Craniopagus Twins and the Possibility of 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). The basic idea, tracing to Wittgenstein and Shoemaker, is that if one accurately discerns some mental state through introspection, one cannot then err in ascribing it to oneself. Many have taken such errors to be logically or conceptually impossible. 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 in fact possible. It turns out that a stronger case can be made for the possibility of introspectively misascribing sensory and perceptual states than for beliefs, desires, and memories. Implications are discussed for some views of the relation between introspection and personal identity.

I.

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. If you discern some mental state through introspection and—using the first person pronoun[1]—ascribe it to yourself, it seems you cannot go wrong in that judgment for the sole reason that you have ascribed the

[1] If a proper name is used, the judgment could fail in the odd circumstance that one is not in fact the referent of that name.

mental state to the wrong person (though you might err in your judgment about what *kind* of mental state was detected—e.g., mistakenly ascribing to yourself an itch instead of a pain).

Let *Introspective Immunity* be the thesis that first-person self-ascriptions of mental states, where introspection provides the grounds, are immune to error through misidentification. The thesis is itself neutral on how introspection works. It is sufficient that introspection be characterized as the usual way 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 ascribed the state to the wrong person (again, assuming one uses the pronoun '*I*').

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 or conceptual (or metaphysical) necessity? Or is their "impossibility" contingent upon things being as they normally are—for instance, upon one's having ordinary human anatomy? Shoemaker (1968) clearly holds the stronger view. 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). Many others agree (Campbell, 1999a, p. 93; Coliva, 2002; Pryor, 1999, p. 297; Smith, 2006, p. 275). Nor is it easy to imagine a situation that would clearly count as one where Introspective Immunity is violated.[2]

[2] There are proposed counterexamples to Introspective Immunity in the existing literature, including Lane & Liang (2011), Rosenthal (2012), and Hogan & Martin (2001). Dissatisfaction with those accounts forms some the impetus for this paper, though there is not adequate space here to properly discuss those proposals.

My project here is to describe a clear counterexample to Introspective Immunity, showing that contingent features of our own anatomies are what normally prevent mistakes of the relevant kind. The argument appeals to an actual case of craniopagus twins who are conjoined at the brain. As unusual as their case may be, it is not obvious, on its face, that it is one where Introspective Immunity might be violated. Thus, much of the discussion will focus on distinguishing interpretations of their situation that challenge Introspective Immunity from ones that do not.

The larger point of this discussion is to gain some clarity on exactly 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-conscious subjects" (2006, p. 274). Intuitively, the impossibility of such errors arises out of some fundamental and distinctive feature of self-consciousness, self-awareness, and (perhaps) of The Self itself. Yet the exact nature of the connection is far from obvious.[3] A key conclusion of this paper will be that a much stronger case can be made for the possibility of misascribing to oneself an introspectively discerned perceptual or sensory state than for an introspectively discerned belief or desire. Whatever lessons Introspective Immunity holds for the connection between introspection and the self, we should not assume it holds the same lessons with respect to all mental states.

In addition, it is worth noting that much of the IEM literature has focused on the question of what it is that guarantees our immunity to errors of misidentification (when we

[3] Though see Wright (2012) for an argument that there is no special connection between judgments that are IEM and self-consciousness. Wright's skepticism on this score is compatible with the view developed here.

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.[4] Often these debates take the absolute necessity of Introspective Immunity as a basic datum that any account of its basis must get right. For example, 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[5]) would not be IEM. However, if this paper's main thesis is correct, then 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.

II.

Krista and Tatiana Hogan are 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 documented in a 2011 *New York Times Magazine* article on the twins. With

[4] See Smith (2006, p. 276) for a clear articulation of this approach to explaining Introspective Immunity.

[5] In Smith's example, a person is wired up so as to feel an itch when someone else (who is really in pain) presses a button; and, further, the person feeling the itch happens to mistake the feeling of itchiness for one of pain. Such a person introspectively seems to be in pain and, on that basis, is able to infer that *someone* is in pain even after concluding that he himself is not in pain (because he believes someone else's correct testimony that what he is really feeling is an itch, and has been told of the unusual wiring and button-pressing situation). This is offered as a counterexample to Wright's thesis that "a claim made on a certain kind of ground involves immunity to error through misidentification just when its defeat is *not* consistent with retention of the grounds for existential generalization" (1998, p. 19).

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.[6] 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. For the thalamus is known to be a multi-modal "relay center" 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 that 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. First, 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 closed, that Tatiana is seeing a doll. It might be that information concerning the visual experience *v* (as of a doll) had by Tatiana

[6] A video depicting some of these capacities, together with Dominus's (2011) article, may be retrieved at the following address: http://www.nytimes.com/2011/05/29/magazine/could-conjoined-twins-share-a-mind.html

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 then misascribe it to 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. 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 token mental state yet err in self-ascribing it.[7]

Thus, in the doll case, we need to suppose that there is only one token experience *v* (as of a doll) between the twins, 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

[7] The TT scenario does not threaten Introspective Immunity for much the same reason that reports of "inserted thoughts" in schizophrenia do not threaten it. When a patient reports that someone else's thought has been inserted into his mind, he can plausibly be interpreted as claiming that he has, *within his own mind*, a mental state that is under the causal control of some external agent (Cf. Campbell, 1999b; Gallagher, 2000). While the patient is presumably wrong about the causal claim, any self-ascription of the mental state remains correct—assuming that "owning" the state amounts to its occurring in one's own mind.

is that it is a *possibility*. Further empirical investigation might 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 in order to challenge Introspective Immunity's claim to being a necessary truth, we simply need a plausible description of a possibility (not an actuality) that challenges it. The appeal to the (quite actual) Hogan twins should be understood in that spirit: as enabling us to form a vivid picture of a *possibility* that might otherwise seem not to exist, by showing us that it is very close to, if not identical with, an actuality.

OT presumes that a distinction can be drawn between a first order visual state and the state or process by which one ordinarily becomes introspectively aware of such a state—that they are *distinct existences*. Shoemaker (1994) denies the coherence of such views. I will not address his reasons for doing so here.[8] Instead I will assume that a "distinct existences" view of introspection *might* be true. There are currently a number of such views on the market, including Higher Order Thought theories (Rosenthal, 2005), Higher Order Perception 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 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

[8] But see Byrne (2005), Rosenthal (2012), Finkelstein (1999), and Lormand (2000), for critiques of Shoemaker on this point.

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.[9]

Even granting OT's possibility, together with a distinct existences view of introspection, we still do not have an obvious challenge to Introspective Immunity. For OT might be a situation where the twins *share* ownership of *v*, in the sense that *v* occurs in both of their minds (For it is *not* being assumed that the only states in one's mind are those realized in one's brain). If they share ownership if *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. However, that presumption can be overturned if there is reason to think that the act of introspectively accessing a mental state (or the ability to do so) is itself what makes it the case that the state occurs in one's mind.

Let us call *Midas Touch* the thesis that *S's* accurate introspective detection of mental state *m*—or *S's* current ability to do so—is what makes it the case that *m* occurs within *S's* mind. We can then ask whether Midas Touch is a necessary condition for *m's* occurring in *S's* mind, a sufficient condition, or both.

[9] It is assumed here that there are two brains between the twins, not one unusual brain. The two brains assumption is supported by the fact that, aside from their thalamic connection, the twins' brains are more or less normal. For in other cases where two body parts or organs are fused at some point, we do not infer that they are really just one.

Considered as a necessary condition, Midas Touch is 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 may have had advocates, relatively few contemporary philosophers and psychologists will be tempted by it. 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 (see, e.g., Carruthers (2011) for a relevant review), and with fact that ordinary introspection can be shown to be unreliable in predictable ways (Hurlburt & Schwitzgebel, 2007).

A more popular view might be that Midas Touch constitutes a sufficiency condition—that *S's* introspecting *m* suffices for *m* to occur in *S's* mind (even if it is not necessary for *m* to occur in *S's* mind). But why think that it does? Provided we have abandoned Midas Touch as a necessary condition on mental state ownership, we are already committed to there being some set of conditions that does not include *S's* being able to introspect *m* that is sufficient for making it the case that *m* occurs within *S's* mind. We should like a reason, then, for why *S's* introspectively accessing *m* should *also,* all by itself, suffice for *m's* occurring within *S's* mind. The next section considers and rejects what I think is the most plausible such reason, while conceding to the contrary perspective that it contains a grain of truth.

III.

Let us call a person's memories, beliefs, desires, and personality traits her *core psychological features*. 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—this is why they are counted as "core" psychological features. Further, it is plausible that what makes some set of core psychological features cohere into those of a single individual is that they bear some special causal (and inferential) relation to 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 one could somehow identify these causal/inferential links among states with the introspective accessibility relation—if they were, somehow, one and the same thing—then one might have a reason for viewing Midas Touch as a sufficiency condition. For the introspectability of these states—now conceived of as their close causal/inferential integration with each other—would be what underwrites the existence of a person in the first place.

That is, instead of (incorrectly) drawing the boundary of a single mind at the set of mental states to which a certain person has introspective access, one might define *a person* in terms of a network of core psychological features that bear a close causal and inferential connection with each other. 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 the sort of intimate causal connection now being equated with introspection. Seen in this light, Midas Touch gains plausibility as a sufficiency condition.

While I think there are real problems with identifying the introspective accessibility relation with the kind of causal relation a person's core psychological features bear to each other,

let us grant the general idea for the sake of argument. For it is important to see that it does not provide reason to accept Midas Touch as a sufficiency condition. A possible interpretation of Krista and Tatiana's situation may still be this: there are, between the two girls, 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. All of this leaves room for the hypothesis that the thalamic bridge allows each of the girl's networks of core psychological features to have an unusually close causal connection to some of the *perceptual* states of her twin—a connection of the kind that each person normally has only to her own perceptual states.

Crucially, perceptual states 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 t1 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 person is introspectively aware of a perceptual state that is not within her own mind, in the way we *might* with belief and desire. Thus, Midas Touch taken as a sufficiency condition 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 implausible thesis that *any* mental state of which one is introspectively aware *must* form a part of one's own mind.

And recall again that, having rejected Midas Touch as a necessary condition, we are already committed to there being some (as yet unarticulated) criterion independent of introspection that is sufficient for a particular mental state's occurring in a certain person's mind. In the case of OT, *v* may fulfill this criterion with respect to Tatiana, but not with respect to Krista. So, there appears no obvious barrier to concluding that OT describes a context where a person is introspectively aware of a mental state that does not occur within her own mind. And if a person could be introspectively aware of such a state, it is reasonable to think she might, in a moment of distraction, misascribe it to herself. Introspective Immunity might be violated.

IV.

Nevertheless, the intuition that Midas Touch articulates a sufficiency condition on mental state ownership is hard to abandon. Some of its attraction may trace to the influence of a broadly Humean conception of selfhood. Hume, when introspectively examining his ideas and impressions—entering "most intimately into what I call *myself*"—could not find anything over and above those very states on which to hang the title of *self*. He thus proposed that selves "are nothing but a bundle or collection of different perceptions, which succeed each other with an inconceivable rapidity, and are in a perpetual flux and movement" (2000/1739, p. 165). However revisionary his position was in its time, even Hume did not contemplate the possibility that some part of his introspected "bundle" may be a part of someone else (and not a part of himself). Of course, if what it is to be a self is just to be some particular bundle of conscious "perceptions"—if there is nothing *more* to it—then there are no competing factors (such as a body and brain) in

virtue of which two parts of a single bundle might not be parts of the same person. Introspective Immunity will be inevitable.

Yet, to the extent that particular selves tolerate periods of unconsciousness—and short of outright eliminating selves (of which Hume is sometimes accused)—there *must be* some further fact that determines whether some bundle at *t1* is a stage in the same self's career as another bundle at *t2*. The irony is that we may still be tempted to view the question of this further fact through a Humean lens, asking: what makes it the case that this one unified *bundle of consciousness* at *t1* is a stage of the same self as some other such bundle five days later? But if some bundles may contain states that are no part of the person who is introspectively aware of them, then this is not quite the right question to be asking.

References

Armstrong, D. (1968). *A Materialist Theory of Mind*. London: Routledge.

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.

Coliva, A. (2002). Thought Insertion and Immunity to Error Through Misidentification. *Philosophy, Psychiatry, & Psychology, 9*(1), 27-34.

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.

Finkelstein, D.H. (1999). On Self-Blindness and Inner Sense. *Philosophical Topics, 26*(1&2), 105-120.

Gallagher, S. (2000). Self-Reference and Schizophrenia: A Cognitive Model of Immunity to Error through Misidentification. In Z. D. (Ed.), *Exploring the Self: Philosophical and Psychological Perspectives on Self-Experience*. Philadelphia: John Benjamins.

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.

Hume, D. (2000/1739). *A Treatise of Human Nature*. Oxford: Oxford University Press.

Hurlburt, R.T., & Schwitzgebel, E. (2007). *Describing Inner Experience? Proponent meets Skeptic*. Cambridge, MA: MIT Press.

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.

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, Ludwig. (1969). *The Blue and Brown Books*. New York: Harper.

Wright, C. (1998). Sefl-Knowledge: The Wittgensteinian Legacy. In C. Wright, B. Smith & C. MacDonald (Eds.), *Knowing Our Own Minds*. 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.