomalous cases, especially instances of alien experience. These experiences are distinctive precisely because self-referencing is explicitly denied by the only person able to report them: those who experience them deny that certain actions, mental states, or body parts belong to self. The relevant actions, mental states, or body parts are sometimes attributed to someone or something other than self, and sometimes they are just described as not belonging to self. But all are *referred away from self*.

The cases under discussion here include somatoparaphrenia, schizophrenia, depersonalization, anarchic hand syndrome, and utilization behavior; the theories employed are higher-order thought, wide intrinsicity, and self-representational. I argue that each of these attempts at explaining or explaining away the anomalies fails. Along the way, since each of these theories seeks at least compatibility with science, I sketch experimental approaches that could be used to adduce support for my position or, indeed, for the positions of theorists with whom I disagree.

In a concluding section, I first identify two presuppositions shared by all of the theorists considered here and argue that both are either erroneous or misleading. Second, I call attention to divergent paths adopted when attempting to explain alienation experiences: some theorists choose to add a mental ingredient, and others prefer to subtract one. I argue that alienation from experience, action, or body parts could result from either addition or subtraction, and that the two can be incorporated within a comprehensive explanatory framework. Finally, I suggest that this comprehensive framework would require self-referencing of a sort, but self-referencing that occurs solely on the level of mechanism, or the subpersonal level. In adumbrating some features of this “subpersonal self,” I suggest that there might be one respect in which it is prior to conscious experience.

2 Subjectivity Theories of Consciousness, Higher-Order Thought, and Belonging

Gennaro (this volume) raises a host of concerns about my interpretation of experimental data concerning some tactile experiences reported by a particular somatoparaphrenia patient.¹ Briefly, the patient (FB) reported that her left hand belonged to her niece, and that she (FB) could not feel tactile sensations in that, “her niece’s” hand. In a series of controlled experiments, however, FB did report recovery of tactile sensation when the left hand was touched after FB had been told that the experimenter was about to touch her niece’s hand, if in fact the left hand was touched. On catch trials, when the hand was not in fact touched, irrespective of whether the experimenter indicated intent to touch FB’s left hand or “her niece’s” hand, FB reported feeling nothing.

My description and interpretation of that case is part of a general critique and rejection of what Billon and Kriegel (this volume) refer to as “subjectivity theories of consciousness,” or SP theories, because all are committed to a *subjectivity principle*. According to this principle, mental states can exhibit phenomenal consciousness only if they involve self-reference. Billon and Kriegel advocate a strong version of this principle, holding that it is metaphysically impossible for any phenomenally conscious state to be instantiated in the absence of subjectivity, the experience of “for-me-ness.” Gennaro (2012, 299–300) advocates a weaker version, holding that it is only necessary for there to be an unconscious higher-order thought with a self-referential component.² But all advocates of an SP theory take self-reference to be a *sine qua non*.

The SP theory that concerns Gennaro is Rosenthal’s (2005) higher-order thought theory of consciousness (HOT).³ Indeed, Mylopoulos (this volume) is also principally concerned with HOT. According to HOT theory, first-order mental states are conscious only if they are targeted by HOTs with the content, “I am in a certain state.” The reference to “I,” understood as the owner of the state, is “unavoidable” (Rosenthal 2005, 342, 347). This necessity claim implies that “being conscious of a state as belonging to someone other than oneself would plainly not make it a conscious state” (Rosenthal 2005, 342). It is in this context that I invoked the case of FB, because it appears that she is only conscious of certain tactile states when those states are *experienced as belonging to someone other than herself*, namely, her niece.⁴

Gennaro argues that somatoparaphrenia does not constitute a challenge to HOT; Billon and Kriegel argue that it poses no challenge for SP theories in general. In crafting their arguments, they do refer to one fact that I agree to be noncontentious: thus far, published studies of somatoparaphrenia include no patient denials that reportable sensations belong to self. Emphasis is invariably placed on denial of limb ownership (e.g., Romano et al. 2014, 1216, table 2). But the point that is relevant to HOT or SP theories in general does not concern somatoparaphrenia per se; instead, it is specific to the experimental data reported by Bottini et al. (2002). In the next paragraph, I direct attention to what I take to be the critical issue, the issue that constitutes a challenge to SP theories of any stripe.⁵

Recall that, in FB’s case, somatoparaphrenia was accompanied by tactile extinction, but she recovered the ability “to perceive tactile stimuli, *provided that these were referred to someone else’s body*” (Bottini et al. 2002, 251).⁶ During the experiment that demonstrated FB’s ability to recover from tactile extinction, she was blindfolded. While blindfolded, she was advised that the examiner would touch *her* left hand. Whenever this was done, FB reported that she felt no tactile sensations. When advised that the examiner was about to touch *her niece’s* hand, however, upon being touched, FB reported feeling tactile sensations. In order to ensure experiment reliability, along with the blindfold and other controls, catch trials were included, trials for which FB was led to expect touches that were not forthcoming. These trials were evenly distributed across three verbal warnings—I am going to touch your right hand, your left hand, and your niece’s hand—and were administered in four sessions, two on one day, two on the next. In not even one of the thirty-six catch trials, nine each per session, did FB respond incorrectly; namely, if no touch was applied, FB reported that she felt nothing. Accordingly, we encounter an explanatory problem—why did FB report feeling

the touch when primed to expect that her niece would be touched, but not when primed to expect that her left hand would be touched?

Because the causal histories of reporting and not reporting tactile sensation are nearly the same—differing only *in whom* FB expected to be touched—sensible contrastive questions that enable elicitation of causal differences can be asked (cf. Lipton 1993). In view of the controls that were in place, it seems the crucial causal difference concerned *who* was to be touched. And only when FB expected that her niece would be touched did she report tactile sensation. In other words, it seems that FB was only conscious of those sensations when she expected the touches to be applied to *someone else's* hand. It is this apparent expectation that the touch causes conscious states only when the touch is expected to be applied to someone other than self that seems to constitute a challenge for HOT and SP theories in general.

One among Gennaro's (this vol.) explicit concerns is that FB might be reporting a belief or a judgment rather than an experience. Since HOT theory allows for the possibility that HOTs can misrepresent first-order sensory states that they are about,⁷ a HOT theorist could argue that FB is reporting a belief or judgment that misrepresents what FB experiences. In this context, since somatoparaphrenia is classified as a delusion, Gennaro raises two concerns about how to interpret delusional reports: one pertains to the endorsement/explanationist distinction. On the endorsement account of delusion, patients are endorsing as veridical the content of unusual experiences; on the explanationist, they are attempting to explain unusual experiences. Gennaro's second concern relates to the distinction between "spontaneous" and "provoked" confabulations, the latter only occurring when patients are questioned directly. Regarding provoked confabulations, Gennaro then proceeds to emphasize FB's initial elusiveness when asked how she could report touches on someone else's hand. Gennaro argues that the endorsement interpretation of FB's reports is more likely appropriate only after the delusional belief has taken firm hold. Before the delusion has taken firm hold, a "provoked explanationist" account should be favored. The implication of Gennaro's interpretation seems to be that at least at first, when FB is responding to touches, it is more likely the case that she is explaining something odd, rather than endorsing as veridical the content of her experiences. If this is the case, then it would weaken the force of my criticism of HOT.

Gennaro's explication of FB's case, however, appears not to address the explanatory problem posed by the experimental results. Among other things, "provoked" is ambiguous: it is the case that in the experimental context, FB

was asked to report touches. So, one could say that she was provoked. But Gennaro's emphasis is not here; instead, he is concerned with the provoking that occurred later, when FB was queried as to how she was able to report touches on someone else's hand. Her simple responses were not problematic; one might even say they were spontaneous. It was only when she was asked to *explain* those spontaneous responses that FB seemed elusive.

Consider as well that Gennaro—along with Rosenthal, Mylopoulos, Billon, and Kriegel—all emphasize the importance of spontaneous, unmediated awareness. Gennaro accurately reports that this is essential to Rosenthal's construal of HOT theory.⁸ Indeed, FB's reports of tactile sensation when her niece's hand was actually touched appear to be instances of spontaneous, unmediated awareness.

In short, it seems that Gennaro's argument turns on an ambiguity and a red herring: a simple response concerning whether a hand was touched is not the type of provoking that Gennaro needs, and emphasizing FB's elusiveness when she was asked to explain how such touches could be reported distracts us from the explanatory contrast problem I've articulated. FB appears to have done nothing more than affirm the occurrence of tactile sensations in some instances while denying their occurrence in others. The only difference between the two cases was in how she was primed: your left hand or your niece's hand. In view of the controls that were in place, as well as the apparent absence of mediation, FB seems to merely be endorsing that a tactile experience was instantiated. That such an instantiation might be unusual and in need of explanation results from the prompting to explain how such a thing could occur. In other words, FB is reporting a sensation, not a belief. If FB was reporting a belief, she should also have done that during the "your niece" catch trials, because she would have been unable to distinguish touch from its absence. But recovery from tactile extinction only occurred (i) when she was primed in the right way and (ii) when she was actually touched.

There is, however, a glaring gap in the data, one that Gennaro, as well as Billon and Kriegel, correctly point out. Bottini et al. did not explicitly inquire as to whether the tactile sensations belonged to FB or to her niece (cf. Feinberg and Venneri 2014). My interpretation of the data is an inference to the best explanation: I am making the case for a conditional claim—if FB feels those sensations belong to her niece, that experience would best explain her ability to recover from tactile extinction only when primed in the right way and only when actually touched. But reasoning in this way obviously does not establish that FB has alienation experiences of the type described by Billon, Kriegel, and Gennaro.

Although I believe this issue might be resolvable on conceptual grounds, it should be emphasized that Rosenthal regards the HOT theory as an empirical theory that is testable (e.g., Lau and Rosenthal 2011). Likewise, Gennaro (2012, 269–302) argues that his “wide intrinsicity view” (WIV) of HOT can inform a neurophysiological research program; Kriegel (2009, 233–265) argues that his self-representational theory can answer some of the principal scientific questions of consciousness. Indeed, when first responding to Rosenthal (Lane and Liang 2010, 500), I acknowledged the desirability of beginning to monitor somatoparaphrenia patients and conducting a refined set of probes in order to more adequately address worries of the type expressed by Gennaro, Billon, and Kriegel. Here I propose a more formal attempt to operationalize HOT and treat somatoparaphrenia as a test case for SP Theories.

The relevant issue concerns what to make of FB’s differential responses to passive touch.⁹ Because I endorse Billon and Kriegel’s view that depersonalization is also a potential counterexample to SP theories, a natural place to look for refined probes of FB-like patients is *The Cambridge Depersonalization Scale* (Sierra 2009, 161–168). Question 22 on that scale typifies the type of item that could easily be adapted to serve as a suitable probe: “When a part of my body hurts, I feel so detached from the pain that it feels as if it was ‘somebody else’s pain.’” But I believe we can do much more than systematize probes of subjective report. A pincers maneuver, whereby data on subjective report and data from objective measures converge, is called for. As for the latter, the means for differentiating self from nonself are already available.

For two decades, evidence for the existence of a mirror neuron system in the motor domain has been accumulating (Kilner and Lemon 2013). Whereas some neurons modulate their activity only when a person observes the actions of others, but not when self performs those same actions, a special class of neurons modulate their activity both when executing an action and when observing that action performed by someone else. In effect, what we see mirrors what we feel. Of more direct relevance to the case of FB, mirroring has recently been identified for the experience of, and the observation of, touch: observation of someone else being touched induces activation of neural circuitry in both the primary (SI) and secondary (SII) somatosensory cortices (Schaefer et al. 2009).

The sensory overlap between what is seen happening to others and what is experienced in self presents an experimental opportunity. It is not the overlap per se that matters, it is the opportunity afforded by mirroring experiments to distinguish between self and nonself. Consider, for example, the findings of Keysers et al. (2004): in an fMRI study of the relationship

between the observation of touch in others and the experience of touch in self, they discovered that the SII was activated both when participants were touched and when they observed someone else being touched. SI, on the other hand, was just weakly activated when observing touch. The authors interpret their findings thus (2004, 342): “It might be that SII activation in the context of weak SI activation only evokes a concept of touch that is relatively detached from our own immediate bodily experience.”

If we were to apply findings of this type to an investigation of subsequent FB-like recoveries of tactile sensation in alien limbs, in order to seek confirmation of my hypothesis that patients only became conscious of certain tactile states when those states are experienced as belonging to someone other than themselves, we would need to proceed in two stages: first, we would attempt to replicate the Bottini et al. experiment, including the blindfold, the catch trials, and other controls. If we were to succeed at that, including successful recovery of tactile sensation on being primed to expect that *someone other than oneself* was to be touched, we should proceed to stage two. Here we could dispense with the blindfold and the catch trials. Instead, having in stage one already confirmed that actual contact and the right prime are necessary for recovery of sensation, we would use fMRI to compare patient (a) reactions when primed to expect that oneself would be touched with (b) reactions when primed to expect that someone other than oneself would be touched.¹⁰ Since we would now allow the patient to observe what was being done, based upon the mirroring effects described above, we would expect overlap between experiencing touch in self and seeing touch applied to someone else. If my working hypothesis is correct, I would expect that when priming to expect someone else will be touched enables recovery of tactile sensation, we will observe SII activation in the context of weak SI activation. In other words, the patient’s experience is that of observing someone else being touched, not of experiencing touch for him- or herself.

My claim is not that this would settle the issue. Rather, my intent—both here and in the remainder of this chapter—is to reorient the debate concerning HOT and all other SP theories, situating them squarely in an experimental context, since I believe we have already entered an era when significant aspects of SP theories are empirically tractable. We have amassed sufficient evidence concerning many types of alien experience, have developed novel experimental methods that enable teasing apart *self* and *nonself* experiences, and have developed technologies that enable application of those methods to both healthy and patient populations. Progress toward settling disagreements regarding how best to explain what I have elsewhere

referred to as problems of belonging (Lane 2014, 54–56) can be made in the laboratory.

As regards FB-like cases, the hypothesis and approach indicated here are not sufficient, but vicarious somatosensory activation has been explored using other technologies as well, such as EEG (Bufalari et al. 2007), magnetoencephalography (Avikainen et al. 2002), and transcranial magnetic stimulation (Bolognini et al. 2011). Moreover, subregions within SI and SII that differentially contribute to distinguishing self from nonself experiences (Keysers et al. 2010) and trait differences that affect responsiveness to observed touch (Schaefer et al. 2012) have already been identified. In sum, an experimental platform from which sophisticated probes of FB-like cases can be launched already exists.

3 A “Something Extra” Self-Representational Hypothesis of Alienation

In their survey of a select set of alienation experiences—schizophrenia, somatoparaphrenia, and depersonalization—when reflecting on the challenge that these pose for SP theories, Billon and Kriegel (this vol.) consider the possibility that the phenomenal difference between alien and nonalien states could either involve the *addition* or the *subtraction* of a phenomenal feature. In building a defense of SP, they aspire to show that alienation experiences are compatible with their preferred theory of phenomenal consciousness, a theory that is distinctive in at least two ways: first, unlike the HOT theory, here conscious states are taken to be conscious because these mental states target themselves. In a word, this is a self-representational position. Second, they emphasize that subjects *experience* these states as belonging to self: that is, awareness of a state as belonging to self “does show up in the subject’s overall phenomenology.”

They observe that the most widely favored subtraction, or “something missing,” position, is that patients possess thoughts (in the case of thought-insertion) that they did not *author*. On this view then, *belonging* and *agency* dissociate: subjects may have thoughts despite lacking a sense of agency for them (Gallagher 2000). But Billon and Kriegel proceed to object that, even for healthy subjects, many thoughts running through our minds come unbidden or feel intrusive, without causing us to experience those thoughts as distinctively alien (cf. Bayne 2010, 156–162, Billon 2013, 296–302, Lane 2012, 279–280). Therefore, it seems that the “something missing” view is explanatorily inadequate.

Motivated by this worry about the “something missing” view, they propose a “something extra” hypothesis of alienation experiences. For

example, in the case of schizophrenic thought insertions, an inserted thought instantiates many properties of a normal thought¹¹ *plus* it instantiates an extra phenomenal property—it *feels* inserted. They begin defense of their position by distinguishing phenomenal consciousness—the object of their concern—from “reflective” consciousness, a distinction that they acknowledge to be similar to Block’s distinction between phenomenal and “access” consciousness (Block 2007, 166–178, cf. Billon 2013, 305–306). In effect, what Billon and Kriegel suggest is that their version of SP theory concerns phenomenal consciousness, and alienation worries can be explained as relevant only to reflective, or access, consciousness. Billon (2013, 307) has written of such patients that they have reflective awareness of “their inserted thoughts, but ... lack phenomenality and subjectivity altogether.”

On the assumption that Billon and Kriegel’s distinction is in most important respects similar to Block’s, it is worthwhile to consider the latter’s original example of access dissociated from phenomenal consciousness, blindsight (Block 2007, 172–173). Blindsight patients, despite suffering from cortical blindness, are able to make accurate forced choices about things presented to those visual fields for which they claim to have no visual experiences. Although blindsight patients apparently have the relevant unconscious “perceptual or quasi-perceptual states,” they seem to lack both phenomenal and access consciousness. Block claims, however, that perhaps there could be “super-blindsighters” who learn to prompt themselves, such that they would guess without being told to do so. Super-blindsighters then would have access (or reflective) consciousness without having phenomenal consciousness. In other words, they would “just know,” despite not having visual experiences, rather in the way that people can just know time or direction without having any relevant conscious experiences.

Let us return to consideration of Billon and Kriegel: they argue that alienation experiences occur when a first-order, nonphenomenal, alien mental state occurs in simultaneity with a second-order, nonalien, phenomenal state that represents the first-order state. The second-order state is claimed to represent the first-order state as belonging to self, albeit without rendering the latter phenomenally conscious.¹² In this way, SP might be saved because the mental state that exhibits phenomenal consciousness also exhibits belonging, while the state that does not exhibit belonging is assigned to a distinct category of consciousness, one that is accessed but nonphenomenal.¹³

In order to cast this in less abstract terms, they propose a thought experiment: imagine that, on awakening one morning, you experience tinnitus. Over time, what was once a meaningless ringing in the ears begins to sound

like a voice, a voice that expresses repressed, unconscious states. Within the plotline of this thought experiment, it develops that you eventually realize a neuroscientist has implanted a device in your brain that monitors repressed states and translates them into phenomenally conscious states; the auditory hallucinations symptomatic of schizophrenia are taken to be one example of translated results. Accordingly, if you are in the presence of someone you have reason to dislike, albeit without being consciously aware of the dislike, the monitor might detect a first-order repressed state that it expresses with the rich auditory phenomenology of an inner voice, such as "Oh, I hate him." Billon and Kriegel further claim that although the reflected-on states belong to you, you experience them as alien.¹⁴

The first concern about this explanatory framework is that it is ad hoc for advocates of a self-representative view to invoke second-order thoughts that somehow translate first-order thoughts into phenomenally conscious states, without making those first-order states phenomenally conscious. Second, one of Billon and Kriegel's goals is to explain the phenomena while taking patient reports more or less at face value, but it is not clear that the reflective-consciousness argument achieves this goal. If the nonphenomenal first-order state is represented by the second-order phenomenal state as belonging to self, then wherefrom comes the experience that the voices one hears are alien?¹⁵ Perhaps the idea is that the alien experience is explainable as owing to the repressed nature of the unconscious state. But if that is the idea, the conjectured relationship between repression and alienation must be explicated, and that would not be easy, because although repressed thoughts or the processes of repression can become conscious, there is no evidence to suggest that such thoughts or processes are thereby experienced as alien (cf. Boag 2010).

Since on this view the first-order state is nonphenomenal (intrinsically), and since it is represented as belonging to the person who experiences the auditory hallucinations, the alien dimension seems to be lost from the subjective experience. SP's metaphysical commitment to "for-me-ness" as appearing in a subject's overall phenomenology can perhaps be defended in this way, but then it is not clear that patient reports of their phenomenology are being taken at face value. In other words, the reflective-consciousness view might help explain how the auditory hallucination "Oh, I hate him" can be experienced as being for-me; it is not clear, however, in what respect this framework helps to explain how a nonphenomenal state represented by the auditory hallucination can be explained, if we are committed to taking seriously patient reports that these voices seem alien.¹⁶ One way of expressing this worry is that it seems a something extra has

been added to the explanans; meanwhile, something has been subtracted from the explanandum.

Nevertheless, since Billon and Kriegel acknowledge that their reflective/phenomenal distinction is similar to Block's access/phenomenal distinction, and since efforts have been made to identify the neural correlates of these two types of consciousness (Block 2005), it should be possible to adduce empirical evidence to assess the neural plausibility of their hypothesis. In particular, some progress has been made with respect to the neural correlates of visual consciousness (Block 2005, 47–48); evidence suggests that the neural basis of *access* involves activation of the superior parietal and the dorsolateral prefrontal cortices. For example, binocular rivalry occurs when two distinct patterns are shown, one to each eye of a participant, in these bi-stable perception experiments. Because the two patterns are so different, the brain does not fuse them; instead, a rivalry ensues. Subjects see one pattern for a few seconds; then the other, for a few seconds; after that, the visual experience shifts back again. Subject reports of conscious contents correlate with activation in these frontal and parietal areas.

The details of bi-stable perception need not detain us here, and it must be admitted that much work remains to be done before one can assert with confidence that the neural correlates of any given instantiation of access consciousness have been identified. But experimental work carried out to date suggests a critical role for the frontoparietal network. A starting point then in the search to seek empirical confirmation of the Billon-Kriegel framework would be evidence of fronto-parietal activity in the absence of phenomenal consciousness. This seems to be a reasonable expectation, at least on the assumption that not all first-order mental states instantiated within schizophrenics are represented by second-order states. And, because the Billon and Kriegel thought experiment concerns first-order states that are repressed, if we take repression as paradigmatic, it is necessarily the case that most first-order states are never expressed on a phenomenal level (cf. Boag 2010, 174), irrespective of whether the relevant phenomenology is extrinsic or intrinsic.

The problem is that not only is there no evidence of fronto-parietal activity in the absence of phenomenal consciousness, but most neuroscientific studies show conscious awareness of an event requires recruitment of widespread brain activation. Frontal and parietal areas, in particular, are implicated (e.g., Baars 2007). In other words, even though it may be conceptually possible for access to occur without phenomenal consciousness, there is nothing to suggest that it is empirically possible.¹⁷ On the contrary, it seems unlikely, at least, that is, if the nonexistence of super-blindsight can be regarded as instructive. If, on the other hand, the suggestion is that

the relevant second-order states are intrinsically wedded to the first-order states represented, then we would need some account of why and how these second-order states sometimes occur.

Taking the repression thought experiment as a literal expression of the main idea, the crux of the issue is that reflective- or access- consciousness seems to implicate fronto-parietal activity, but this pattern of activity seems to play no distinctive role in repression. What evidence there is concerning repressed impulses, thoughts, or desires does not implicate fronto-parietal activity (Berlin 2011). We can, however, infer that the amygdala plays an important role in repression (Berlin 2011, 15), and separate evidence suggests that hyperactivation of receptive language areas in the left temporal lobe mediates auditory verbal hallucinations (Hugdahl et al. 2012).¹⁸ If we consider the Billon-Kriegel hypothesis against this backdrop, it would seem that the repressed first-order state must somehow involve the amygdala, and the translation of that state must somehow involve the left temporal lobe. But if these regions are the ones that we can, with some measure of confidence, claim to mediate repression and hallucination, it remains unclear how the seemingly essential fronto-parietal activity could be assimilated to explanation of alien experience.

In sum, three principal worries append to the reflective-consciousness hypothesis as it applies to first-rank schizophrenic symptoms. First, it is ad hoc. In Ptolemaic fashion it adds an ingredient—one that carries a not inconsiderable amount of conceptual baggage—to the self-representational theory of consciousness solely in order to account for a phenomenon that self-representational theory itself is unable to explain. Unfortunately, unlike Ptolemy's epicycles, the something extra here seems not to increase our explanatory leverage. Second, by failing to specify how repression is related to alienation, it seems to have omitted the phenomenon that is in need of explanation. Although the conceptual footwork might save the theory, the cost seems to have been a sacrifice of the phenomenon that stands in need of explanation. And, third, the hypothesis seems to lack neural plausibility. Just as there is no empirical evidence to suggest that Block's super-blindsight is ever instantiated, so too there is no empirical evidence to suggest that the neural substrates of reflective and phenomenal consciousness ever dissociate or interact in the requisite way.

So this strategy seems not to ease the explanatory burden of accounting for schizophrenia's first-rank symptoms. But might it help with other types of alienation? Billon and Kriegel suggest that the reflective-consciousness hypothesis can be extended to depersonalization. Indeed, they argue that it might work even better here, because "some depersonalized patients

explicitly affirm that their alien states are unconscious even though they are conscious *of* them." And it is in fact the case that some patients describe their experiences in such terms.

But as Billon and Kriegel observe, the feeling that one's "emotional phenomenology is blunted or absent" is a *far more common* symptom of depersonalization. It seems these symptoms of depersonalization are mediated by "fronto-limbic suppression" (Sierra 2009, 146): that is, the amygdala, the anterior insula, and perhaps other limbic areas are suppressed due to abnormal prefrontal regulatory activity. Not only is this suppression hypothesis supported by abundant experimental evidence, it would also help explain most of depersonalization's symptoms.

Nevertheless, the symptom that is of most concern to Billon and Kriegel, however, does seem to be consistent with their reflective-consciousness hypothesis, so it warrants close examination. Recall that, according to their hypothesis, it can be the case that "alien states are unconscious even though (patients) are conscious *of* them" (this vol.). Bearing this in mind, take note of the patient's descriptions—"I suddenly wonder: is it really me here? Is it really me walking?" This is followed by what they regard as a highly significant passage: "Then I make enormous efforts in order to apply my consciousness to this unconsciousness ... in order to realize that I am making the walking movements." Is this truly a case wherein the patient has access to a nonconscious state, somewhat like a super-blindsighter, perhaps the equivalent of forcing oneself to guess?

First, take note that the patient's description concerns the act of walking. Second, further note that currently the most influential theory of depersonalization is the "two-neural-network" model (Sierra 2009, 146). The first network is the fronto-limbic suppression network described above. The second network involves parietal regions that seem to mediate the experience of embodiment and agency. For example, elevated activation in the angular gyrus and decreased activity in the posterior insula have been observed in patients who report the absence of agentive feelings. Those patients who exhibit decreased activity in the posterior insula report so striking an absence of agentive feeling that when they move it seems "they are watching the movements of another person."

Now once again consider (i) the hypothesis that "alien states are unconscious even though (patients) are conscious *of* them" along with (ii) the patient who, while walking, wonders whether it is really self who is walking, and who must exert strenuous effort in order to be certain that it is in fact self. It appears to be the case that what troubles this patient is the *loss of agency*. My suggestion here is that this case is adequately explained

by the second component of the two-neural-network model. Naturally, were we to encounter a patient who described symptoms in these terms, in order to confirm my hypothesis, we would want to check angular gyrus and posterior insula activity. But if the *absence* of agency is the critical factor here, then the Billon and Kriegel attempt at developing a “something extra” explanation fails.

It might appear to be the case that now I am omitting something important from the explanandum, to wit—the patient’s effort “to apply my consciousness to this unconsciousness.” But this omission is only apparent. For quotidian instances of action, I see no reason to presuppose that a conscious agentive state is instantiated (Lane 2014, 64–69).¹⁹ On this characterization then, one could still say that the patient is conscious of something unconscious, at least in the sense that the patient is aware that something is amiss, an awareness that precipitates inferring the absence of a typically unconscious ingredient. Therefore, we can still take the patient’s description at face value, but the hypothesis is markedly different than that of Billon and Kriegel. Here I am not positing a reflective- or access- state that is somehow independent of a phenomenal state. Instead, I am suggesting that there are agentive mental states (whether conscious or not), and that these can go missing when the angular gyrus exhibits elevated activity and the posterior insula, decreased activity. This interpretation might even be a bit closer to Block’s example of access- without phenomenal- consciousness: there do seem to be times when we “just know” that something is missing or that something has changed.²⁰

Regarding the explanandum, one final point remains to be made. Billon and Kriegel have also assembled a selection of depersonalization self-descriptions that suggest the patients are “totally unconscious” in a way that is “particularly amenable” to their hypothesis. It is the case that one patient records “I’m like a zombie”; another, “I am in emptiness.” Indeed these descriptions might be amenable to the “totally unconscious” characterization. But it is obvious that these are tropes, attempts at applying natural language to experiences for which it was not designed. My claim is that not all tropes are alike: some characterize the phenomenon in question with greater accuracy than do others. Of the examples cited by Billon and Kriegel, I submit that a more accurate trope is “it is not me who feels.” This seems not to be a denial that conscious feelings are instantiated; rather, it seems to be an assertion that there are conscious feelings of which I am directly aware that do not belong to self.²¹

I believe this is not just a matter of cherry-picking. The principal reason is that converging lines of evidence suggest that being uniquely situated to

report instantiation of a conscious experience dissociates from belonging or personal ownership (Lane 2012, Klein 2014). Unlike the tropes cited by Billon and Kriegel, these descriptions appear to have identified a distinctive conscious experience that has previously gone underreported. Neither “I’m like a zombie” nor “I am in emptiness” can be said to enjoy such a felicitous fit with an emerging body of scientific description.

Regarding the explanans, likewise, one final point remains to be made. Sierra (2009, 143) opines, regarding the first component of the two-neural-network model of depersonalization, that “an ‘emotion coloring’ mechanism is likely to be a major contributor to feelings usually described in terms of ‘immediacy.’” Billon and Kriegel, along with most others who weigh in on consciousness, take “seemingly immediate access” be part of the explanans of “consciousness,” part of what makes a state conscious. I believe this view to be mistaken. Sierra here suggests one component of felt immediacy; I believe there are others (e.g., Lane 2012, 258–259). Felt immediacy, I submit, does not so much inform as regards what makes an experience conscious, as it does regarding to whom that conscious experience belongs.

4 A “Something Missing” HOT Hypothesis of Alienation

Mylopoulos’s concern (this volume) is to explain alien action, in particular schizophrenic passivity experiences, anarchic hand syndrome (AHS), and utilization behavior.²² As her point of departure, she presupposes that in nonpathological or nonaberrant cases there is a “subjective sense of performing the action at the time”—what is variously referred to as “phenomenology of agency,” “agentive awareness,” or “action consciousness”—something that she takes to be missing when alien actions occur.²³ Action here is taken to be a bodily event that is “suitably caused by an appropriate goal state or set of goal states (e.g., intentions, desires, reasons, motor commands).” Action consciousness, then, is determined by some property—an “agentive quality”—that “belongs to” the bodily movement or the goal state associated with the action. She considers various candidates for the missing “agentive quality”—for example, proprioception or sensory attenuation—but dismisses both. Instead, she derives a proposal for understanding agentive quality from a version of the HOT theory according to which action consciousness results from “thinking that one is A-ing on the basis of an intention in action to A.” One supposed virtue of HOT theory in this context is that it can explain the phenomenal character that she presupposes to be “central to action consciousness,” because all that

matters is whether the relevant HOT “represents oneself as acting.” In sum, if a HOT of this kind is missing, then action consciousness is missing, and the attribution of agency can go awry.²⁴

In developing her hypothesis, Mylopoulos draws a distinction between two types of action control, intentional and sensorimotor. For the latter, action is guided primarily by motor commands that specify fine-grained features of movement—for example, angle of trajectory and grip aperture. For intentional, movements are “guided by way of intentions that relate to any agent’s practical beliefs and desires.” She also emphasizes that the relevant sense of “intentional” here is “intention in action,” a concept borrowed from Searle that underscores *present*-directedness; in other words, intentions so understood are “the direct mental antecedents of bodily movements involved in action.” As this distinction applies to the pathologies considered here, the claim is that alien movements are guided by motor commands, but not by “intentions in action at the level of intentional control.”

Although acknowledging that in most cases the sensorimotor and the intentional interact, Mylopoulos argues the two can dissociate and that AHS and utilization behavior are examples of such dissociation. AHS and utilization behaviors, on this view, seem not to be preceded by intentions in action. In a word, *something is missing*. It is for this reason—the absence of intentions in action—that these behaviors are felt to be alien. And, the absence of intention in action is due to the absence of a higher order thought “that one is A-ing on the basis of an intention in action to A.”²⁵

One among the reasons that Mylopoulos recruits HOT theory for her attempt to explain these aberrant behaviors is that—like Gennaro, Billon, and Kriegel—she accurately notes that, on Rosenthal’s theory, thoughts of the right sort must be direct or noninferential (e.g., Rosenthal 2002, 408–411). Applying this hypothesis then to AHS, she says that subjects feel the behaviors to be alien because they lack the feeling of immediacy, the sense that awareness of these behaviors is “subjectively unmediated.” And this absence of immediacy is due to the absence of a HOT “that one is A-ing on the basis of an intention in action to A.” For patients with AHS, the anarchic behaviors are just what they observe, from the outside; hence, awareness of these behaviors is indirect or mediated.

Why should this matter? Mylopoulos emphasizes that speed counts. If formulation of a belief about action were dependent upon *observing* self in action, such mediation, or the inferences that need to be made, would cause us to “lose valuable time.” But here is where the difficulties begin.

First, Mylopoulos emphasizes “the *subjectively* unmediated, intention-based higher-order thoughts.”²⁶ But the *subjective sense* that something is

unmediated does not imply that it is unmediated. More importantly, since on Rosenthal’s theory HOTs can radically misrepresent (Lane and Liang 2008), there is no necessary connection between how things seem and the objective passage of time. What Mylopoulos needs—something HOT theory does not provide—is objective rather than subjective speed and efficiency. Note too that Rosenthal (e.g., Lau and Rosenthal 2011, 366) takes pains to emphasize that HOT theory is neutral as regards whether conscious awareness “adds significant utility or immediate impact on behavior and task performance.” He proceeds to emphasize that for most cognitive and perceptual tasks, performance does not depend upon higher-order representations; in fact, “because conscious awareness can differ even if all first-order representations remain completely unchanged, such awareness itself might serve little function.” If speed is a critical issue, and if HOT theory cannot help with explaining speedy action responses in ecological contexts, then one wonders why insist upon a “subjective sense of performing the action at the time,” especially given that consciousness is notoriously slow (Dehaene 2014, 115–160).

Second, to support her view that AHS results from sensorimotor rather than intentional action control, which the agent can know of only on the basis of “conscious observation,” she cites the case of JC, who appears only to be aware of anarchic movements when he sees them. Of course this concerns a straightforward empirical issue, but at the current stage of knowledge there is no consensus regarding the facts of the matter. Indeed, Marcel (2003, 81–82) describes a patient who “was aware of the anarchic actions performed by his left hand even when out of his sight.” So even though conscious observation was not involved, the patient reported feeling “as if someone else was doing the actions.”

Third, the degree of efficiency that can be achieved when depending on observation of one’s own body, while acting, has yet to be fully explored. But the case of IW’s neuropathy is instructive in this regard (McNeill et al. 2010). When he was nineteen years old, as the apparent result of an auto-immune reaction triggered by a severe fever, IW suffered a deafferentation of his body from the neck down: that is, he underwent a nearly complete loss of peripheral sensory feedback, including tactile and proprioceptive sensations. Although neither was he paralyzed nor was his motor system affected, unless he could look to see what his body was doing, he had no control over what his body did. In a word, visual observation of his body became essential to the exercise of control over his body—everything from simple sitting up or walking, to the complex actions required for his vocation.

Although the initial prognosis was that IW would be confined permanently to a wheelchair, he designed a program aimed at relearning how to move. Using vision to guide him, while carefully thinking about each motion, he practiced moving his body, repetitively and in various combinations—different trajectories, distances, and velocities. Despite this constant need to maintain visual contact with his body and the environment, especially with objects that he must manipulate, IW has learned how to perform at levels all but indistinguishable from healthy persons.

The point of citing IW's case is not to deny the distinction that Mylopoulos draws between intentional and sensorimotor control. Indeed, Mylopoulos agrees that in many cases "these two types of control interact in rich ways." But I believe she overstates the role that dissociation between these two, or the idea of "losing contact" with self, might play in explaining phenomena like AHS. No doubt IW's effective training incurs a cognitive and perceptual cost in the distribution of mental resources—he must devote resources to motor activity that most of us can reserve for other purposes. Nevertheless, he is capable of maintaining contact with and control over his body, visually. The degree to which IW recovered suggests that intentional and sensorimotor control might be so seamlessly connected to one another that Mylopoulos's explanatory strategy is blocked.

Fourth, Wegner's (2002) Ouija-like experiments that suggest we can quite easily be fooled into thinking that we are acting when that could not possibly be the case are also relevant here. What most concerns with regard to Mylopoulos's "intention in action" hypothesis is Wegner's (2002, 179) suggestion that much of human behavior "seems to occur without much influence by intentions, especially when the behavior is not particularly discrepant from prior beliefs." Wegner's (2002, 180) experiments seem to show that many of our reported intentions are "post hoc inventions" or "fabrications" that depend more upon an idealized image of self as agents than upon actual choice of actions "with foreknowledge and in accord with our conscious intentions." As for the case of IW, it seems that vision plays *more* of a role than Mylopoulos's hypothesis implies; Wegner's experiments, on the other hand, seem to show that intention in action plays *less* of a role than Mylopoulos's hypothesis suggests.

Finally, perhaps we are at a moment in time when the hypothesis can be operationalized and tested. First, although strictly speaking, HOT theory does not specify neuroanatomical detail (Lau and Rosenthal 2011, 366), the relevant higher-order representations might be mediated in virtue of dorsolateral prefrontal (DLPFC) activity (Lau and Rosenthal 2011, 367–370).²⁷ Second, since fMRI studies have shown that intentions in action can be

analyzed to a sufficiently fine-grained level of detail such that brain activity can distinguish among distinct grasping actions prior to the initiation of those actions (Gallivan et al. 2011), it should be possible to devise a suitable action-involving paradigm. Third, neurodisruption techniques (e.g., transcranial magnetic stimulation [TMS]) could be applied in order to determine whether or not targeting a specific area (e.g., the DLPFC) interferes with the hypothesized HOT in such a way as to induce an experience of alien action. And, fourth, TMS has already been successfully applied to the DLPFC, effectively creating patterns of neurodisruption (Bilek et al. 2013).

Briefly, TMS makes it possible to assess the role of brain regions in cognitive activity, by creating virtual lesions: that is, it can be targeted to specific brain areas, causing temporary disruptions of activity, thereby providing information about the functional relevance of a brain region (e.g., the DLPFC).²⁸ Sidestepping many technical details for the nonce, what I am suggesting is that repetitive TMS be applied within an action paradigm, to the DLPFC, and at different parameters, in order to both enhance and inhibit cognitive processes. The hypothesis to be tested is Mylopoulos's application of HOT theory to action: if that hypothesis is correct, parameters that enhance relevant cognitive processes should enhance the feeling of control, and those that inhibit should diminish the feeling of control or engender a sense of alien action.

In fact, recently, Dienes and Hutton (2013) have taken the first steps in this direction. Their concern was not with action per se; instead, they were testing a HOT interpretation of the cold control theory (CCT) of hypnotic suggestion. The specific hypnotic suggestion in this instance was magnetic hands; experimenters suggest to subjects that their hands are like magnets, and are thereby able to attract or repel one another. According to the HOT interpretation of CCT, hypnotic responses are constituted by (i) intentions to perform actions, even though (ii) hypnotized subjects think they do not intend those actions. The phenomenology of hypnosis, the feeling that the arms move by themselves, results then from intentions of which one is not aware. To express this idea in terms of HOT theory: "hypnotic response is all due to the formation of inaccurate ... HOTs" (Dienes and Hutton 2013, 387). On the assumption that the left DLPFC mediates the formation of *accurate* HOTs, Dienes and Hutton applied low-frequency repetitive TMS to the left DLPFC, predicting that the resulting disruption of activity would enhance the hypnotic effect, because by hypothesis disruption would increase the likelihood of inaccurate HOTs being formed (cf. Rounis et al. 2010). Indeed, the authors did find some evidence to indicate that TMS-induced neural disruption of left DLPFC activity enhanced hypnotic

response, thereby suggesting inaccurate HOTs might be playing a role in the alien experiences associated with hypnotic experiences.²⁹

The intent in citing this experiment is not to imply that Mylopoulos's hypothesis has already been tested. The intent is only to add substance to the claim that the hypothesis is already empirically tractable. Doubtless though, a suitable test would need to consider the relationship between the formation of inaccurate HOTs along with the formation of actual intentions,³⁰ and both of these in the context of a paradigm devoted not to suggestion as such, but to action.

It goes without saying that many additional details would need to be taken into consideration. What is more, as is true for all empirical hypotheses, the results of one set of experiments could be used to adduce support for or against a hypothesis, but such results could not be claimed to settle the matter once and for all. Nonetheless, I think the time has arrived that we should begin adding a set of constraints to the many conjectures on offer in this vicinity. The constraints I have in mind are those that are commonplace when one seeks to operationalize and test in experimental settings.

5 Concluding Remarks and the Subpersonal Self

My first major concern is that each of the theories discussed above presupposes that consciousness includes an element of self-reference. I have argued that this presupposition is erroneous whether that self-referential element is conscious or not, but my focus here is on the gratuitous explanatory burden created by theories that attempt to explain both (i) the experience of alienation and (ii) the experience of belonging. If it were the case that strong evidence existed for the simultaneous occurrence of (i) and (ii), this would then be a necessary explanatory challenge. We should not subtract from the explanandum for the sake of convenience. But although it is clear that (i) occurs in certain pathological, illusory or otherwise atypical conditions, it is not obvious that (ii) is a necessary component of all conscious experiences. In fact, there are reasons to be wary of presupposing the latter (e.g., Prinz 2012, 213–240).

Kriegel (2009, 121, n. 32) acknowledges that he is aware of no experimental evidence showing that “phenomenal consciousness involves for-me-ness.” But he (2009, 175) “cannot envisage what it would be like to have a phenomenology lacking the kind of inner awareness that constitutes for-me-ness.” Although I do not share his intuitions, I do share his concern that when considering the phenomenology of consciousness theorists should avoid glib rejection of one another's phenomenological pronouncements,

if these are based solely upon dissonance between what different persons envisage. After all, as Kriegel rightly avers, “there are certainly facts of the matter pertaining to phenomenology,” and these are critical to getting clear about the explanandum.³¹

Where we disagree is on how to regard these issues from the perspective of scientific explanation. Although there is much that distinguishes early views of explanation from more recent views, the two converge on a willingness to allow for the addition of entities or processes whose existence is uncertain, just so long as they enhance explanatory adequacy; such liberality though is usually not extended to the phenomena under scrutiny (e.g., Hempel 1965, Craver 2007). As this relates to alien experiences, I submit that many of the difficulties we encounter when trying to achieve adequate explanation derive from having inflated the explanandum.³²

Concerning the explanandum, Kriegel pitches his view as a conditional (2009, 67): “if the phenomenology has the features I say it does, *then* self-representationalism is true.” As to whether the antecedent of this conditional is true, two empirical points can be made. First, there appear to be an ever-increasing number of counterexamples to the for-me-ness feature of the phenomenology. The seeming counterexamples I have in mind are not restricted to somatoparaphrenia, passivity experiences, or depersonalization: they include visual experiences (Zahn 2008), nonpathological “switching” away from self and back in a way that corresponds to changes in hemispheric dominance (Gott et al. 1984), episodic memory (Klein 2014, 103–109), fibromyalgia pain (Valenzuela-Moguillansky 2013), pain asymbolia (Klein forthcoming) and so forth (cf. Lane 2012). But we can do more than cherry pick examples from the scientific literature; we can design experiments. Recall that Kriegel “cannot envisage what it would be like to have a phenomenology lacking the kind of *inner awareness* that constitutes *for-me-ness*.”³³ Now consider the possibility that (i) the neural substrate of “inner awareness” supervenes on our primary interoceptive system (Craig 2003), and (ii) it is in virtue of a representation of the primary system realized in that anterior insula that we have “the feeling that ‘I am’” (Craig 2009, 65). If Kriegel's “inner awareness” approximates the neural substrate that Craig identifies as mediating interoception, and if “the feeling that ‘I am’” approximates “for-me-ness,” then it should be possible to begin testing whether, for example, anterior insula activation that is hypothesized to be essential for (2) ever fails to occur when we have conscious experiences. Once again, the claim is not that this would settle the issue, but it would enable us to begin moving beyond comparison of what one another envisage as possible.³⁴

My second major concern is that each theory of consciousness considered in this chapter—HOT, wide intrinsicity, self-representational, or HOT as applied to agency—emphatically takes *seeming* directness, immediacy, or noninferential access to be an integral component of conscious experience.³⁵ Moreover, they link experienced immediacy to the position that phenomenally conscious states all “contain a crucial self-referential element” (Gennaro 2006, 221). Although it may be the case that, somehow understood, *experienced immediacy* plays a role along the lines suggested by these theories, experienced immediacy does not entail self-reference or belonging.

For those whose intuitions incline them to endorse the presupposed link between immediacy and self-reference, consider the example of craniopagus twins who are connected at the thalamus (cf. Lane 2014, 55). Although these sisters have not been tested under controlled conditions, they seem to share interoceptive or introspective access to conscious experiences such as thirst. But shared access does not imply inability to distinguish between sensations “that belong to self and those that belong to her sister” (Bor 2012, 29).³⁶ Whereas typically when reflecting on whether two persons could share conscious experiences philosophers have had to resort to thought experiments (e.g., O’Brien 2007, 206), craniopagus twins appear to show that not only what Hirstein (2012) refers to as “mindmelding” can occur, but also that it can dissociate from belonging. It seems to be the case that one sister can be aware of thirst without referring that sensation to self.³⁷

Felt immediacy, however, does not only dissociate from belonging during interoception or introspection; it can be experienced even when we are observing the external world and exhibit no confusion as regards belonging attributions. We are extremely sensitive to certain signals in the external world, such as changes in the scleral field size (eye white area) of conspecifics: that is, we quickly and efficiently detect fear and certain other emotions in others (Tsuchiya et al. 2009, Hardee et al. 2008, Yang et al. 2007). But we do not misattribute the conscious experience of fear to self. Naturally, the neural substrates of personal and observed fear overlap in important respects, but in quotidian circumstances we can easily distinguish between the two. The essential point is that it is not obviously the case that self-reference is intrinsically bound to *seeming* immediacy, directness, or noninferential access.

I suspect that the ill-advised emphasis placed on felt immediacy arises because SP theories, albeit in distinct ways, link consciousness to self-reference, somehow understood, which in turn suggests a link to self-knowledge. And knowledge of self’s mental states is commonly claimed to be

epistemically direct, immediate, or noninferential (e.g., Macdonald 2009, 741). But some cases suggest that conscious states can occur when felt immediacy is absent (e.g., Zahn et al. 2008, Sass and Parnas 2003, 438). In sum, my concern is not only that felt immediacy does not entail belonging; I also suspect that the presupposition of a link between self-reference and consciousness misleads us in our efforts to characterize that which we hope to explain.

My third major concern is that amid the debate over whether it is preferable to add or subtract ingredients from the explanans, perhaps we are overlooking an alternative, ecumenical option. Elsewhere I have argued that distinct “molecular” or “isomeric” arrangements play a role in causing alienation experiences (Lane 2012, 2014). What happens in quotidian cases is that mental states cluster in specific ways—for example, pain’s sensory-discriminative component seems intrinsically bound to its affective-motivational component. In a word, when we feel pains we don’t like them. But sometimes the affective-motivational component is *subtracted* from the overall experience (e.g., pain asymbolia) such that we no longer care. The pains don’t bother us; they might even feel alien. Alternatively, in the quotidian case, actions exhibit intentional binding and sensory attenuation. In a word, when we act, sensory experiences are either altogether absent or vanishingly thin. But sometimes something is *added* to actions: that is, sensory experiences are accentuated. When these sensory experiences are added, actions can be felt to be alien (e.g., passivity experiences). In short, the addition and subtraction hypotheses do not conflict with one another; instead, they can be combined in a comprehensive model of alienation experiences.

What I am suggesting is that when standard expectations concerning how mental states should cluster are confounded, the likelihood of alienation experiences is high. I say the likelihood is high, because alienation is not guaranteed by the mere awareness that mental states are clustering in unexpected ways. For example, we can be aware that mental states are clustering in atypical ways when we observe people with whom we are intimately familiar, as in Capgras syndrome. Capgras syndrome appears to be importantly analogous to pain asymbolia, in that sensory-discriminative components are intact, but the affective-motivational components are missing. Unlike pain asymbolia, however, here the problem seems to be that a person with whom we are deeply familiar seems alien. It is not an instance wherein conscious experiences that we host seem not to belong to self. Therefore, I previously recommended that when seeking to explain alienation experiences, not only do we require awareness that mental states are clustering in unexpected ways, we also need to distinguish among stimuli

(either extero- or interoceptive) that are related to self, to persons with whom we are familiar, and to persons or things that are related neither to self nor to an intimate. Only when an awareness of confounded expectations is accompanied by self-referencing stimuli do alienation experiences occur.

It may appear that I am now preparing to recant on my rejection of the role that self-reference is proclaimed to play in SP theories, but that is not the case. First, I am only claiming that self-reference, paradoxically, plays an important role in alienation experiences. Second, Kriegel's (2009) self-reference is part of the conscious experience and, though Gennaro's (2012) is not part of the conscious experience, it is characterized as an unconscious psychological phenomenon. But the type of self-reference that matters to explaining alienation experiences is neither conscious nor psychological; instead, it is a neuronal or subpersonal process. Baldly, this is a subpersonal self.

Northoff and Bermpohl (2004) and Northoff et al. (2006) have previously argued that stimuli, irrespective of sensory modality, if related to self, are processed in virtue of neural activity in the brain's cortical midline structures (CMS). Many additional findings have followed in the wake of this discovery: First, Northoff et al. (2010) have adduced evidence to suggest that high resting state activity (RSA)—activity that occurs in the absence of external stimuli—is prevalent throughout the brain and that it can both shape and be shaped by stimulus-induced activity. Second, Qin and Northoff (2011) have shown that the perigenual anterior cingulate cortex (PACC) is specifically involved in the processing of self-referential stimuli, in a way that clearly distinguishes self from that which is familiar or that which concerns stimuli of other types. Third, Schneider et al. (2008) discovered “overlap” between RSA and self-referential activity in anterior CMS, such as the PACC; in these regions, high self-referential activity correlates with less deviation from RSA, relative to low self-referential activity.

Building upon these and related findings, Huang et al. (2014) have investigated vegetative state (VS) patients. Although VS patients exhibit no indication of purposeful behavior, language comprehension, awareness of sensations, or of self, an *active* paradigm was used. The questionnaire comprised both self-referential (e.g., “Have you been to Taiwan?”) and nonself-referential (e.g., “Are there sixty minutes in one hour?”). The “task” itself consisted of four fMRI scanning runs, and each run comprised twenty self-referential and twenty nonself-referential questions. Findings were striking: first, the greater the PACC signal change during self-referential, compared to nonself-referential questions, the higher the degree of consciousness.³⁸ Second, RSA was lower for the patients than for the control subjects. And,

third, two of the patients who exhibited the highest signal changes in the PACC recovered two months after the fMRI scanning.

Obviously, much more would need to be said about this experiment and its findings, and experimental probes of this kind are in need of refinement. But the principal implications of relevance here are these: because of the RSA's unique relationship with the self, if self-referential questions are to elicit distinctive neural activity, RSA must have achieved a minimum threshold. If that threshold has been achieved, even for patients with serious disorders of consciousness, PACC signal changes evoked in response to self-referential questions can be used to predict the degree of consciousness. And, strikingly, these PACC signal changes—these distinctive reactions to self-referencing stimuli—might serve as predictive markers of future capacities for consciousness.

Above I argued that theories of consciousness that presuppose conscious or psychological self-reference all fail to adequately explain alienation experiences. But studies of VS patients suggest that self might be related to consciousness in a manner previously unrecognized. It seems that self is *prior* to our experience of the world: recall, self overlaps with the resting state, the state in which subjects are not being exposed to external stimuli.³⁹ What is more, self, understood in this neuronal sense, seems to be essential to conscious experience: the *subpersonal self* exhibited in VS patients—PACC activation in response to self-referential questions when subjects appear to be wholly unconscious—is, potentially, an indicator of whether the capacity for consciousness is recoverable. Simplifying, self is intrinsically related to RSA; RSA is a precondition for self-reference; self-referencing precedes recovery of consciousness. In this respect, the subpersonal self is essential to conscious experience. If a slogan is called for, self is prior to consciousness.

I opened this essay with an epigraph from Rilke: “nothing alien happens to us, but only what has long been our own.” There is, I believe, a sense in which this is true. Alienation experiences are robust conscious experiences. But if the model I propose is true, they only occur when stimuli interact with the brain in a way that implicates self. On a neuronal level this self-referencing occurs when stimuli interact with the brain in such a way as to evoke just minimal deviation from RSA; it is in virtue of this close matching of external stimuli to the RSA (the brain's intrinsic activity) that stimuli are perceived as and judged to be self-referencing. Because neuronal self-referencing is essential to the having of alien experiences, it can truly be said that what seems alien has long been our own. Indeed this is why alienation experiences are robust: expectations concerning what to

expect from self-referential stimuli are confounded, and the confounding of expectations occasions a thickening of conscious experience.

Acknowledgments

I express my heartfelt gratitude to Rocco Gennaro, Alexandre Billon, Uriah Kriegel, and Myrto Mylopoulos for their generosity in allowing me to read early and penultimate versions of their manuscripts. I am also grateful to David Rosenthal for many helpful discussions concerning his higher-order thought theory of consciousness, as well as to Georg Northoff and Pengmin Qin for discussions regarding technical aspects of experimental research on what I refer to as the subpersonal self. Funding for this research was, in part, provided by National Science Council of Taiwan research grants, 100-2410-H-038-009-MY3, 102-2420-H-038-001-MY3, and 104-2420-H-038-001-MY3.

Notes

1. Previously, I discussed this case in Lane and Liang 2010, 2011 and Liang and Lane 2009. Rosenthal (2010) responded to those interpretations, concerning their possible implications for his higher-order thought theory of consciousness, and developed a response that is partially endorsed by Gennaro (this volume). I later argued that this case can be understood within a comprehensive explanatory framework for *belonging* or *mental ownership*, a relationship between selves and conscious experiences, whether those experiences pertain to mental states, actions, or bodies (Lane 2012, 2014).
2. For more detail concerning Gennaro's views on the kind of self-reference essential to conscious experience, see 2012, 103–134, 220–221. I return to this issue in the manuscript's concluding section.
3. It should be noted that Gennaro's version of HOT theory differs from Rosenthal's in certain important respects; for a summary of the former's "wide intrinsically view" (WIV) see Gennaro 2012, 55–59. What matters for our purposes here is that Gennaro is committed to a self-referential condition: for example, "what makes mental states conscious is *intrinsic* to conscious states ... a kind of *inner* self-referential and relational element is ... present *within* the structure of such states" (2012, 55).
4. See Liang and Lane 2009 and Lane and Liang 2010.
5. Billon and Kriegel also suggest that degree of methodological rigor might be a concern here. With this too, I am in agreement, and that is why I take the Bottini et

al. (2002) case to be just a starting point or motivation for further inquiry (Lane and Liang 2010, 500). I expand on this point below.

6. Italics added.
7. HOT theory even allows for the possibility of what I have elsewhere dubbed "radical confabulation" (Lane and Liang 2008)—representation of first-order states that do not even exist. See Gennaro 2012, 59–70, for some of his views on misrepresentation and "targetless" HOTs.
8. As for the passage from Lane and Liang (2010, 498) that concerns Gennaro, the point I intended there was not to deny that this is an essential component of Rosenthal's theory. Instead, my point was that even if spontaneity and the absence of mediation are critical to understanding consciousness, this alone does not imply that FB's HOT represents recovered tactile sensations as belonging to her.
9. For the distinction, as revealed by functional magnetic resonance imaging (fMRI), between active and passive touch, see Gardner and Johnson 2013, 522.
10. Applying touches to a patient's hand such that the patient can see what is being done, while undergoing fMRI, does pose many technical difficulties. But that these can be overcome is suggested by the methodology adopted by Ehrsson et al. (2004) in their fMRI study of the rubber hand illusion.
11. Billon (2013, 307) describes inserted thoughts thus: They differ "from the patient's ordinary thoughts by being phenomenally unconscious for him ... [they are] akin to sentences, images, unconscious computational processes, and other people's conscious thoughts." He adds, however, that they differ from sentences, images, and so forth, in that they are "apparently in the patient."
12. According to Billon (2013, 307), "this does not mean that there is no phenomenology associated with thought insertion, only that it is, so to speak, an extrinsic, or a 'second-order phenomenology.'"
13. When arguing that there can be more than one type of consciousness, one phenomenal and the other not, Billon and Kriegel emphasize that what the two share is immediacy, or "immediate access." Elsewhere I have argued that although "immediate access" is relevant to these discussions, the lack of immediacy is one factor that can contribute to the experience of alien mental states (Lane 2012, 257–267; cf. Klein 2014, 101–103). Below I expand on this concern about the presumed significance of immediacy.
14. Billon (2013, 304) cites an actual instance of thought insertion that might be adduced to support the distinction they adopt here: according to one frequently cited description of thought insertion, it is like having "a screen" in the middle of one's skull on which pictures can be flashed, conveying thoughts that seem not to belong to self.

15. Elsewhere Billon (2013) emphasizes that, in order to develop an adequate explanation, we must allow for the possibility that two sets of features—the “phenomenal” and the “spatial”—are dissociable. There are many problems with this approach, not the least of which is the requirement that we accept the explanatory utility of distinguishing between intrinsic and extrinsic phenomenology. But setting these thorny conceptual issues aside, consider only the weight assigned by Billon to the spatial dimension and how this relates to the first-rank symptoms of schizophrenia. He says that “in me” is plausibly understood as “always where I am, that follows me around” (Billon 2013, 303–304). Arguably, giving emphasis to spatial phenomenology helps, in a rather straightforward way, to make sense of thought insertion. But ego-boundaries are permeable in two directions: some thoughts are inserted, while others “fly” to others, who can “catch” them (Mullins and Spence 2003, 294). In the latter case it is less clear how we should understand the spatial phenomenology, especially given Billon’s (2013, 296) expressed intent “to save the meaningfulness and the intelligibility of the patients.”

16. I should add that although I am in sympathy with interpretations of patient reports as reflecting actual patient phenomenology, there is need for caution when interpreting schizophrenic speech. Language disturbance—in particular incoherent or desultory content—is one of the primary behaviors by which diagnosis is made (Hyman and Cohen 2013, 1391). An example of this “loosening of associations” characteristic of much schizophrenic speech is: “I don’t think they care for me because two million camels ... 10 million taxis ... Father Christmas on the rebound.”

17. Here Billon (see 2013, 306, fn. 4) might insist that what is accessible or “reportable” bears no necessary connection to phenomenal consciousness.

18. Note that this role identified for receptive language areas in the left temporal lobe fits well with subjective reports, since the “near-universal experience” is of these voices “as being like an auditory percept” (Garrett and Silva 2003, 454). But it should also be noted that my claim in the text concerns empirical evidence about *receptive* areas, and that Billon (2013, 311) claims (a) the same does not hold for *productive* areas and (b) auditory cortex activation has only been observed “in a subgroup of patients.” Of course, getting straight about the facts of the matter in this vicinity will require much more work but, as regards (a), the subjective reality of auditory hallucinations correlates with activity in Broca’s area (Raij et al. 2009) and, even when subjective reality is not controlled for, activity in the right hemisphere homolog of Broca’s area correlates with auditory hallucinations (Sommer et al. 2008). As regards (b), one reason why auditory cortex activation might have been observed only in a subgroup of patients is a failure to adequately distinguish between the relative significance of resting state vis-à-vis stimulus-related activity (cf. Northoff 2013, 351; Kompus et al. 2011).

19. I expand on this point below.

20. As an analogy, consider commonplace awareness of change in a person or place, but change in features that typically go unnoticed. Such aspects of a person or place can be noticed when absent, even if we are unable to articulate or identify what is missing. Applying this idea to action, note that subliminally-processed stimuli can induce motor cortex activations enabling accurate responses to a target (Dehaene 2014, 129). Now consider the patient who wonders, “Is it really me walking?” My suggestion is that a degree of incompatibility between anticipation mechanisms and motor responses might be sufficient to cause patients to become aware of a disturbance to agency. What Poincaré dubbed the “subliminal self” can fail, thereby drawing attention to its absence, and causing mobilization of “enormous efforts in order to apply ... consciousness” toward rectifying motor responses that ordinarily do not require intervention by a conscious, agentive self (cf. Dehaene 2014, 86).

21. I expand upon this point in the concluding section.

22. All three are alien in the sense of seeming not to be controlled by self, but only for the passivity experiences of schizophrenics is it commonplace to attribute agency to some external source. For AHS this is less common. Utilization behavior refers to the compulsive grabbing and use of objects without regard to need or social situation, for example, reaching for and eating food even when not hungry or when the food belongs to someone else (Rizzolatti and Kalaska 2013, 883).

23. Mylopoulos is here endorsing a distinction between the phenomenology of bodily motion and of agency. As Horgan (2011, 64) characterizes the difference, the former involves a visual, kinesthetic, and so forth what-it’s-like; agency, however, places emphasis upon the “what-it’s-like of *self as source*.”

24. I do not agree with Mylopoulos that action consciousness occurs in ordinary cases; at best we *just know* that self performs a given action (Lane 2014, 64–65), and this seems to be what Block intends by suggesting the conceptual possibility of access- without phenomenal- consciousness. If this distinction does apply here, there is no agentive quality or phenomenal character that need concern us. In other words, what Mylopoulos suggests has gone missing, was not there in the first place. Neither do I endorse her dismissal of sensory attenuation as relevant to these discussions; there seems to be abundant experimental data to show that, pace Mylopoulos, sensory attenuation does occur in persons who have passivity experiences (e.g., Shergill et al. 2005 and Teufel et al. 2010), who have just been awakened from REM sleep (Blagrove et al. 2006), and so forth.

25. I here devote most of my attention to AHS because Mylopoulos takes that to be the example that most clearly supports her hypothesis. Concerning utilization behaviors, however, it is worth noting that recent experimental results show, at least for some of its manifestations, movements are more “rationally integrated into the agent’s psychology at the time” than Mylopoulos suggests. For example, Besnard et al. (2010) have adduced evidence to suggest that the frequency of utilization behaviors depends upon their relationship with the content of a task (involving a verbal-

ized script). If the task counts as an important part of the agent's psychology at the time, then many utilization behaviors are "rationally integrated."

26. Italics added.

27. Note that the conjectured association between HOTs and PFC activity is not universally endorsed (Gennaro 2012, 279–280).

28. For a succinct review of some relevant technical matters, see Rossi et al. 2009.

29. Because my concern here is only to argue that the Mylopoulos hypothesis is empirically tractable, I will not discuss methodological worries or the degree to which the data can plausibly be claimed to support the hypothesis that HOTs are misrepresenting actual intentions.

30. A possible neural substrate for intention-to-act has been suggested by Andersen and Cui 2009.

31. That the issues here are difficult is undeniable, and it may surprise some to know that it is not only philosophers who worry about for-me-ness. The cognitive neuroscientist, Revonsuo (2006, 32), for example, observes that "The study of consciousness is, first and foremost, the study of the world-for-me. ... That is the principal explanandum for the science of consciousness. Accordingly, the most fundamental concept in the study of consciousness should capture the essence of the world-for-me." By way of contrast, some philosophers whose main concern is the "phenomenal self" allow that "there are occasions—when we drift into reverie, or similar conditions—when we are not particularly aware of ourselves as selves at all" (Dainton 2008, 147).

32. Among my reasons for being disinclined to endorse Kriegel's view of subjective character is that he consigns it to "fringe" or "peripheral" consciousness (2009, 47–52). The result seems to be a compounding of our explanatory burden, adding not one but two contentious posits to the explanandum.

33. Italics added.

34. I am aware of the pitfalls of trying to operationalize before we have achieved sufficiently mature levels of conceptual sophistication (on the philosophical side) and methodological as well as technological sophistication (on the scientific side). And it is unlikely that anyone will soon pass a Turing test of consciousness science (cf., Revonsuo 2006, 300–303). But every intellectual choice involves a degree of risk: I am betting that we have already achieved levels of sophistication—on both sides—such that we are poised to eclipse "sketches of, or promises for, arguments" (Kriegel 2009, 312).

35. Kriegel's (2006, 156) view is distinctive in that he holds "the awareness we have of our conscious states is immediate simply because it *really* is unmediated."

36. I regard this as an instance wherein a person can "host" a conscious experience without that experience belonging to the person who is introspectively aware of it (cf. Lane 2012, 260). This is a somewhat different sense of "host" than that employed by Kriegel (2009, 8) when he asserts that "conscious experiences are not states that we may *host*..."

37. For a detailed analysis of this case and its possible relevance to "immunity to error through misidentification," see Langland-Hassan (forthcoming).

38. To distinguish among various degrees of consciousness, minimally conscious patients were included in the sample. Assessment included use of the Coma Recovery Scale (Revised).

39. If "self" so used still seems to intimate the mental, then substitute, "this organism, here, now."

References

- Andersen, R., and H. Cui. 2009. Intention, action planning, and decision-making in parietal-frontal circuits. *Neuron* 63:568–583.
- Avikainen, S., N. Forss, and R. Hari. 2002. Modulated activation of the human SI and SII cortices during observation of hand actions. *NeuroImage* 15:640–646.
- Baars, B. 2007. Attention and consciousness. In *Cognition, Brain, and Consciousness: Introduction to Cognitive Neuroscience*, ed. B. Baars and N. Gage. New York: Academic Press.
- Bayne, T. 2010. *The Unity of Consciousness*. New York: Oxford University Press.
- Berlin, H. 2011. The neural basis of the dynamic unconscious. *Neuro-psychoanalysis* 13:5–31.
- Besnard, J., P. Allain, G. Aubin, F. Osiurak, V. Chauvire, F. Etcharry-Bouyx, and D. Le Gall. 2010. Utilization behavior: Clinical and theoretical approaches. *Journal of the International Neuropsychological Society* 16:453–462.
- Bilek, E., A. Schafer, E. Ochs, C. Esslinger, M. Zangl, M. Plichta, U. Braun, et al. 2013. Application of high-frequency repetitive transcranial magnetic stimulation to the DLPFC alters human prefrontal-hippocampal functional interaction. *Journal of Neuroscience* 33:7050–7056.
- Billon, A. 2013. Does consciousness entail subjectivity? The puzzle of thought insertion. *Philosophical Psychology* 26:291–314.
- Blagrove, M., S.-J. Blakemore, and B. Thayer. 2006. The ability to self-tickle following rapid eye movement sleep. *Consciousness and Cognition* 16:285–294.

- Block, N. 2005. Two neural correlates of consciousness. *Trends in Cognitive Sciences* 9:46–52.
- Block, N. 2007. On a confusion about a function of consciousness. In *Consciousness, Function, and Representation: Collected Papers*, vol. 1. Cambridge, MA: MIT Press.
- Boag, S. 2010. Repression, suppression, and conscious awareness. *Psychoanalytic Psychology* 27:164–182.
- Bolognini, N., A. Rossetti, A. Maravita, and C. Miniussi. 2011. Seeing touch in the somatosensory cortex: A TMS study of the visual perception of touch. *Human Brain Mapping* 32:2104–2114.
- Bor, D. 2012. *The Ravenous Brain: How the New Science of Consciousness Explains Our Insatiable Search for Meaning*. New York: Basic Books.
- Bottini, G., E. Bisiach, R. Sterzi, and G. Vallar. 2002. Feeling touches in someone else's hand. *Neuroreport* 13:437–443.
- Bufalari, I., T. Aprile, A. Avenanti, F. Di Russo, and S. Aglioti. 2007. Empathy for pain and touch in the human somatosensory cortex. *Cerebral Cortex* 17:2553–2561.
- Craig, A. D. 2003. Interoception: The sense of the physiological condition of the body. *Current Opinion in Neurobiology* 13:500–505.
- Craig, A. D. 2009. How do you feel—now? The anterior insula and human awareness. *Nature Reviews: Neuroscience* 10:59–70.
- Craver, C. 2007. *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. New York: Oxford University Press.
- Dainton, B. 2008. *The Phenomenal Self*. New York: Oxford University Press.
- Dehaene, S. 2014. *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*. New York: Viking.
- Dienes, Z., and S. Hutton. 2013. Understanding hypnosis metacognitively: rTMS applied to left DLPFC increases hypnotic suggestibility. *Cortex* 49:386–392.
- Ehrsson, H., C. Spence, and R. Passingham. 2004. That's my hand! Activity in premotor cortex reflects feeling of ownership of a limb. *Science* 305:875–877.
- Feinberg, T., and A. Venneri. 2014. Somatoparaphrenia: Evolving theories and concepts. *Cortex*. doi:10.1016/j.cortex.2014.07.004.
- Gallagher, S. 2000. Self-reference and schizophrenia: A cognitive model of immunity to error through misidentification. In *Exploring the Self: Philosophical and Psychological Perspectives on Self-Experience*, ed. D. Zahavi. Philadelphia: John Benjamins.

- Gallivan, J. P., D. A. McLean, K. F. Valyear, C. E. Pettypiece, and J. C. Culham. 2011. Decoding action intentions from preparatory brain activity in human parieto-frontal networks. *Journal of Neuroscience* 31:9599–9610.
- Gardner, E. P., and K. Johnson. 2013. Touch. In *Principles of Neural Science*, 5th ed., ed. E. Kandel, J. Schwartz, T. Jessel, S. Siegelbaum, and A. Hudspeth. New York: McGraw Hill.
- Garrett, M., and R. Silva. 2003. Auditory hallucinations, source monitoring, and the belief that “voices” are real. *Schizophrenia Bulletin* 29:445–457.
- Gennaro, R. 2006. Between pure self-referentialism and the extrinsic HOT theory of consciousness. In *Self-Representational Approaches to Consciousness*, ed. U. Kriegel and K. Williford. Cambridge, MA: MIT Press.
- Gennaro, R. 2012. *The Consciousness Paradox: Consciousness, Concepts, and Higher-Order Thoughts*. Cambridge, MA: MIT Press.
- Gott, P., E. Hughes, and K. Whipple. 1984. Voluntary control of two lateralized conscious states: Validation by electrical and behavioral studies. *Neuropsychologia* 22:65–72.
- Gullivan, J., D. McLean, K. Valyear, C. Pettypiece, and J. Culham. 2011. Decoding action intentions from preparatory brain activity in human parieto-frontal networks. *Journal of Neuroscience* 31:9599–9610.
- Hardee, J., J. Thompson, and A. Puce. 2008. The left amygdala knows fear: Laterality in the amygdala response to fearful eyes. *Social Cognitive and Affective Neuroscience* 3:47–54.
- Hempel, C. 1965. *Aspects of Scientific Explanation*. New York: Free Press.
- Hirstein, W. 2012. *Mindmelding: Consciousness, Neuroscience, and the Mind's Privacy*. New York: Oxford University Press.
- Horgan, T. 2011. From agentive phenomenology to cognitive phenomenology: A guide for the perplexed. In *Cognitive Phenomenology*, ed. T. Bayne and M. Montague. New York: Oxford University Press.
- Huang, Z., R. Dai, Z. Yang, D. Liu, J. Hu, L. Gao, W. Tang, et al. 2014. The self and its resting state in consciousness: An investigation of the vegetative state. *Human Brain Mapping* 35:1997–2008.
- Hugdahl, K., E.-M. Loberg, L. Falkenberg, E. Johnsen, K. Kompus, R. Kroken, M. Nygard, R. Westerhausen, K. Alptekin, and M. Ozgoren. 2012. Auditory verbal hallucinations as aberrant lateralized speech perception: Evidence from dichotic listening. *Schizophrenia Research* 140:59–64.

- Hyman, S., and J. Cohen. 2013. Disorders of thought and volition: Schizophrenia. In *Principles of Neural Science*, 5th ed., ed. E. Kandel, J. Schwartz, T. Jessel, S. Siegelbaum, and A. Hudspeth. New York: McGraw Hill.
- Keyzers, C., J. Kaas, and V. Gazzola. 2010. Somatosensation in social perception. *Nature Reviews: Neuroscience* 11:417–428.
- Keyzers, C., B. Wicker, V. Gazzola, J.-L. Anton, L. Fogassi, and V. Gallese. 2004. A touching sight: SII/PV activation during the observation and experience of touch. *Neuron* 42:335–346.
- Kilner, J., and R. Lemon. 2013. What we currently know about mirror neurons. *Current Biology* 23:R1057–R1062.
- Klein, C. Forthcoming. What pain asymbolia really shows. *Mind*.
- Klein, S. 2014. *The Two Selves: Their Metaphysical Commitments and Functional Independence*. New York: Oxford University Press.
- Kompus, K., R. Westerhausen, and K. Hugdahl. 2011. The “paradoxical” engagement of the primary auditory cortex in patients with auditory verbal hallucinations: A meta-analysis of functional neuroimaging studies. *Neuropsychologia* 49: 3361–3369.
- Kriegel, U. 2006. The same-order monitoring theory of consciousness. In *Self-Representational Approaches to Consciousness*, ed. U. Kriegel and K. Williford. Cambridge, MA: MIT Press.
- Kriegel, U. 2009. *Subjective Consciousness: A Self-Representational Theory*. New York: Oxford University Press.
- Lane, T. 2012. Toward an explanatory framework for mental ownership. *Phenomenology and the Cognitive Sciences* 11:251–286.
- Lane, T. 2014. When actions feel alien—an explanatory model. In *Communicative Action*, ed. T. W. Hung. Singapore: Springer Science+Business Media.
- Lane, T., and C. Liang. 2008. Higher-order thought and the problem of radical confabulation. *Southern Journal of Philosophy* 46:69–98.
- Lane, T., and C. Liang. 2010. Mental ownership and higher-order thought. *Analysis* 70:496–501.
- Lane, T., and C. Liang. 2011. Self-consciousness and immunity. *Journal of Philosophy* 108:78–99.
- Langland-Hassan, P. Forthcoming. Introspective misidentification. *Philosophical Studies*.
- Lau, H., and D. Rosenthal. 2011. Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences* 15:365–373.

- Liang, C., and T. Lane. 2009. Higher-order thought and pathological self: The case of somatoparaphrenia. *Analysis* 69:661–668.
- Lipton, P. 1993. Contrastive explanation. In *Explanation*, ed. D. H. Ruben. New York: Oxford University Press.
- Macdonald, C. 2009. Introspection. In *The Oxford Handbook of Philosophy of Mind*, ed. S. Walter, A. Becherhmann, and B. McLaughlin, 741–766. Oxford: Oxford University Press.
- Marcel, A. 2003. The sense of agency. In *Agency and Self-Awareness: Issues in Philosophy and Psychology*, ed. J. Roessler and N. Eilan. New York: Clarendon Press.
- McNeill, D., L. Quaeghebeur, and S. Duncan. 2010. The man who lost his body. In *Handbook of Phenomenology and Cognitive Sciences*, ed. S. Gallagher and D. Schmickin. New York: Springer.
- Mullins, S., and S. Spence. 2003. Re-examining thought insertion: Semi-structured literature review and conceptual analysis. *British Journal of Psychiatry* 182:293–298.
- Northoff, G. 2013. *Unlocking the Brain*, vol. 2: *Consciousness*. New York: Oxford University Press.
- Northoff, G., and F. Bermpohl. 2004. Cortical midline structures and the self. *Trends in Cognitive Sciences* 8:102–107.
- Northoff, G., A. Heinzel, F. Bermpohl, H. Dobrowolny, and J. Panksepp. 2006. Self-referential processing in our brain: A meta-analysis of imaging studies on the self. *NeuroImage* 28:440–457.
- Northoff, G., P. Qin, and T. Nakao. 2010. Rest-stimulus interaction in the brain: A review. *Trends in Neurosciences* 33:277–284.
- O’Brien, L. 2007. *Self-Knowing Agents*. New York: Oxford University Press.
- Prinz, J. 2012. *The Conscious Brain: How Attention Engenders Experience*. New York: Oxford University Press.
- Qin, P., and G. Northoff. 2011. How is our self related to midline regions and the default mode network? *NeuroImage* 57:1221–1233.
- Raij, T., M. Valkonen-Korhonen, M. Holi, S. Therman, J. Lehtonen, and R. Hari. 2009. Reality of auditory hallucinations. *Brain* 132 (Pt. 11): 2994–3001.
- Revonsuo, A. 2006. *Inner Presence: Consciousness as a Biological Phenomenon*. Cambridge, MA: MIT Press.
- Rilke, R. M. 1954. *Letters to a Young Poet*, rev. ed. New York: W. W. Norton.
- Rizzolatti, G., and J. Kalaska. 2013. Voluntary movement: The parietal and premotor cortex. In *Principles of Neural Science*, 5th ed., ed. E. Kandel, J. Schwartz, T. Jessel, S. Siegelbaum, and A. Hudspeth. New York: McGraw Hill.

- Romano, D., M. Gandola, G. Bottini, and A. Maravita. 2014. Arousal responses to noxious stimuli in somatoparaphrenia and anosognosia: Clues to body awareness. *Brain* 137:1213–1223.
- Rosenthal, D. M. 2002. Explaining consciousness. In *Philosophy of Mind: Classical and Contemporary Readings*, ed. D. Chalmers. New York: Oxford University Press.
- Rosenthal, D. M. 2005. *Consciousness and Mind*. New York: Oxford University Press.
- Rosenthal, D. M. 2010. Consciousness, the self, and bodily location. *Analysis* 70: 270–276.
- Rossi, S., M. Hallett, P. Rossini, and A. Pascual-Leone. 2009. Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. *Clinical Neurophysiology* 120:2008–2039.
- Rounis, E., B. Maniscalco, J. Rothwell, R. Passingham, and H. Lau. 2010. Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. *Cognitive Neuroscience* 1:165–175.
- Sass, L., and J. Parnas. 2003. Schizophrenia, consciousness, and the self. *Schizophrenia Bulletin* 29 (3): 427–444.
- Schaefer, M., B. Xu, H. Flor, and L. Cohen. 2009. Effects of different viewing perspectives on somatosensory activations during observation of touch. *Human Brain Mapping* 30:2722–2730.
- Schaefer, M., H.-J. Heinze, and M. Rotte. 2012. Embodied empathy for tactile events: Interindividual differences and vicarious somatosensory responses during touch observation. *NeuroImage* 60:952–957.
- Schneider, F., F. Bermpohl, A. Heinzel, M. Rotte, W. Tempelmann, C. Wiebking, H. Dobrowolny, H. Heinze, and G. Northoff. 2008. The resting state and our self: Self-relatedness modulates resting state neural activity in cortical midline regions. *Neuroscience* 157:120–131.
- Shergill, S., G. Samson, P. Bays, C. Frith, and D. Wolpert. 2005. Evidence for sensory prediction deficits in schizophrenia. *American Journal of Psychiatry* 162:2384–2386.
- Sierra, M. 2009. *Depersonalization: A New Look at a Neglected Syndrome*. New York: Cambridge University Press.
- Sommer, I., K. Dierkeren, J. Blom, A. Willems, L. Kushen, K. Slotema, M. Boks, et al. 2008. Auditory verbal hallucinations predominantly activate the right inferior frontal area. *Brain* 131 (Pt. 12): 3169–3177.
- Teufel, C., A. Kingdon, J. Ingram, D. Wolpert, and P. Fletcher. 2010. Deficits in sensory prediction are related to delusional ideation in healthy individuals. *Neuropsychologia* 48:4169–4172.

- Tsuchiya, N., F. Moradi, C. Felsen, M. Yamazaki, and R. Adolphs. 2009. Intact rapid detection of fearful faces in the absence of the amygdala. *Nature Neuroscience* 12:1224–1225.
- Valenzuela-Moguillansky, C. 2013. Pain and body awareness: An exploration of the bodily experience of persons suffering from fibromyalgia. *Constructivist Foundations* 8:339–350.
- Wegner, D. 2002. *The Illusion of Conscious Will*. Cambridge, MA: MIT Press.
- Yang, E., D. Zald, and R. Blake. 2007. Fearful expressions gain preferential access to awareness during continuous flash suppression. *Emotion* 7:882–886.
- Zahn, R., J. Talazko, and D. Ebert. 2008. Loss of the sense of self-ownership for perceptions of objects in a case of right inferior temporal, parieto-occipital, and precentral hypometabolism. *Psychopathology* 41:397–402.