

For citation purposes, please see published version *Philosophical Studies* (2015), which has some important additions and clarifications.

## **Introspective Misidentification**

Peter Langland-Hassan

University of Cincinnati

*Abstract:* It is widely held that introspection-based self-ascriptions of mental states are “immune to error through misidentification” (IEM), relative to the first person pronoun. Many have taken such errors to be logically impossible, arguing that the immunity holds as an “absolute” necessity. Here I discuss an actual case of craniopagus twins—twins conjoined at the head and brain—as a means to arguing that such errors are logically possible and, for all we know, nomologically possible. An important feature of the example is that it is one where a person may be said to be introspectively aware of a mental state that occurs outside of her own mind. Implications are discussed for views of the relation between introspection and mental state ownership, and between introspection and epistemic criteria for the “mark of the mental.”

### **1. Introduction**

It is not the normal thing to see a canary and wonder who is seeing it, or to feel a pang of hunger and wonder whose it is. But are there situations where such questions reasonably arise?

Wittgenstein thought not, writing in *The Blue Book* that, when someone complains of a headache, “To ask ‘are you sure it is *you* who have pains?’ would be nonsensical” (1969, p. 67). Sydney Shoemaker (1968) distilled Wittgenstein’s observation into the now familiar claim that certain judgments are “immune to error through misidentification” (IEM). To say that a judgment of the form ‘*x* is *F*’ is IEM, in Shoemaker’s sense, is to say that one cannot go wrong in that judgment simply because one has misidentified *x* as the thing that is in fact *F* (1968, p. 557). The range of judgments that are IEM is debated, though introspection-based self-ascriptions of mental states are widely agreed to be among them. It is commonly held that if introspection grounds a judgment of the form, “I am in mental state *m*,” then that judgment cannot be false for the sole reason that the wrong subject was identified as the owner of the

mental state. It is consistent with this idea that one might err in an introspective judgment about what *kind* of mental state one is in. The judgment, “I am in pain” might be mistaken because one is really feeling an itch. But there is no getting *right* the kind of state one is introspecting, while simultaneously getting *wrong* that it belongs to oneself. At least, so goes a widely shared intuition.

To give this idea a name, let *Introspective Immunity* be the following thesis:

*Introspective Immunity*: self-ascriptions of mental states, where the first person pronouns (‘I’ or ‘me’) are used, and where introspection provides the grounds, are immune to error through misidentification.

The thesis itself says nothing about how introspection works. Introspection can be neutrally characterized as the usual way (or ways) in which people gain knowledge of their own current mental states. If one judges oneself to be in a particular mental state, and if introspection provides the grounds for the judgment, then, according to Introspective Immunity, one cannot err in the judgment *simply* because one has misidentified the owner of the state as being oneself.

But what is the modal force of ‘cannot’ in the preceding sentence? Are such errors flat out impossible, the principle holding as a kind of logical necessity? Or is their “impossibility” contingent upon things being as they normally are—for instance, upon one’s having ordinary human anatomy? Shoemaker (1968) defends the stronger view, with respect to at least some mental predicates. He explains that “in being [introspectively] aware that one feels pain one is, *tautologically*, aware, not simply that the attribute *feel(s) pain* is instantiated, but that it is instantiated *in oneself*” (1968, p. 564; first emphasis mine). In his term, the judgments in question are “absolutely” immune to error through misidentification (p. 566). Many others agree that certain introspective judgments are absolutely IEM (Campbell, 1999a, p. 93; Coliva, 2002; Pryor, 1999, p. 297; Smith, 2006, p. 275). While some are explicit in equating Shoemaker’s “absolute” impossibility with logical impossibility (Coliva, 2002, p. 28), others fail to specify the notion of possibility in play (Pryor, 1999, p. 297). But it seems they must have

both logical and metaphysical possibility in mind. At least, Introspective Immunity (and IEM) has never, to my knowledge, been defended as a necessary *a posteriori* principle—as something that at one time seemed false or contingent but which we have discovered, *a posteriori*, to be necessarily true. Nor do there seem to be any defenders of Introspective Immunity who hold it to be logically, but not metaphysically necessary (or vice versa). Rather, the idea seems to be that we know *now* that Introspective Immunity cannot be violated—not here or in any logically or metaphysically possible world. This is why Wittgenstein found it simply “nonsensical” to question whether a pain one feels is one’s own. And, indeed, it is very difficult to imagine a situation that would clearly count as one where Introspective Immunity is violated.<sup>1</sup>

Nevertheless, my project here is to describe a counterexample to Introspective Immunity, showing that contingent features of our own anatomies may be all that prevent mistakes of the relevant kind. What I aim to establish is that violations of Introspective Immunity are logically possible and, further, that there is no reason to think that such violations are not metaphysically and nomologically possible as well. The argument appeals to an actual case of craniopagus twins—two sisters—who are conjoined at the brain. Their situation (what we know of it) is used to make vivid the logical possibility of a scenario where Introspective Immunity is violated, by showing that it might be what is occurring in an actual case. In showing that an actual situation might be one where Introspective Immunity is violated, I aim also to show that there is no obvious nomological or metaphysical barrier to violations of Introspective Immunity. That said, I stop short of arguing that their situation *is* one where Introspective Immunity is violated. We do not, at present, know enough about the twin’s anatomy, or about the neural mechanisms of underlying introspection, to establish that stronger conclusion.

“Counterexample games” aside, the real motivation of this paper is to gain some clarity on what hangs on the truth or falsity of Introspective Immunity. Joel Smith speaks for many in proposing that the (supposed) impossibility of introspective errors of misidentification “is a feature of self-ascriptions that marks their centrality to our conception of ourselves as self-

---

<sup>1</sup> There are proposed counterexamples to Introspective Immunity in the existing literature, including Lane & Liang (2011), Rosenthal (2012), and Hogan & Martin (2001). These will be discussed below.

conscious subjects” (2006, p. 274). Intuitively, the impossibility of such errors arises out of some fundamental and distinctive feature of self-consciousness, introspection, and mentality generally. Yet the exact nature of the connection is far from obvious.<sup>2</sup> Challenging Introspective Immunity requires treating a number of perplexing issues, including the relation between mental state integration and personhood, between a owning a brain state and owning a mental state, between introspection and “the mark of the mental,” and between phenomenology and moral appraisal. I will treat each of these in the course of this essay.

Finally, it is worth noting that much of the IEM literature has focused on the question of what it is that guarantees our absolute (or “logical”) immunity to errors of misidentification (when we have it), a common answer being that introspective self-ascriptions do not involve any act of identification of the kind that might possibly go awry.<sup>3</sup> Often these debates take the absolute necessity of Introspective Immunity as a datum that any account of its basis must get right.<sup>4</sup> However, if this paper’s main thesis is correct, explanations of IEM (to whatever extent it holds) should not be counted defective simply because they allow for possible violations of Introspective Immunity. It is a contingent feature of our world that such errors are exceedingly rare.

## 2. Telepathy and Thought Insertion

Before defending a counterexample to Introspective Immunity (in Section 4), it will help to make some clarifications about the target thesis, and to say why other proposed counterexamples fall short.

First, a clarification: philosophers sometimes distinguish between cases of a person seeing an *x* and cases of a person having a visual experience *as of* an *x*. Cases of seeing require a certain kind of success. If a person *sees* a lion, then there really must be a lion with which the person is in appropriate causal-perceptual contact. Visual experiences *as of* lions, on the other

---

<sup>2</sup> Though see Wright (2012) for an argument that there is no special connection between judgments that are IEM and self-consciousness.

<sup>3</sup> See Smith (2006, p. 276) for a clear articulation of this approach to explaining Introspective Immunity.

<sup>4</sup> For instance, Smith (2006, p. 279) dismisses Wright’s(1998) explanation of why introspective self-ascriptions of mental states are IEM on the grounds that, if it were correct, some judgments of the form ‘I am in pain’ (made in *very* unusual circumstances) would not be IEM.

hand, do not require concurrent perceptual contact with lions. So, hallucinations of lions qualify as visual experiences as of lions, and not as cases of seeing a lion. Corresponding distinctions are available across the sense modalities. Let us call the more demanding, success-connoting kinds of mental states (such as seeings and hearings) “factive” states, and the less demanding, hallucination-tolerating kinds of states “non-factive” mental states.

While not everyone will like these distinctions, I want to work with them in order to face what I take to be the most plausible, and widely endorsed, version of Introspective Immunity. Consider, then, the following (imaginary) scenario involving telepathy.<sup>5</sup> Suppose that Kruno is a telepathic detective. His special talent is to identify kidnapers by gaining telepathic access to the minds of kidnapped children. Kruno closes his eyes and, via telepathy, has a visual experience that mirrors the one that a recently kidnapped boy is having...he sees the face of the kidnapper! Or does Kruno merely have a visual experience *as of* the face of kidnapper? It depends on whether the telepathic connection is deemed sufficient for factive seeing. Let us assume it is not.<sup>6</sup> Suppose now that it is late at night and Kruno has been drinking; unbeknownst to him, his telepathic powers suddenly engage. He seems to see a suspicious man’s face in the shadows at the back of the bar. Not realizing he is using telepathy, he judges, through introspection, “I see a suspicious man.” In fact it is a kidnapped child who sees the man, not the detective. Kruno merely has a visual experience *as of* a suspicious man. So his judgment—“I see a suspicious man”—is false. And one might think that, in this unusual case, the introspection-based judgment is false *for the sole reason* that he has misidentified the person who sees a suspicious man as himself. For the child *does* see the man in question. This may seem to challenge Introspective Immunity.

It is here that the defender of IEM can reply that the thesis only holds as a *logical* necessity with respect to non-factive mental states. It is only when the detective judges, through introspection, “I am having a visual experience as of a suspicious man,” (or, more likely, “I *seem to see* a suspicious man”), and where the judgment is false for the sole reason that he

---

<sup>5</sup> Imagined cases of telepathy are sometimes offered as counterexamples to the absolute necessity of Introspective Immunity (Hogan & Martin, 2001).

<sup>6</sup> If we assume it is sufficient, then there is no appearance of a challenge to IEM—a judgment of the form “I see the kidnapper,” made by Kruno, will be true. Here I want to consider (and dismiss) a possible counterexample that could arise if telepathy is not sufficient for factive seeing.

has misidentified the person undergoing the experience as himself, that we would have a counterexample to Introspective Immunity. In this case, it seems plausible that Kruno is indeed having a visual experience as of a suspicious man. So the more cautious judgment—"I am having a visual experience as of a suspicious man"—is true. Kruno can then arguably be seen as having made an error of *mispredication* when he judged: "I see a suspicious man." Like the person who mistakes one of his itches for a pain, the detective mistakes his visual experience as of a suspicious man for a case of seeing a suspicious man. No threat to IEM there; for the judgment is not false *for the sole reason* that he has misidentified the subject of an otherwise proper predication.

My point in contemplating this case is to make clear that my target is the more cautious (and more attractive) version of Introspective Immunity, relativized to non-factive mental states. These are the kinds of judgments that people are most likely to find *absolutely* immune to error through misidentification. So my goal will be to describe a plausible case where a person uses introspection to judge, e.g., "I am having a visual experience as of an *x*," and is mistaken for the sole reason that he has misidentified the subject of the experience.

A second point of clarification—of emphasis, really—is that the version of IEM under discussion pertains only to *self*-ascriptions of mental states, where a first person pronoun is used, and where introspection provides the grounds. This is worth emphasizing because Shoemaker's IEM thesis is sometimes characterized (incorrectly, in my view) as holding that *all* ascriptions of mental states (whether to self or other), where introspection provides the grounds, are immune to errors of misidentification (see, e.g., Campbell, 1999b, p. 610; Lane & Liang, 2011, p. 87). This understanding of IEM makes no essential appeal to the first person pronoun. All of the weight is put on the fact that the mental state in question is detected through introspection. The state's being so detected is considered sufficient for its not being ascribed to the wrong person. Let us call this the *subject-term-neutral* version of IEM, as on this version of IEM there is no stipulation that the ascription must incorporate the first person pronoun. According to the subject-term-neutral version of IEM, a counterexample occurs if someone accurately introspectively detects one of her own mental states as being the *kind* of

state it is—a sensation of pain, say—yet wrongly ascribes it to someone else. Cases of thought insertion in schizophrenia—where a patient will claim that another person’s thoughts have been “inserted” into her own mind—are sometimes offered as actual instances (Campbell, 1999b, p. 610). More recently, Lane & Liang describe a patient F.B. (from a case study by Bottini *et al.* (2002)) who suffered from post-stroke somatoparaphrenia (denying ownership of her hand). Her reports seem to fit the same pattern, as she ascribes the sensations she feels in her own hand to her niece.

Whatever we make of these cases, they are not counterexamples to any version of IEM that is relativized to self-ascriptions using the first person pronouns.<sup>7</sup> For the misidentifications occur when the introspectively detected state is ascribed to someone *other than* the person making the relevant judgment and ascription. My interest is in Introspective Immunity and its connection to the first person pronoun. And it is easy to see that the subject-term-neutral version of IEM makes a very different sort of claim than Introspective Immunity. It claims, in effect, that as a matter of logical necessity, a person will never introspectively discern one of her own mental states and then ascribe it to someone else. Well, perhaps no one could possibly do that (notwithstanding reports of thought insertion). But it is not clear why sincere utterances would have that kind of limit.

Understanding how Introspective Immunity could be violated presents a puzzle of a different kind. Recall that Introspective Immunity is not violated simply by one’s mistaking the kind of mental state one is in. A counterexample requires a person to *correctly* introspectively discern a particular mental state (as being the kind of mental state it in fact is), yet fall into error *simply* because he has—with the first person pronoun—self ascribed it. Satisfying both conditions requires a confluence of worldly affairs that is difficult to imagine. Whereas

---

<sup>7</sup> Lane & Liang offer another example, involving the “Body Swap” illusion, that does arguably involve a self-ascription of the right form to challenge Introspective Immunity. They hold that when subjects experience the illusion of shaking hands with themselves, “their experiences involve misrepresentation of action awareness” (2011, p. 92). Although it was really the experimenter who was shaking their hand, “the subjects misrepresented themselves as the agent of the action” (p. 92). They take the case to violate IEM because the subjects “represent themselves as agents when plainly they are not” (*ibid.*). However, in one sense, the test subjects really are agents. For they do in fact shake hands with the experimenter. So they are right to the extent they judge, through introspection, that they are shaking hands. Where they go wrong is in judging that they are shaking *their own* hands. However, it is only an unusual sort of *perceptual* experience—seeing their own bodies on a head-mounted screen—that leads to the mistaken part of the judgment. So in this case the judgment as a whole is not grounded in introspection in the way it would need to be to challenge Introspective Immunity.

violations of the subject-term-neutral version of IEM require only that someone sincerely denies ownership of an introspected mental state, violating Introspective Immunity requires a person *in fact* to be introspectively aware of a mental state that does not occur within her own mind. Only once it is clear how *that* can happen can we see how a person might go on to mistakenly self-ascribe such a state.

Thus, most of what follows (in Section 4) will aim at describing a case where it is plausible that a person is indeed introspectively aware of a (non-factive) mental state that occurs outside of her own mind. That is the tough nut to crack. Once that piece is in place, it is by comparison easy to see how a person might mistakenly ascribe such a state to herself.

### **3. Dissociative Identity Disorder**

David Rosenthal (2012) has recently offered a counterexample with the right form to challenge Introspective Immunity. His case involves Dissociative Identity Disorder (formerly known as Multiple Personality Disorder). In Dissociative Identity Disorder (DID), one or more personalities—called “alters”—inhabit a single body. Part of what distinguishes one alter from another is the distinctive set of memories to which it has access. But, according to Rosenthal, “some alters have access not only to their own memories and experiences, but to those of other alters as well,” creating “room for confusion about which state belongs to which alter” (2012, p. 43). Extending that idea to the case of sensations, Rosenthal suggests that we should be able to imagine a situation where one alter has introspective access to the pain of another, without also sharing in ownership of the pain. “When one alter but not another actually has a pain,” he writes, “it may happen that the other alter is sometimes wrong in being aware of itself as being in pain because, and only because, the alter that is aware of itself as being in pain misidentifies the other alter as itself” (2012, p. 43).

The example is suggestive, and on target concerning the requirements for violating Introspective Immunity. However, questions can be raised about its coherence. First, it requires that DID is *not* simply a disorder where a single person has multiple personalities, or perceives himself to have multiple identities, but rather a disorder where multiple *people* share a single body. For an error of the right kind can only occur when one person is introspectively

aware of a mental state that belongs (exclusively) to someone else. This view of DID entails that any treatment reducing the number of alters inhabiting a body will end at least one person's existence. That is very far from the psychiatric community's perspective on the matter. In their "Guidelines for Treating Dissociative Identity Disorder in Adults, Third Revision" the International Society for the Study of Trauma and Dissociation emphasize that "although the DID patient has the subjective experience of having separate identities, it is important for clinicians to keep in mind that the patient is not a collection of separate people sharing the same body," but rather "a single person who experiences him or herself as having separate alternate identities" (2011, pp. 132, 120). Therefore, "treatment should move the patient toward better integrated functioning [among the personalities] whenever possible" (p. 132). At a minimum, Rosenthal is in the awkward position of arguing that mainstream psychiatry is operating under fundamental misconceptions about the nature of DID.

Second, given that the two alters share the same brain, we have to ask what it is that might make a particular pain realized in that brain belong to *only one* of the alters. For if they *share* ownership of the pain, neither can err in self-ascribing that pain. Rosenthal answers that "a state belongs to one alter rather than another if it is suitably integrated with a range of other states characteristic of a particular alter, and not other alters" (p. 43). What does the "suitable integration" of a pain with states characteristic of a particular alter consist in, if not the capacity of an alter to be introspectively aware of it? Rosenthal does not offer further details. Without those details, it is hard to see why such a pain would not be suitably integrated within the psychologies of both alters.

But the problem goes deeper still: appealing exclusively to "suitable" integration as a criterion for ownership gives rise to a difficult dilemma. If we set the bar for suitable integration relatively low, then both alters will inevitably own the mental state; if we set the bar higher, we risk the absurd conclusion that many of our own poorly integrated mental states (e.g., dim memories, suppressed beliefs and desires, subliminal perceptions, etc.) are not, in fact, our own mental states. To illustrate this problem, let Alter 1 be an alter who owns a certain pain, and Alter 2 be an alter who is merely introspectively aware of it. It seems likely that the pain will be better integrated within Alter 2's psychology than will be many of Alter 1's

own poorly integrated mental states (such as his repressed beliefs, subconscious perceptions, and so on). If integration is all there is to ownership, and if Alter 1 owns his poorly integrated states, then Alter 2 must own the (poorly integrated) pain as well. On the other hand, if integration is not all there is to ownership, then we have no reason to think that DID is a situation where only one alter might own the mental state in question.

What this dilemma suggests is that integration is *not* all there is to ownership. My inability to report on one of my own mental states, and its relative lack of impact on my other mental states, does not entail that it is not my mental state after all. Knowing the facts about degree of integration does not seem to settle the issue with respect to mental state ownership. With that in mind, I turn to a case that I think would constitute a genuine counterexample to Introspective Immunity.

#### **4. Craniopagus Twins**

Krista and Tatiana Hogan are (actual) craniopagus conjoined twins, attached at the head. Unlike most craniopagus twins, they are also attached *at the brain*, sharing what one researcher calls a “thalamic bridge”—a neural pathway from one girl’s thalamus to the other’s. This is the full extent of their physical connection; each girl has an otherwise distinct body of her own.

By all appearances, their neural “bridge” allows for some remarkable abilities, many of which Susan Dominus documents in a 2011 *New York Times Magazine* article on the twins. With her eyes covered by her mother’s hand, Krista seems able to report on what kind of object (a toy pony) has been raised before Tatiana’s eyes; facing the opposite direction, Tatiana knows when (and where) Krista is being tickled.<sup>8</sup> Each twin seems to know what the other is seeing or feeling, and perhaps even thinking, in a way others cannot. Each seems to know these things through introspection.

As Dominus notes, no controlled studies have confirmed these capacities. This is partly due to the young age of the twins (they were four at the time of writing) and to the fact that they are not guinea pigs, after all. Yet brain imaging clearly shows a connection at or near the thalamus of each girl, and such a connection fits with the many unusual anecdotal reports of their behavior.

---

<sup>8</sup> A video depicting some of these capacities, together with Dominus’s (2011) article, can be retrieved at: <http://www.nytimes.com/2011/05/29/magazine/could-conjoined-twins-share-a-mind.html>

For the thalamus is known to be a multi-modal “relay center” near the brain stem, running between early sensory inputs and higher processing areas. It is reasonable to think that sensory signals initially registered in the brain of one twin might travel, via the thalamic link, to the brain of the other. Note that their brains are otherwise relatively normal. Each twin has two cerebral hemispheres connected by a corpus callosum. Their neural overlap appears to begin and end at their thalami.

Going forward, I will assume it is a live hypothesis that the twins do, in fact, have a kind of immediate access to experiences that occur in each other’s minds—a kind of access that others do not have. Note that, if true, this would not in itself pose a challenge to Introspective Immunity. For one twin might be “immediately” aware of the other’s experiences only in the sense that she sometimes tokens, in her own mind, “copies” of the experiences enjoyed by her twin. Consider the case where Krista correctly reports, with her eyes covered, that Tatiana is seeing a toy pony. It might be that information concerning the visual experience  $v$  (as of a toy pony) had by Tatiana is sent, via the thalamic link, to Krista’s brain, where  $c$ , a corresponding copy of  $v$ , is realized in Krista’s visual cortices. Call this the Two Tokens (TT) scenario. If Krista is able to report that Tatiana sees a doll by being introspectively aware of  $c$ , we do not have a clear challenge to Introspective Immunity. For (arguably) Krista is not introspectively aware of a mental state that in fact occurs outside of her own mind (such that she could misidentify its owner as herself). Rather, she is introspectively aware of one of her own mental states and, in virtue of its “matching” a state of her sister, is able to say what her sister is seeing.

The situation so envisioned is similar to the case of Kruno, the telepathic detective, considered above. In imagined cases of telepathy generally, the natural (though not inevitable) assumption is that the telepath knows what someone else is seeing or thinking by tokening, in her own mind and brain, a copy of the same type of experience had by that person. What we need to challenge Introspective Immunity is a case where a person is introspectively aware of a *token* mental state that does not occur within her own mind; only then will it be clear how a person could be introspectively aware of a non-factive mental state yet err in self-ascribing it.

Returning to the twins, a first step in that direction is to suppose that there is only one token experience  $v$  (as of a toy pony) between them, and that they are both introspectively aware

of it. Call this the One Token (OT) scenario. In OT,  $v$  is realized by patterns of activity within Tatiana's visual cortex. However, the thalamic link allows for direct causal contact between  $v$  and introspective processes realized in Krista's brain. In this way, Krista is able to be introspectively aware of  $v$ , and to thereby judge that Tatiana sees a doll (other sensorimotor cues make it evident to Krista that her own eyes are closed).

To be clear, the claim is not that we have good reason to believe that OT is *in fact* what occurs when Krista, with her eyes closed, reports on what Tatiana is seeing. Rather, the proposal is that it is a logical possibility (and, for all we know, a nomological possibility as well). Further empirical investigation might even show it to be the best supported account of their situation. Moreover, even if it does not—if, say, neuroimaging, combined with what we know about thalami, supports TT instead—we still have a clear picture of how OT could correctly characterize some possible set of neutrally conjoined twins, not too different from Krista and Tatiana. For keep in mind that, in order to challenge Introspective Immunity's claim to being a logical necessity, we simply need a plausible description of a logical possibility (not an actuality) that challenges it. The value of appealing to the (quite actual) Hogan twins is that it allows us to form a vivid picture of a logical possibility, by showing us that it is close to, and perhaps even identical to, an actuality.

The OT scenario as so far described offers a first ingredient to challenging Introspective immunity. It is a case where one person (Krista) is introspectively aware of a token mental state that does not occur within her own *brain*.<sup>9</sup> My argument will be that this also may be a scenario where  $v$  does not occur within Krista's *mind*. If that is right, then, in a moment of distraction, or given confusing perceptual cues, Krista might become introspectively aware of  $v$  and erroneously self-ascribe it. That is, she might judge, through introspection, "I am having a visual experience as of a doll," and be wrong for the sole reason that she has misidentified the owner of the state (her sister) as herself.

That is the shape of the counterexample I have in mind. Most of what follows consists in its defense. But let me again be clear that the error of misidentification is envisioned as a logically possible error. Nothing in the published reports concerning the Hogan twins suggests that they

---

<sup>9</sup> It is assumed, at this point, that there are two brains between the twins, not one unusual brain. That point will be argued for below (end of Section 5.1).

have *in fact* made any errors of misidentification. The point of the example is that it offers a way of seeing how such an error *could* occur.

#### 4.1 *Distinct Existences views of introspection*

Several issues must be addressed to make it plausible that OT could indeed be a situation where Krista is aware of a mental state that occurs outside of her own mind. First, OT presumes a distinction can be drawn between a first order visual state (*v*) and the state or process by which one ordinarily becomes introspectively aware of such a state—that the two are *distinct existences*. Shoemaker famously (1994) denies the coherence of such views. I will not have space to address his reasons for doing so here.<sup>10</sup> Instead I will assume that a “distinct existences” view of introspection *might* be true—that at least one such view is not logically impossible. There are many such theories on the market, including Higher Order Thought theories (Rosenthal, 2005), Higher Order Perception/Monitoring views (Armstrong, 1968; Lycan, 1996; Nichols & Stich, 2003), global broadcasting-plus-mindreading-module views (Carruthers, 2011), and “outside-in” inferentialist views (Byrne, 2005; Dretske, 1999). The list is not exhaustive. Despite considerable

---

<sup>10</sup> I will, however, say something here about his “vicious regress” argument against distinct existences views of introspection. Distinct existences views are naturally paired with the idea that introspection must involve the identification of mental states as belonging to oneself, just as judgments about one’s body formed through perception involve identifying perceived things (e.g. a bleeding knee) as being one’s own. However, Shoemaker insists that if this is how introspection works, it gives rise to “a vicious infinite regress” with respect to our capacity for self-knowledge: “Identifying something as oneself would have to involve...finding something to be true of it that one independently knows to be true of oneself...This self-knowledge might in some cases be grounded on some *other* identification, but the supposition that every item of self-knowledge rests on an identification leads to a vicious infinite regress” (p. 561).

If one is convinced that that IEM is a logical necessity, the vicious regress problem may arise; if one doubts the absolute necessity of IEM, then the regress problem poses no additional challenge to distinct existences views (I think that Shoemaker means to grant this much, though I am not sure others will agree). For note that ordinary perceptual knowledge does not fall prey to a vicious infinite regress, even if perceptually identifying some thing *x* as *y* (e.g., that man as Bill Clinton) has to involve finding something true of it that one already knew to be true of *y* (e.g., that Clinton has certain facial features). There is no vicious regress here because we can initially pick up the reference of ‘Bill Clinton’, and learn some of his characteristics, through someone else’s demonstration. Of course, if I initially learned that Clinton has certain facial features from my friend’s pointing to him and saying, “That is Clinton,” then my subsequent judgments about Clinton will not be immune to errors of misidentification. For something could have gone wrong in the initial identification (e.g. I may have mistook who my friend was pointing at), infecting my judgments thereafter. But, of course, knowledge of Bill Clinton does not require immunity to such errors of misidentification. By the same token, so long as we accept that judgments about one’s own mental states are not absolutely immune to errors of misidentification, there is no reason that self-knowledge cannot consist in a succession of judgments involving identification, together with some (fallible) acts of introspective demonstration, or application of a theory of mind (to oneself).

See Byrne (2005), Rosenthal (2012), Finkelstein (1999), and Lormand (2000), for responses to Shoemaker’s other arguments (involving, e.g., “self-blindness”) against distinct existences views.

differences among them, they all agree that introspection involves a state or process that is ontologically distinct from the states of which it makes one aware. These views can all countenance the possibility that an introspective process might occur within Krista's brain that allows her to be introspectively aware of  $v$ , where  $v$  is a first order visual state realized in Tatiana's brain.

What if one rejects distinct existences views of introspection? First, in rejecting outright the possibility of a distinct existences view of introspection, one cannot then object to such views on the specific grounds that they would entail the contingency of Introspective Immunity—for that is what is in question. Some other reason for rejecting distinct existences views would be needed. Second, and more importantly, our question is whether Introspective Immunity is a logical necessity. The relevant question with respect to introspection is thus whether a distinct existences view is logically possible. Only if no such view is logically possible can we preserve the absolute necessity of Introspective Immunity by placing a (logical) ban on distinct existences views. And, whether one finds it plausible or not, is there any more *comprehensible* account of introspection than Armstrong's (1968) classic "inner sense" view? If such a view is logically possible, then the OT scenario (as so far described) is as well.

Granted, those who are convinced (on independent grounds) that all distinct existences views are logically impossible may not be moved. Nevertheless, from that perspective, one can still take what follows as an argument that preserving Introspective Immunity *requires* a prior acceptance of the logical impossibility of distinct existences views of introspection. As Introspective Immunity is rarely (if ever) defended on the *prior* grounds that distinct existences views are logically impossible, this will be an interesting conclusion enough.

#### 4.2 *Sharing the experience*

Even granting the OT scenario's possibility, together with a distinct existences view of introspection, we still do not have an obvious challenge to Introspective Immunity. For OT might necessarily be a situation where the twins *share* ownership of  $v$ , in the sense that  $v$  occurs in both of their minds (again, it is *not* being assumed that the only states in one's mind are those realized in one's brain). If they share ownership of  $v$ , then it would not be possible for either twin to

erroneously self-ascribe  $v$  and (in so doing) violate Introspective Immunity. So, something must be said to make it plausible that they *might not* share ownership of  $v$ .

Now, as  $v$  occurs only in Tatiana's brain, and as the overlap between the brains is minimal, we have at least *some* reason to think that it only occurs in Tatiana's mind. After all, a natural presumption is that a person owns all and only the mental states that are realized in her brain (I will come back to this idea). However, that presumption can be overturned if there is reason to think that a person's introspectively accessing a mental state (or her ability to do so) guarantees that the state occurs in her mind.

There are three ways to interpret this last idea. It could be that  $S$ 's introspecting  $m$  "guarantees" that  $m$  occurs in her mind in the sense that 1)  $S$ 's introspecting  $m$  is a necessary condition for  $m$ 's occurring in her mind, 2)  $S$ 's introspecting  $m$  is a sufficient condition for  $m$ 's occurring in her mind, or 3)  $S$ 's introspecting  $m$  is a necessary and sufficient condition for  $m$ 's occurring in  $S$ 's mind.

Considered as a necessary condition, the claim is clearly too strong. It does not allow there to be *any* mental states, of any kind, that are a part of one's mind yet of which one cannot, at will, become introspectively aware. While historically this view has had advocates, relatively few will accept it today. It conflicts the existence of anything like a Freudian unconscious, with robust results in cognitive psychology indicating that people are often unaware of their true motives, biases, and beliefs (Carruthers (2011, Ch. 11) contains a nice review), and with the fact that ordinary introspection can be shown to be unreliable in predictable ways (Hurlburt & Schwitzgebel, 2007). If  $S$ 's ability to introspect  $m$  is not a necessary condition on  $m$ 's occurring in  $S$ 's mind, then neither is it a necessary and sufficient condition.

However, a more promising view might be the second above, that  $S$ 's ability to introspect  $m$  suffices for  $m$  to occur in  $S$ 's mind, even if it is not necessary for  $m$  to occur in  $S$ 's mind. While this has some intuitive plausibility, it is important to see how strong a claim it really is. It suggests that introspection has the power to transform a mental state that otherwise would not have been a part of one's mind into a proper part of *oneself*. In the case of the twins, the idea would be that Tatiana's visual experience  $v$  becomes a proper part of Krista at the moment Krista becomes

introspectively aware of it. This is not unlike having the power to turn objects to gold with a touch of one's hand. For that reason, I will call the following principle 'Midas Touch':

*Midas Touch:* Subject *S*'s becoming introspectively aware of *m* suffices for *m*'s occurrence within *S*'s mind.

Midas Touch is explored—and is refutation sought—in the next section.

## 5. Integration and Ownership

Recall that we have rejected *S*'s ability to introspect *m* as a *necessary* condition on *m*'s occurring in *S*'s mind. A person can be the owner of mental states to which she lacks introspective access. Accepting this much commits one to holding that there is some set of conditions that does *not* include *S*'s being able to introspect *m* that is *sufficient* for *m*'s occurring within *S*'s mind. What might those sufficient conditions be? I offer 'Brain Based' as an answer:

*Brain Based:* if mental state *m* is realized in *S*'s brain and has at least some causal and inferential interaction with *S*'s other mental states, then *m* occurs in *S*'s mind.

(Note that Brain Based is *not* offered as a necessary condition on ownership). Brain Based allows that poorly integrated mental states are nevertheless owned by the individuals in whose brains they occur. If Brain Based is indeed a sufficiency condition on ownership, we are in need of a special reason for why *S*'s introspectively accessing *m* should *also*, all by itself, suffice for *m*'s occurring within *S*'s mind.

The most likely reason one might give traces back to the idea of integration. If mental state *m* is closely causally and inferentially integrated with subject *S*'s other mental states, we have at least some reason to think that *m* occurs within *S*'s mind. This is one of the guiding principles both of neo-Lockean accounts of personal identity (Parfit, 1984), and of the "extended mind" program in cognitive science, which proposes to extend the boundaries of cognition beyond the traditional barriers of skull and brain. Extended mind theorists will count any information-bearing state with the right kind of causal and inferential integration with a person's other mental states as part of his or her mind, regardless of where (inside or outside the skull) or in what (brain

or iPhone) those states are realized (Clark & Chalmers, 1998). Putting these ideas in the form of a claim about ownership, let *Strong Integration* the following thesis:

*Strong Integration*: if mental state  $m$  is strongly causally and informationally integrated with subject  $S$ 's other mental states, then  $m$  occurs within  $S$ 's mind.

We can leave open for now how much integration is required for strong integration. If  $S$ 's ability to introspect  $m$  guarantees that  $m$  will be strongly integrated with respect to  $S$ 's other mental states, then Strong Integration will help vindicate Midas Touch.

My first response here is that Strong Integration (as a condition on ownership) is more plausible with respect to some kinds of mental states than others. Specifically, it is most plausible with respect to what we can call *core psychological features*. These are a person's memories, beliefs, desires, intentions, and personality traits. Fast, widespread, and permanent changes in the core psychological features inhering in a body and brain raise the question of whether the same *person* has continued to exist (at least insofar as we find psychological continuity theories of personal identity plausible). This is why they are counted as "core" psychological features. And what makes some set of core psychological features cohere into those of a single individual is that they bear some special causal (and inferential) integration with each other. Indeed, one could reasonably hold that the very existence a person at some point in time *depends* upon there being some such causally and inferentially integrated network of states in place.

If that is what persons fundamentally *are*, it will be hard to see how any core psychological state that has a place within such a network could not be a proper part of the person it helps realize. For it will be hard to see what factor could matter *more* in our appraisals of who owns the mental state. Considerations of brain location—highlighted by the case of the twins—will appear secondary to causal and inferential integration.

Granting these points about the link between ownership and integration for core psychological features, we need not conclude that they hold for *all* kinds mental states. A possible interpretation of the twins' situation may still be this: there are, between the two twins, two *distinct* networks of core psychological features: two distinct sets of beliefs, desires, intentions,

memories, and personality traits, where the states within each set are causally linked in the way necessary (we are assuming) for single personhood.<sup>11</sup> All of this leaves room for the hypothesis that the thalamic bridge allows each of their (distinct) networks of core psychological features to have an unusually close causal connection to some of the *perceptual* states of their twin—a connection of the kind that each person normally has only to her own perceptual states.

Crucially, perceptual states (and sensations generally) are not themselves core psychological features. Fast, radical, and permanent changes in the contents of perceptual states do not suggest changes in personal identity. So, whereas there is some reason to think that a certain belief might form a part of *S*'s mind *just because* it is closely causally and inferentially integrated with other beliefs (for what else could it be for there to exist a single person at a certain time than for there to be such a network?), this reasoning does not extend to perceptual states. We do not lose our grip on the very notion of personhood in holding that a perceptual state is tightly integrated within the mind of someone whose perceptual state it is not, in the way we *might* with beliefs, desires, and (quasi)-memories. Thus, Strong Integration can be seen as overextending a reasonable, neo-Lockean, thought about single-personhood's requiring an intimately causally integrated set of core psychological features to the far less obvious thesis that *any* mental state that is well integrated with one's core psychological features *must* form a part of one's own mind.

We can turn now to the Extended Mind approach to ownership to see if it might better support Midas Touch. We may assume that defenders of the Extended Mind approach will hold that Strong Integration is a sufficiency condition on the ownership of all kinds of mental states (to the extent that they exclude some states, the approach offers no support for Midas Touch with respect to those states). For I think we can grant, for the sake of argument, that Strong

---

<sup>11</sup> As part of the OT scenario, I am assuming that Krista and Tatiana do in fact have two distinct sets of core psychological features, and therefore count as two distinct people. Of course, if there were one just one causally and inferentially integrated set of core psychological features between them, we might conclude that they are in fact only one person with an unusual body. In that case, there would be no threat to IEM. Many of the anecdotal reports from Dominus, however, suggest that there are indeed two sets of preferences between the twins (one reportedly loves ketchup, while other hates it and tries to scrape it off her own tongue while her sister is eating it!). However, some of the anecdotes also suggest that their identities are more closely fused (Krista says, "I have two pieces of paper," while she is holding one, and her sister is holding one). For the purposes of this paper, we can simply stipulate that the propositional attitudes of each twin are causally and inferentially isolated from each other, and that they only have introspective access to each other's perceptual states. This is certainly an empirically (and logically) possible interpretation of their situation.

Integration is indeed a sufficiency condition on the ownership of all kinds of mental states. Even so it is doubtful that *S*'s introspecting *v* is sufficient for the strong integration of *v* within *S*'s mind. Here we can look to Clark & Chalmers' (1998) work itself for support. They reasonably grant that not all instances of causal interaction between internal and external representational states will count as the right kind of coupling for the states to be considered parts of the same mind (or for them to be counted as mental states at all). "What really counts," they emphasize, "is that the information is easily available when the subject needs it" (p. 15). In short, "*reliable* coupling is required," for a causally integrated information-bearing state to count as one of the subject's own mental states (p. 11, emphasis in original).

As reliability implies the *consistent* availability of a state (or faculty) to the subject when she needs it, we can see that the kind of strong integration that makes *v* a part of *S*'s mind will be a *diachronic* matter—it is something that can only hold and be assessed over an extended period of time. Introspection, on the other hand, is a synchronic matter. A person's becoming introspectively aware of a mental state is something that occurs in an instant, as evidenced by our ability to become introspectively aware of fleeting thoughts and imaginings that themselves last only a moment. This means that *S*'s ability to introspect *v* does not by itself entail that the reliable integration required for ownership is in place.

In light of this, we can tweak OT to guarantee that it is a situation where the diachronic conditions necessary for strong integration do not hold, but where Katrina is nevertheless introspectively aware of *v*. We can simply note that it is a logical (and, for all we know, nomological) possibility that the thalamic link provides for a kind of intermittent, unreliable access by each twin to the other's visual experiences. It is only occasionally (and unpredictably) that each twin's experiences become introspectively available to the other. This may in fact be the case with Katrina and Tatiana. In this situation Katrina might become introspectively aware of *v*, while the diachronic requirements on strong integration (and ownership) are not met. And if that is possible, then it is also possible that Katrina might, in a moment of inattention or confusion, misascribe *v* to herself. This would be a violation of Introspective Immunity, as it would be a case where Krista's introspection-grounded judgment is false for the sole reason that she has

misidentified the owner of  $v$  (her sister) as being herself. And it is a violation consistent with accepting Strong Integration as a blanket condition on the ownership of all kinds of mental states.

### 5.1 Individuating brains

Recall that, when addressing Rosenthal's case of Dissociative Identity Disorder, I noted that very often we count persons as the owners of mental states that are not well integrated within their overall psychologies. This suggested that there is more to ownership than Strong Integration—that if a high level of integration of a state with one's other core psychological states is sufficient for ownership, it is not necessary. That "something more," was expressed in the Brain Based principle above. Where a mental state  $m$  is realized in  $S$ 's brain and has at least some causal interaction with  $S$ 's other mental states, it will count as being owned by  $S$ . Yet this opens the door to a new objection that will have already occurred to some: how do we know that the visual cortices in which  $v$  is realized—so far described as being only Tatiana's—are not also parts of Katrina's brain? If they are, then, even if  $v$  is not strongly integrated with Katrina's core psychological features, it should still count as being owned by Katrina (for the same reason that other of Katrina's poorly integrated states count as her own—they occur in her brain).

Here we have to consider principles for individuating brains. We can recall that each twin has two cerebral hemispheres, and that each set of hemispheres is fused to the other only at their thalami. While the twins may well share part of their brains (their thalami), this does not suggest that they share *every* part of their brains—even when there is a causal connection between two parts. Consider how we individuate other bodily organs: often two organs are fused at some point (the stomach and small intestine, say), and there are relatively direct causal interaction between those parts, without parts of one organ becoming parts of the other. If the twins were attached only at the shoulder, no one would conclude that one girl's entire arm was also the other's. So there is no general principle deriving from how we individuate bodily organs that would entail that Tatiana's visual cortices are part of Krista's brain, even if there is causal cross-talk between the two.

One might nevertheless object that, in the special case of brains, our principles of individuation are (or should be) such that, wherever there is reliable causal interaction between

two representational states and/or cognitive capacities, the areas realizing them should count as parts of the same brain. In my view, this mistakes criteria for individuating minds for criteria for individuating biological organs (it turns the extended mind thesis into the extended brain thesis). But let us grant the proposal for the sake of argument. It would entail that, wherever there is extended mind (in the Clark & Chalmers sense) there is also extended brain. But then the converse will also be true: where there is not extended mind, there will also not be extended brain. This means that the reply above concerning the synchronic nature of introspection is again relevant. The kind of reliable causal connection between two neural regions necessary for counting both as parts of the same brain need not hold between *v* and Krista's introspective processes. The relation between them may be intermittent and unreliable, though still allow for synchronic acts of introspection. When the causal relation is of this sort, and where the existing anatomy offers no presumption in favor of their being parts of the same brain, it is very hard to see why the two neural regions could be counted parts of the same brain.

The presumption that Tatiana is the sole owner of *v* therefore stands. This entails that Krista could become introspectively aware of *v*—a visual experience as of a toy pony—yet make an error of misidentification in judging, “I am having a visual experience as of a toy pony.”

## **6. Phenomenological considerations**

My assumption is that, in the OT scenario, what it is like for Krista to be introspectively aware of *v* (Tatiana's visual experience as of a toy pony) is more or less the same as what it is like for Tatiana to be introspectively aware of the same token experience. There may be differences in their overall phenomenologies, of course. One may simultaneously feel her feet cooling in a tub of water, the other warming in the sun. But, with respect to the contribution of *v*, my assumption is that their phenomenologies are more or less the same. The sort of introspective relation Krista holds to *v* entails that *v* will contribute to her phenomenology in whatever way introspected states typically contribute to one's phenomenology. So it is not a sort of “blind sight” counterpart to introspection—where she simply finds herself with opinions about Tatiana's experiences, for reasons she knows not—that I have in mind in saying that Krista is introspectively aware of an experience that occurs outside of her own mind.

Does *v* nevertheless occur within Krista's stream of consciousness? If we hold that every state within one's stream of consciousness is *ipso facto* a part of one's mind, then the answer is no. This paper's claim could then be put as: one can be introspectively aware of a state, and have that state contribute in the normal way to one's phenomenology, without that state occurring in one's stream of consciousness. If instead we allow that one can, in extraordinary circumstances, have a state in one's stream of consciousness that one does not own (just as one can have objects in one's field of view that one does not own), then the answer may that Krista does indeed have *v* in her stream of consciousness. I favor the second approach, which allows for the occurrence of experiences "in" one's stream of consciousness that do not occur within one's mind. And I take the preceding sections to be arguments for that possibility. Moving to the language of "streams" or "fields" of consciousness (and their owners) does not by itself offer new leverage on the basic question we have been discussing all along.

One might still object as follows: "Look, if the experience contributes to her phenomenology, then she is *having* the experience. And if she is having the experience, then it occurs in her mind" (Cf. Coliva, 2002, p. 28). My response: what do we mean by *having* an experience? If, in extraordinary circumstances, one can have an experience that does not occur within one's mind, then, yes, she is having the experience; if one cannot, then, no, she is not having the experience—even though she is introspectively aware of it, and it contributes to her phenomenology (the same remarks apply, *mutatis mutandis*, to the notions of "enjoying" or "undergoing" an experience). The flat-footed assertion that "having" or "enjoying" a mental state entails its occurrence in one's mind simply begs the question at issue.

But have I begged the question in the opposite direction, simply asserting that an introspected state—a state one *has* or *enjoys*—need not occur within one's mind? Certainly not. I have given independent reasons for why an certain introspected state one "enjoys" might not occur in one's own mind (*viz.*, it is not in one's brain, and is not well integrated with one's other psychological states), and have considered and dismissed the most plausible independent reasons for thinking that all introspected states must occur within one's own mind (considering, e.g., extended mind theories, and neo-Lockean conceptions of personhood). Of course, one could still maintain that the logical necessity of Introspective Immunity (and of Midas Touch) is a primitive

matter, admitting of no outside justification. But, in that case, *any* challenge to Introspective Immunity will amount to “begging the question” against the principle; for then it will have no outside support of the kind that could be undermined without assuming the falsity of the principle. In such a situation, it is perfectly acceptable to question the truth of a principle by offering competing considerations.

That said, there is at least one argument grounded in phenomenology that has some bite against the present proposal. On the sort of view I have been proposing, it should be possible not only for Krista to be introspectively aware of a visual experience that is not her own, but also of a *pain* that is not her own. And if it is granted that the states of which one is introspectively aware contribute, in the normal way, to one’s phenomenology, then I seem committed to holding that Krista might feel a pain that is not her own. That does not in itself seem to me any more problematic than the sort of cases already considered. However, it might be insisted that, if Krista can literally feel Tatiana’s pains, then we have just as strong a moral duty to her as to Tatiana to avoid causing Tatiana pain.<sup>12</sup> But then, why would we have a moral duty to Krista to remove a pain that is not her own? The answer cannot simply be that Krista suffers at the thought of Tatiana feeling pain; Krista will deserve our concern in this situation even if she relishes the idea of Tatiana feeling pain. The basic idea—which I can only sketch here—is that there may be constitutive links among personhood, being an object of moral concern, and the capacity for feeling pain. These links *might* create problems for the claim that Krista could literally feel a pain that is not her own.

My response is to hew to the line already defended and to insist that, in unusual circumstances such as this, we can have a moral duty to a person to stop pains that are not his or her own. What matters is whether one feels a pain. And we have a *prima facie* duty to end pains that a person feels, whether or not the pains are owned by that person. Of course, people usually cannot feel anyone’s pains but their own; but, in very unusual circumstances such as this, things could be different. If this entails an outright absurdity, I do not see it.

Further, it is important to note that many mental states with phenomenal character—such as visual experiences—do not present the same kind of challenge. A person is not a subject

---

<sup>12</sup> Thanks to XX for pressing this objection.

of moral concern simply because he or she has (or is introspectively aware of) visual experiences with a certain phenomenal character. True, if Krista feels distress at watching horror movies, we might, out of consideration, ask Tatiana not to watch a horror movie (on the assumption that Krista might become introspectively aware of Tatiana's ensuing visual experiences and feel distress). However, the reason Krista becomes an object of moral concern here is not due to the phenomenology of the visual experience itself, but due to her own feelings of distress that will ensue from her "enjoying" that visual phenomenology. Unlike the case of pain, a horror film *aficionado* might have the very same visual phenomenology as we are envisioning Krista to have, without suffering in the least. It is the cognitive effects that visual experiences have on people, not the visual experiences themselves, that make people objects of moral concern. Thus, if there is anything to the objection from moral appraisal, it applies only to a subset of the (non-factive) states to which one might have introspective access and which contribute to one's phenomenology.

## 7. Conclusions

I think a strong case has been made for viewing the OT scenario as a logically possible situation where a person is introspectively aware of a non-factive mental state that is not her own. We should only conclude that  $v$  belongs to Krista if her introspecting  $v$  is by itself sufficient for  $v$ 's being her own. And I have argued that there is no good reason to think that it is. We should reject Midas Touch. Further, as  $v$  occurs only in Tatiana's brain, and is not well integrated with anyone else's core psychological features, we should conclude that she is its sole owner. This means it is possible that Krista might (in a moment of inattention) mistake  $v$  as being her own and falsely judge, "I am having a visual experience as of a toy pony." Introspective immunity is therefore not a logical necessity, *pace* Shoemaker and others. Further, we are left with no reason to see Introspective Immunity as a nomological necessity either.

What broader lesson does this hold for self-knowledge and introspection? Here I only have space to comment on one upshot that seems to me significant. Shoemaker closes his classic 1968 essay with some reflections on the significance of there being some introspective judgments that are absolutely IEM: "It has often been held to be one of the defining features of the realm of

the mental,” he notes, “that each person knows of his own mental or psychological states in a way in which no other person could know of them” (p. 565). Here Shoemaker is referencing *epistemic accounts* of the mark of the mental, where what is distinctive about the mental is held to be the way in which mental states are known about by their owners. Epistemic accounts are no longer widely endorsed as comprehensive criteria for the mental (for one thing, they seem to rule out the possibility of subconscious and introspectively inaccessible mental states). However, Shoemaker is not proposing that we accept a comprehensive epistemic account. Rather, he thinks that we can “put what is true” in such views by saying that “there is an important and central class of psychological predicates, let us call them ‘P\*-predicates,’ each of which can be known to be instantiated in such a way that knowing it to be instantiated in that way is equivalent to knowing it to be instantiated in oneself” (p. 565). These P\* predicates are presumably things like ‘believe’, ‘want’, ‘see’, and ‘feel’. If knowing such predicates to be instantiated through introspection really is “equivalent” to knowing them to be instantiated in oneself, then we have a neat explanation of Introspective Immunity (see Coliva (2012) for a more recent articulation of the same idea).

While Shoemaker grants that there are (non-P\*) mental and psychological predicates, such as “is highly intelligent,” where knowing them to be instantiated is not equivalent to knowing them to be instantiated in oneself, he nevertheless sees a path from P\* predicates to definitions of mentality in general. Says Shoemaker:

I think that those [psychological predicates] which are not P\*-predicates are classified as psychological predicates only because they are related in certain ways to those which are; e.g., they are predicable only of things of which some P\*-predicates are also predicable (p. 566).

Shoemaker’s idea is that, while there may be psychological states the self-ascription of which is not IEM, these only occur in the context of a being who bears psychological states of the P\* sort, whose self-ascription *is* absolutely IEM. In short, according to Shoemaker, all ascriptions of mental states to an organism are parasitic on the existence of some states in that organism that are absolutely IEM. If that is right, then there is an epistemic mark of the mental after all, albeit one step removed. Understanding how there can be any mental states whatsoever will require understanding how a subject can know about certain of his own mental states in a way that gives

him absolute immunity to errors of misidentification. If that is right, then we really won't understand the nature of the mental generally, until we understand how it is that we can make the error-immune judgments in question.

One way to challenge this claim is to show that there are judgments involving *non*-mental predicates that are *also* absolutely immune to errors of misidentification. Evans (1982), Wright (1998, 2012), and others have pursued deflationary accounts of IEM along those lines, arguing many "direct" perceptual judgments—"That was fast"; "You are very close"—also involve a lack of identification, and are, accordingly, IEM (relative to the subject term used). If that is right, then a judgment's being IEM will have nothing special to do with mentality. I have approached the issue from the opposite direction, attempting to show that ordinary introspective judgments are not absolutely IEM after all, even when they involve non-factive mental states like perceptual experiences. If they are not, then there is no reason to think that the essence of the mental lies in the kind of epistemic security associated with IEM.

We can visualize the debate in the following way. On the view of mentality that Shoemaker recommends, the entire set of a person's mental states can be represented as a blue circle, and the set of mental states to which that person has introspective access (and that are P\* states), as a red circle falling wholly within the boundary of the blue. We can add to this that only the states within the red circle are featured (or are poised to be featured) in the person's stream of consciousness. For Shoemaker and many others, it is an absolute necessity that the boundaries of the red circle fall wholly within the blue circle. Working from this perspective, one might conclude, with Shoemaker, that the having of red-circle states grounds the having of blue-circle states. On the view I have proposed, the boundaries of the red circle may (in extraordinary cases) extend beyond those of the blue: token mental states may contribute to a person's stream of consciousness without occurring in that person's mind. This view does not carry with it the same implication that understanding mentality in general will require first understanding the nature of introspection, or of self-ascriptions that are absolutely immune to errors of misidentification.

**Acknowledgements:** *Versions of this paper were presented at the Fifth Consciousness Online conference and at the 2013 Southern Society for Philosophy and Psychology Conference. Special*

*thanks to Elizabeth Schechter, Annalisa Coliva, John Schwenkler, Joel Smith, and Serife Tekin for their valuable criticisms and suggestions.*

## References

- Armstrong, D. (1968). *A Materialist Theory of Mind*. London: Routledge.
- Bottini, G., Bisiach, E., Sterzi, R., & Vallar, G. (2002). Feeling touches in someone else's hand. *Neuroreport*, *13*(2), 249-252.
- Byrne, A. (2005). Introspection. *Philosophical Topics*, *33*(1), 79-104.
- Campbell, J. (1999a). Immunity to Error through Misidentification and the Meaning of a Referring Term. *Philosophical Topics*, *26*(1&2), 89-104.
- Campbell, J. (1999b). Schizophrenia, the space of reasons, and thinking as a motor process. *The Monist*, *82*(4), 609-626.
- Carruthers, P. (2011). *The Opacity of Mind: An Integrative Theory of Self-Knowledge*. Oxford: Oxford University Press.
- Clark, A., & Chalmers, D. (1998). The Extended Mind. *Analysis*, *58*(1), 7-19.
- Coliva, A. (2002). Thought Insertion and Immunity to Error Through Misidentification. *Philosophy, Psychiatry, & Psychology*, *9*(1), 27-34.
- Coliva, A. (2012). Which 'Key to All Mythologies' About the Self? A Note on Where the Illusions of Transcendence Come from and How to Resist Them. In S. Prosser & F. Recanati (Eds.), *Immunity to Error Through Misidentification* (pp. 22-45). Cambridge: Cambridge University Press.
- Dominus, S. (2011, May 29, 2011). Could Conjoined Twins Share a Mind? *The New York Times Magazine* -- URL: <http://www.nytimes.com/2011/05/29/magazine/could-conjoined-twins-share-a-mind.html>.
- Dretske, F. (1999). The Mind's Awareness of Itself. *Philosophical Studies*, *95*(1-2), 103-124.
- Evans, G. (1982). *The Varieties of Reference*. Oxford: Oxford University Press.
- Finkelstein, D. H. (1999). On Self-Blindness and Inner Sense. *Philosophical Topics*, *26*(1&2), 105-120.
- Hogan, M., & Martin, R. (2001). Introspective misidentification: An I for an I. In A. Brook & R. C. DeVidi (Eds.), *Self-reference and Self-awareness* (pp. 205-214). Philadelphia: John Benjamins Publishing Company.
- Hurlburt, R. T., & Schwitzgebel, E. (2007). *Describing Inner Experience? Proponent meets Skeptic*. Cambridge, MA: MIT Press.
- International Society for the Study of, T., & Dissociation. (2011). Guidelines for treating dissociative identity disorder in adults, third revision. *J Trauma Dissociation*, *12*(2), 115-187. doi: 10.1080/15299732.2011.537247

- Lane, T., & Liang, C. (2011). Self-Consciousness and Immunity. *The Journal of Philosophy*, 108(2), 78-99.
- Lormand, E. (2000). Shoemaker and 'inner sense'. *Philosophical Topics*, 28(2), 147-170.
- Lycan, W. (1996). *Consciousness and Experience*. Cambridge, MA: MIT Press.
- Nichols, S., & Stich, S. (2003). *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding of Other Minds*. Oxford: Oxford University Press.
- Parfit, D. (1984). *Reasons and persons*. Oxford Oxfordshire: Clarendon Press.
- Pryor, J. (1999). Immunity to Error through Misidentification. *Philosophical Topics*, 26(1&2), 271-304.
- Rosenthal, D. (2005). *Consciousness and Mind*. Oxford: Clarendon Press.
- Rosenthal, D. (2012). Awareness and Identification of Self. In J. Liu & J. Perry (Eds.), *Consciousness and the Self: New Essays* (pp. 22-50). Cambridge: Cambridge University Press.
- Shoemaker, S. (1968). Self-knowledge and self-awareness. *Journal of Philosophy*, 19, 555-567.
- Shoemaker, S. (1994). Self-Knowledge and 'Inner Sense': Lecture II: The Broad Perceptual Model. *Philosophy and Phenomenological Research*, 54(2), 271-290.
- Smith, J. (2006). Which Immunity to Error? *Philosophical Studies*, 130, 273-283.
- Wittgenstein, L. (1969). *The Blue and Brown Books*. New York: Harper.
- Wright, C. (1998). Self-knowledge: The Wittgensteinian Legacy. In C. Wright, B. C. Smith & C. MacDonald (Eds.), *Knowing Our Own Minds* (pp. 13-45). Oxford: Clarendon Press.
- Wright, C. (2012). Reflections on Francois Recanati's 'Immunity to error through misidentification: what it is and where it comes from'. In S. Prosser & F. Recanati (Eds.), *Immunity to Error through Misidentification* (pp. 247-280). Cambridge: Cambridge University Press.