Philosophical Psychology
2011, 1–17, iFirst

Routledge
Taylor & Francis Group

What certainty teaches

Tomas Bogardus

Most philosophers, including all materialists I know of, believe that I am a complex thing—a thing with parts—and that my mental life is (or is a result of) the interaction of these parts. These philosophers often believe that I am a body or a brain, and my mental life is (or is a product of) brain activity. In this paper, I develop and defend a novel argument against this view. The argument turns on certainty, that highest epistemic status that a precious few of our beliefs enjoy. For example, on the basis of introspection, I am certain that I am not in fierce pain right now. But if I am a complex thing like a body or a brain, then introspection might be a causal series of events extended in time. And any such process could go awry. So, if introspection is such a process, then I could gain good evidence that the introspective process has gone awry and that I am, contrary to appearances, feeling fierce pain right now. Therefore, the view that I am a complex thing like a body or a brain forces open the possibility that I cannot be certain that I am not feeling fierce pain right now. Since that is clearly not an open possibility, it follows that I am not a complex thing. I conclude by responding to three objections.

Keywords: Certainty; Incorrigibility; Infallibility; Introspection; Materialism; Self-Intimation

1. Introduction

In 1962, Kurt Baier argued that materialism about the mind entails introspective fallibility, e.g., that you might be wrong that you are in pain at a time when it seems to you that you are in pain. Since very many philosophers at the time accepted introspective infallibility, this was a significant result. In response, many materialists like David Armstrong (1963) squared their shoulders, accepted the implication, and haven't looked back. These days, it's hard to find a philosopher who accepts that introspection is infallible.

Even so, most philosophers these days accept that, though introspection is fallible, it's not *hyper*fallible. Introspection might get it somewhat wrong, they think, but

Tomas Bogardus is finishing his doctoral dissertation at the University of Texas at Austin.
Correspondence to: Tomas Bogardus, WAG 421, Mailcode C3500, University of Texas at Austin, Austin, TX 78712, USA. Email: tomasb@mail.utexas.edu

ISSN 0951-5089 (print)/ISSN 1465-394X (online)/11/0001-17 © 2011 Taylor & Francis
DOI: 10.1080/09515089.2011.579419

introspection can't get it radically wrong. It couldn't be, for example, that I am actually experiencing fierce pain right now though it introspectively seems to me that I am not. In this paper, I will extend Baier's argument to target this popular view. I will argue, roughly, that each of the standard accounts of introspection on which it is mechanistic—that is, a causal series of events extended in time—entails introspective hyperfallibility. If any one of the standard accounts of introspection is right, then we never have introspective certainty—every last one of our introspective beliefs is defeasible. This implication, I take it, is far less palatable than the one Baier pointed out.

Furthermore, I will argue that any version of what I call "the Complex View"—including the standard materialist view on which people are complex material objects like bodies or brains—entails that introspection might be mechanistic. What follows, I will argue, is that the Complex View forces open the possibility that none of our introspective beliefs is certain for us. Philosophers who believe that is not an open possibility will consider this to be a powerful argument against the Complex View—they will think certainty teaches us that the standard materialist view of human persons is false, and that human persons are simple.

2. The Main Argument

I am some thing. Many people think that I am a complex thing—a thing with parts—and that my mental life is (or is a result of) the interaction of (some of) these parts. Which complex thing am I? Perhaps I am a body, or more plausibly some part of a body such as a brain, or perhaps some special part of a brain. Other people think that I am not a complex thing at all. Rather, these people say that I am a simple thing—a thing with no parts—and my mental life is a basic activity of this simple thing, not a result of the interaction of any parts.[1]

In this paper, I will develop and defend a novel argument against the Complex View:

(1) If the Complex View is true, then I cannot be certain[2] that introspection[3] is not a causal series of events extended in time.
(2) If introspection is a causal series of events extended in time, then I could gain good evidence that I am feeling fierce pain right now.[4]
(3) For any propositions p and q, if (*i*) I cannot be certain that p is false, and (*ii*) p entails q, then I cannot be certain that q is false.
(4) So if the Complex View is true, then I cannot be certain that I could not gain good evidence that I am feeling fierce pain right now.

If you think it is more likely that the Complex View is false than that you cannot be certain that you couldn't gain good evidence that you're in fierce pain, then (1)–(4) constitute an argument for the conclusion that the Complex View is false. Let me now explain and motivate the premises and then respond to three objections.

3. Introspection as a Causal Series of Events Extended in Time

Visual perception is a process, a series of events extended in time whereby one comes to have beliefs about how the world is. We commonly take this to be a *causal* series of events extended in time: each member of this series of events is merely nomically sufficient for the next, and each member of this series could occupy a point or duration of time distinct from that of any other member in this series. For example, we think some story like this is now true of you: there is a surface with markings before you. This surface and these markings cause light to be reflected in a certain way into your eyes. This reflected light causes certain events on your retinas. These retinal events cause certain events in your optic nerve. These optic nerve events cause certain events in your visual cortex. Then you enjoy a visual experience, which represents (among other things) that there is a white surface with black markings before you.

Like visual perception, introspection is a process, a series of events extended in time. Unlike visual perception, however, introspection is a process by which one becomes aware not of the external world, but rather of the phenomenal character of her own experiences. For example, we think some story like this is now true of you: you have a visual experience, which represents that there is a white surface with black markings before you. Then you attend to some of the phenomenal character of your—the whiteness, say. And then, somehow, you end up with the belief (or awareness, or perception) that you are having a visual experience as of white. Many people think that introspection, like visual perception, is a *causal* series of events, involving some sort of *mechanism*. For example, Byrne writes: ''unless it's magic, I must have some sort of mechanism (perhaps more than one) for detecting my own mental states—something rather like my visual, auditory, and gustatory systems, although directed to my mental life'' (2005, p. 80) I will first discuss some mechanistic views of introspection before arguing for the premises of the main argument.

4. Some Mechanistic Views of Introspection

My introspective awareness that I am having a visual experience as of white stands in some relation to that visual experience itself. There are many views of introspection on which some causal mechanism takes the first-order state as input and delivers the introspective state as output. On these views, some sort of temporally extended causal chain leads from the first-order state to the introspective state.

Armstrong's and Lycan's ''Inner Sense Model'' of introspection is one of these views. According to Lycan, ''introspection is the operation of an internal attention mechanism that monitors experiences and produces second-order representations of their properties'' (2003, p. 26). These second-order representations are importantly similar to ordinary *perceptions*, and thus this view has become known as the ''Higher-Order Perception'' (HOP) view of introspection. Lycan says introspection makes us aware of our experiences and their properties, as perception makes us aware of

external objects (like bottle rockets) and their properties. That is, introspection is a mechanism that delivers second-order perceptions that there is an experience that there is such and such, just as visual perception is a mechanism that delivers first-order perceptions that there is such and such. For my purposes, it is important to note only that (*i*) if I am a brain, I cannot be certain that introspection does not work this way, and (*ii*) on this view introspection is a temporally extended causal chain—mediated by this internal attention mechanism—leading from experiences to distinct second-order representations of their properties.

Consider now a "Higher-Order Thought" (HOT) view of introspective awareness advocated by Nichols and Stich (2003), Rosenthal (2004), and Shoemaker (1994), among many others. On this view, introspection is the process by which we come to have non-perceptual second-order self-attributions about our first-order mental states. According to Shoemaker, these second-order states are beliefs, and the brain state associated with the first-order mental state *causes* the brain state associated with the belief about it. According to Rosenthal, a mental state is conscious only if it is accompanied by a distinct, occurrent HOT. Right now, I am conscious that there is a white surface before me: the visual representation is accompanied by that HOT. In introspection we become conscious of our consciousness; that HOT itself comes to have—via a causal process extended in time—an accompanying HOT. I become conscious that I am conscious that there is a white surface before me.

A variant HOT theory was put to me by Michael Tye (personal communication, November 2006), though I do not know how seriously he takes it. On what we may call a "Read-Write Model," there is a consciousness-compartment (C-box) in the mind, in addition to a belief-compartment (B-box). When one's experience represents that *p*, the sentence 'p' (in the language of thought) is inscribed in the C-box. Introspection is a mechanism and one of its jobs is to read sentences inscribed in the C-box and write corresponding sentences in the belief-box, e.g., "I am aware that *p*." For my purposes, it is important to note only that (i) if I am a brain, I cannot be certain that introspection does not work this way, and (ii) on all these variations of HOT theory, introspection is a temporally extended causal chain leading from some type of first-order state to a distinct second-order representation of it.

Finally, let us consider the "Same-Order Monitoring Theory" (SOMT) advocated by Kriegel (2007),[5] and by some accounts Franz Brentano. According to Kriegel (2007, p. 370), a visual experience as of green (call that mental state "M") is a "complex" of a visual representation of green (call that "M1") bundled with the awareness of M1, i.e., an appropriate representation of M1 (call this state "M2"). In virtue of being represented by M2, M1 is conscious, and M is a visual *experience* as of green rather than a mere representation. Kriegel seems to agree that the first-order state does not represent itself *as* being represented. Rather, that is what introspection does.

There are a few plausible ways introspection might work on Kriegel's view. It may be that a higher-order representation either of M2 or of M is not part of the complex M, but is brought about by some causal mechanism. Alternatively, it may be that the

complex target state M *comes to have as a constituent* a representation of M2, via some causal mechanism (Rosenthal, 2004, p. 33). For my purposes, it is important to note only that (i) if I am a brain, I cannot be certain that introspection does not work this way, and (ii) on any of these plausible SOMT views of introspection, introspection is a temporally extended causal chain leading from some type of first-order state to a distinct representation of it.

5. Support for Premise (1) in the Main Argument

Recall the first premise of the Main Argument:

(1) If the Complex View is true, then I cannot be certain that introspection is not a causal series of events extended in time.

Let me now support this premise. I take it that none of the views discussed in the previous section is obviously false, at least on the assumption that I am a brain. After all, if I suppose that I am a brain and that my mental states supervene on the physical states of that brain, then it may be that the physical states on which the mental states that constitute introspection supervene are a temporally extended causal chain. If so, it may be that the supervening mental states that constitute introspection are a temporally extended causal chain. The theories discussed in the previous section are reasonable explanations of how introspection might work, given these assumptions.

To put it somewhat more vividly, suppose you are a brain. Given that assumption, there is evidence I could give you to make it reasonable for you to believe that introspection works as, for example, the Read-Write Model suggests. Say I pop open your skull and show you the goings-on therein: if you just are that brain, and your mental life is intimately related to various events in that brain, couldn't it be that there is, say, a read-write introspective mechanism in there? On the assumption that you are a brain, you cannot rule out this theory from the armchair—this theory is clearly broadly logically possible. And the same goes for the other theories of introspection as well, each of which suggests that introspection is a temporally extended causal chain. On the assumption that you are a brain, none of these theories is a priori knowably false. You may believe that one or more are false, but this belief can't be *absolutely* certain for you. Since this is all premise (1) in the main argument claims, we should accept it.

6. Support for Premise (2) in the Main Argument

Recall the second premise of the Main Argument:

(2) If introspection is a causal series of events extended in time, then I could gain good evidence that I am feeling fierce pain right now.

Let me now support this premise. Consider first visual perception. Because visual perception involves a causal series of events extended in time—because, that is, it is

mechanistic—it is subject to radical correction, and *none* of its deliverances is certain for you. After all, it's metaphysically possible for any causal series of events to go (very) awry, and to lead to a (very) statistically abnormal, improper, or inapt result.[6] And in any case of visual perception, I could present you with evidence that would make it reasonable to believe that the causal chain has in fact gone (very) awry, and that your visual experience (really badly) misrepresents the way the world is.

For example, right now, though your visual experience represents that there is a white surface with black markings before you, there is evidence I could give you that would make it reasonable for you to believe that your mechanism of visual perception has malfunctioned, and that actually there is only a red surface with green markings before you (so the causal chain has gone awry), or to believe that actually there is no surface and no markings at all (so the causal chain has gone *very* awry). The well-rehearsed stories involve the usual suspects: malevolent neurosurgeons from Alpha Centauri, an Evil Demon, hallucinogenic drugs, etc. In general, since visual perception is mechanistic, for any possible visual experience E you may have, though E represents that the world is a certain way, it could be that the world is different from how E represents it to be, even radically different. No matter how things visually seem, we recognize the possibility that things are not as they seem. Things may even be *very* different from how they seem to me, if someone is tampering with my visual mechanism in the right way.

Why should the same inference not hold in the case of introspection, if it too is mechanistic?[7] If introspection is a causal series of events extended in time, and any causal series of events could go (very) awry, introspection is also subject to radical correction, and none of its deliverances is certain for you. On any occasion of operation, the physically-realized introspective mechanism could malfunction, and could deliver (very) false self-ascriptions, second-order beliefs, second-order perceptions, or whatever output your favored theory suggests. If introspection is only in the business of delivering the awareness or thought or perception or belief that *I am experiencing that p*, then if I come to believe that my introspective mechanism is malfunctioning, or that the causal process has gone awry, then the deliverances of that mechanism are subject to radical correction. Just as I accept the possibility that things are very different from how they visually seem (since visual perception is mechanistic), I ought to accept the possibility that things are very different from how they introspectively seem, if introspection is mechanistic.

Consider the Read-Write Model of introspection discussed above. Assuming it is operating according to a good design plan, if the mechanism is functioning properly and reads the sentence "p" in the C-box (the Consciousness-Box), it writes "I am aware that *p*" in the B-box (the Belief-Box). But it is in principle possible to manipulate the mechanism such that it is no longer functioning properly, such that it for example reads "p" in the C-box and writes "I am aware that not-*p*" in the B-box. Assuming that my brain realizes this mechanism, a sufficiently clever neurosurgeon could in principle manipulate my introspective mechanism in this way.

Therefore, according to the Read-Write Model, and on the assumption that the Complex View is true, on any occasion in which my introspective mechanism has inscribed "I am aware that not-*p*" in my B-box, I could gain evidence which would make it reasonable for me to believe that I am actually aware that *p*. It might go like this, on the assumption that the Complex View is true: first, I gain evidence that makes it reasonable to believe that I am the victim of a fiendishly clever neurosurgeon. Then, I gain evidence that makes it reasonable to believe that I have an introspective read-write mechanism, and that this neurosurgeon is causing it to malfunction in so that, though "there is fierce pain" is written in my C-box, only "I am aware that it's not the case that there is fierce pain" is written in my B-box.[8] In such an instance, it would be reasonable for me to believe that my introspective beliefs are radically false. Though it would surely introspectively seem that I am not in fierce pain, in this case I would have good reason to believe that things are not as they introspectively seem.

And so it follows that, on the Read-Write Model, I could gain evidence that would make it reasonable for me to believe that I am feeling fierce pain right now. Similar considerations apply to the other versions of HOT including Shoemaker's model,[9] the Inner Sense Model,[10] and SOMT.[11] In fact the point generalizes to any view of introspection according to which it is a causal series of events extended in time. Therefore we should accept premise (2) of the main argument. Having now supported that premise, let me move on to premise (3).

7. Support for Premise (3) in the Main Argument

Recall the third premise of the main argument:

> (3) For any propositions *p* and *q*, if (i) I cannot be certain that *p* is false, and (ii) *p* entails *q*, then I cannot be certain that *q* is false.

First, a preliminary note about "entails" as it appears in (ii). For the move from (1) and (2) to (4) to be valid, the "entails" in clause (ii) of premise (3) must refer to whatever sort of entailment relation is claimed to hold between the antecedent and consequent of premise (2). There is no algorithmic way of settling the claim made by (2), as there is, by contrast, with claims of first-order entailment. In this way, the consequence relation claimed by (2) is akin to the relation claimed by the proposition that *for any x, if x is a prime minister, x is not a prime number*. I take it that we have epistemic faculties that at least *can* deliver certainty regarding matters such as these, matters which we have no algorithmic method of settling.[12]

If so, then (3) can be proven indirectly: assuming that (3) is false results in a contradiction. To see this, suppose first that you cannot be certain that some proposition *p* is false, i.e., that your epistemic faculties cannot deliver certainty that *p* is false. Suppose further that *p* entails some other proposition *q*. (You may or may not believe that *p* entails *q*.) Now suppose that, contrary to (3), you *can* be certain that *q* is false.

It follows obviously that in such a case you at least *can* be certain that p is false. All it would take is for your epistemic faculties to deliver certainty that p entails q, and the certainty of modus tollens. You may as a matter of fact not realize that p entails q, and you may not believe that p is false. Nevertheless it is true that you *can* be certain that p is false. But then we stumble onto a contradiction. Attempting to construct an instance in which the antecedent of this conditional is true while the consequent is false results in absurdity. So we should accept that (3) is true.

Consider also the following proposition, which is logically equivalent to (3)[13]:

> (3*) For any propositions p and q, if (i) I can be certain that p is true, and (ii) p entails q, then I can be certain that q is true.

Think about an instance in which the antecedent is true: for some p and q, p entails q and your epistemic faculties at least *can* deliver certainty that p is true. In this case, you cannot be certain that q is true only if your epistemic faculties cannot even in principle deliver certainty that p entails q, or the certainty of modus ponens. Yet surely you at least *can* be certain of those things. So we should accept (3*) and its equivalent: (3) itself.

8. What to Do with (4) in the Main Argument

Premise (4) follows from premises (1)–(3):

> (4) So if the Complex View is true, then I cannot be certain that I could not gain good evidence that I am feeling fierce pain right now.

What (4) tells me, substantially, is that either the Complex View is false or I can't be certain that my belief that I am not experiencing fierce pain right now is indefeasible. I cannot rationally deny both of these; at least one is true. Which option I take should be determined by which I find more credible. If I find the antecedent of (4) more credible than the negation of the consequent, I should run a modus ponens. If on the other hand I find the negation of the consequent more credible than the antecedent, I should run a modus tollens. If I find the antecedent and the negation of the consequent equally credible, I ought to remain agnostic.

For what it's worth, I am strongly inclined to deny the consequent. There is very little I find more credible than that I can be certain that my belief that I am not experiencing fierce pain right now is indefeasible. Though I'm occasionally bothered by skeptical arguments aimed at every bit of my knowledge of the external world, I'm never bothered by parallel skeptical arguments aimed at every bit of my knowledge of the inner world, so to speak. I am supremely confident that I could not receive any compelling evidence that, contrary to appearances, I really am experiencing fierce pain right now. My experience may change—I may suddenly step in a rusty bear trap, for example—but given my current experience, surely nothing could defeat my belief that I am not experiencing fierce pain. I am far more confident of this than I am of the suggestion that, for example, I am a brain and my

mental life is (or is a product of) the interaction of (some of) that brain's parts. And so I have a powerful argument against the Complex View, and in favor of the Simple View. What certainty teaches me, then, is that I am simple. Perhaps the same goes for you, the reader.

9. Objections

9.1. Tu Quoque

I use the main argument to support the Simple View. Some readers have suspected that I richly deserve a tu quoque response, since to them the inference in premise (1) seems equally valid in the case of the Simple View. That is, these objectors urge the plausibility of:

> (1*) If I am a *simple* thing, then I cannot be certain that introspection is not a causal series of events extended in time.

And if (1*) is plausible, then of course the main argument could be turned against the Simple View. If (1) and (1*) are equally plausible, then whatever motivation the Main Argument originally produced for the Simple View is neutralized by this revised argument.

My response is that (1*) is not plausible, or at least not as plausible at (1). The reason, ultimately, is that there is nothing about the Simple View that forces open the possibility of mechanistic introspection, while the same is not true of the Complex View. Mechanistic introspection may be ruled out from the armchair on the Simple View, but not on the Complex View.

Let me show you how. First, we'll suppose that the Simple View is true. The Simple View does not have much content—it's just the denial of the Complex View. Now, our best bet to rule out the possibility of mechanistic introspection is, I believe, by running a modus tollens on something like premise (2) in the main argument. It would go like this: clearly, if introspection were mechanistic, then my belief that I'm not in fierce pain right now would be defeasible—all I'd need is evidence that the mechanism malfunctioned. But since my belief that I'm not in fierce pain right now is clearly indefeasible, it follows that introspection is not mechanistic.

Notice well that, as we close the door on the possibility of mechanistic introspection on the Simple View, it freely glides shut. Nothing in the nature of the Simple View forces open the broadly logical possibility of mechanistic introspection. The Simple View is light on content, and so it presents no obstacle to a modus tollens on premise (2). And so, in this way, one in fact can be certain that introspection is not mechanistic, on the Simple View. Therefore, (1*) is implausible and—if the original premise (1) of the Main Argument is more plausible—the tu quoque objection fails.

Premise (1) in the main argument says that, if the Complex View is true, then we can't be certain that introspection is not mechanistic. At this point, you may be wondering whether premise (1) in the main argument really is plausible, and whether

I didn't just sneak it by you back there in section 4. After all, you might think, if it's so easy to be certain that introspection is not mechanistic on the Simple View, couldn't we likewise gain that certainty on the Complex View?

I think not, and here's why. Let's try to be certain that introspection is not mechanistic on the Complex View. First, we'll suppose that the Complex View is true. In contrast to the Simple View, the Complex View comes with heavy baggage. It entails that each of us is a complex thing—a thing with parts—and each of our mental lives is (or is a result of) the causal interaction of (some of) these parts. Now, again, our best bet to rule out the possibility of mechanistic introspection is by running a modus tollens on something like premise (2) in the main argument. But notice here that, as we try to shut the door on the possibility of mechanistic introspection, we encounter some firm resistance: the nature of the Complex View gets in the way, insisting on the possibility of mechanistic introspection. For consider the set of possible worlds in which we are complexes like brains, and our mental lives supervene on the causal interaction of our parts. Isn't it obvious that, within this set of worlds, there are worlds in which introspection is a causal series of events extended in time? Consider also that, on the Complex View, I already know that other of my physically-realized belief-producing processes *are* mechanistic, e.g., perception. My introspective beliefs *may* result from a similar mechanism in a brain, if I just am a brain. We cannot shut the door on that broadly logical possibility from the armchair, given the hypothesis that I am a complex object like a brain.

In sum: aided by something like the main argument, one can come to see clearly that mechanistic introspection is impossible, in a way perfectly compatible with the Simple View. And so (1*) is not plausible. On the Complex View, however, mechanistic introspection clearly seems broadly logically possible, and so it cannot be ruled out from the armchair. And so (1) is quite plausible. Therefore, since (1*) is less compelling than (1); the views are not on a par, contrary to what the tu quoque objection insists. Reflection on introspective certainty provides, therefore, a compelling argument against the Complex View, an argument that cannot return the favor to the Simple View.

9.2. Constitution and Incorrigibility

It has been suggested in the philosophical literature that the phenomenal character of a subject's experience is "taken up" into her corresponding introspective beliefs—that some of those very abstract objects that constitute the representational content of her visual experience also (at least partially) constitute the representational content of her introspective belief about that experience. Some say phenomenal concepts are "quotational": they are said to "include" or "contain" the very phenomenal properties they refer to (see, for example, Block, 2006; Chalmers, 2003).

Assuming sense can be given to its metaphors, such a theory would presumably secure an incorrigibility thesis along the lines of the one discussed in Jackson (1973): that *S believes at t that he is in pain at t* via introspection broadly logically

guarantees that *S is in pain at t*. Similarly, that *the Statue of Liberty is on the pedestal at t* broadly logically entails that *matter is on the pedestal at t*. And, importantly, introspection could work this way and yet be a temporally extended causal chain of events.

And so an objector might urge that premise (2) is false, saying "look, here's an account of introspection according to which it is a temporally extended causal chain, and yet according to which one may be certain about some introspective beliefs. Given the constitutive relation between the experience and the belief, the subject's belief has a very high epistemic status—the belief couldn't be false. So this belief may rightly be said to be *certain* for the subject. No evidence would make it reasonable for her to believe that she is in fierce pain."

The objection includes something like the following steps:

> *Constitution*: introspective beliefs are at least partly constituted by the states they are about.

Therefore,

> *Incorrigibility*: that S introspectively believes she is in phenomenal state P at time t broadly logically guarantees that S is in P at t.

Therefore,

> *Certainty*: we have an explanation of the certainty of at least some introspective beliefs.

And,

> *Compatibility*: this explanation is compatible with mechanistic introspection.

Therefore,

> *Counterexample*: premise (2) in the Main Argument is false.

In what follows, I will give three critical responses to this objection.

My first response to this objection is that the inference from Certainty and Compatibility to Counterexample is invalid. Even if we grant that this constitution story is compatible with mechanistic introspection, and that it can explain the certainty of introspective beliefs like "I am aware that there's fierce pain," how does the story go with respect to what we may call "negative" introspective beliefs, such as "I am aware that there's *no* fierce pain?" Right now, I have that belief and it is true. Yet it cannot be that some constituent of my experience is "taken up" into the introspective belief, since the introspective belief accurately represents that a certain phenomenal quality, namely fierce pain, is *absent* from the content of my experience.[14] So the objector has not yet provided an account according to which introspection is mechanistic and yet I can be certain that I am aware that there's *no* fierce pain. And so premise (2) is unchallenged, since the proposed constitution story does not apply to negative introspective beliefs. Indeed, my main argument intentionally concerns a negative introspection belief in order to sidestep just such an objection.

My second response is that a theory that entails anything like Incorrigibility has highly implausible consequences. Such a theory would entail, for example, that the following case is impossible:

> *Paint Store*: I am at a paint store, looking at samples. I hold a maroon color sample in the center of my visual field, which thereby tokens only maroon. Nevertheless, I misidentify the color and believe that I am visually experiencing scarlet in the center of my visual field.

Surely this story is coherent.[15] (My wife testifies that this is a common occurrence in my own life.) Yet, according to this constitution theory, that *I introspectively believe that I am experiencing scarlet in the center of my visual field* broadly logically guarantees that *I am experiencing scarlet in the center of my visual field*. So, if this constitution theory is right, Paint Store is incoherent. Yet since Paint Store is coherent, we should reject this constitution theory. So this theory cannot offer us an objection to the main argument.[16] The objection stalls at the first steps: Incorrigibility above is false, and therefore so is Constitution, and so again the objection does not reach the conclusion that premise (2) in the main argument is false.

My third and final response is that nothing in the neighborhood of Incorrigibility is sufficient to explain the kind of epistemic certainty that attends some of our introspective beliefs. That is, for example, the inference from *if S believes she's in pain at t then she is in pain at t* to the conclusion that *S is certain that she's in pain at t* is invalid, and so the move from Incorrigibility to Certainty above fails. For consider that the number of hairs on your head is either even or odd. Suppose it's even here in the actual world, @. Given this supposition, it's true that if you believe the number of hairs on your head is even in @, then the number of hairs on your head is even in @. And that conditional is necessarily true, given that the proposition in question is a true world-indexed proposition. Across modal space, any creature with that belief believes truly; we have here necessary reliability and, indeed, incorrigibility.

And yet, despite the incorrigibility of this belief, it scarcely follows that this belief would be certain for you, were you to hold it. You wouldn't be entitled to believe it, come what may. You lack any grounds at all for that belief, let alone the absolutely indefeasible grounds required for certainty. Evidently, the sort of incorrigibility thesis that the constitution story secures has little or nothing to do with justification and certainty. It follows, therefore, that even if an incorrigibility thesis in this neighborhood were true—and I argued above that it isn't—it wouldn't by itself be enough to explain how I might be certain that I'm not in fierce pain right now, even if introspection is mechanistic. Since the move from Incorrigibility to Certainty above is invalid, the objection to premise (2) in the main argument again fails.

9.3. Constitution and Self-Intimation

In the previous section, we considered a theory of introspection on which introspective beliefs are partly constituted by the first-order states they are about.

Such a theory would secure an incorrigibility thesis, as I said. But one might also wonder whether the constitution relation holds in the other direction. That is, one might wonder whether first-order phenomenal states like fierce pains are themselves partly composed of introspective awareness or belief. This view is not unprecedented in the literature,[17] and it would secure an intuitively attractive self-intimation thesis: necessarily, if a subject is in fierce pain, then she's aware that she is.[18]

It's not immediately obvious how such a self-intimation thesis might challenge the main argument. Exactly which premise(s) would it call into question? It holds most promise of providing a counterexample to premise (2) in the main argument: an account of mechanistic introspection on which, nevertheless, my belief that I'm not in fierce pain right now is indefeasible.

The objector may reason this way: "listen, on this view I just sketched for you, it's obvious that, necessarily, if a subject is in fierce pain, she's aware that she is (or at least she would be, were she to introspect). Since by introspection you can be certain that you don't believe that you're in fierce pain, you can now run a modus tollens with that self-intimation thesis and conclude—with certainty—that you're not in fierce pain. This account of introspection secures the self-intimation thesis and thereby explains the certainty of your belief that you're not in fierce pain, all the while being compatible with introspection's being mechanistic. Therefore, premise (2) in the main argument is false."

This is a clever and interesting objection to the main argument. However, it passes the buck in a way that renders it, in the end, unsuccessful. Let's retrace the dialectic. Initially, I was wondering how I might be certain that I'm not in fierce pain, if introspection is mechanistic. The objection now under consideration begins with this advice: "in order to be certain that you're not in fierce pain, start by being certain that you don't believe you're in fierce pain." But, of course, my initial concern will reemerge at this higher level—the problem has been merely kicked upstairs, and the main argument can be redeployed against this new introspective belief. For how might I be certain that I don't believe that I'm in fierce pain, if introspection is mechanistic? The same considerations about the fragility of causal process that first led me to worry about the epistemic status of my introspective belief that I'm not in fierce pain apply just as strongly to my introspective belief that I don't believe that I'm in fierce pain. Sure, it *seems* like I don't believe that I'm in fierce pain, but there remains a possibility—remote, to be sure, but real nonetheless—that my introspective mechanism is malfunctioning, delivering the belief that I don't believe I'm in fierce pain, when I really *do* believe that. Therefore, the objection has not yet explained how I may be certain that I'm not in fierce pain right now, if introspection is mechanistic.

In other words, the objection promises me certainty that I'm not in fierce pain, but only if I first gain certainty that I don't believe I am. Yet the objection does not explain how I might be certain that I don't believe I am in fierce pain. In this way, the objection passes the buck in an unsatisfactory way; it writes me a check that I cannot cash. So, ultimately, the objection does not succeed in explaining how I might be certain that I'm not in fierce pain even if introspection is mechanistic. And so, in the

end, it does not succeed in providing a counterexample to premise (2) in the main argument.

Secondly, this objection has it that my justification for believing that I'm not in fierce pain right now is *inferential*, and in addition that this inference crucially depends on my knowledge of the self-intimation thesis. This strikes me as implausible on both counts: first, many people lack a belief in the self-intimation thesis, and yet are nevertheless certain—entitled to believe, come what may—that they're not in fierce pain right now; children, for example. Second, even supposing that a person grasps and believes the self-intimation thesis, he may lack the cognitive resources to perform the somewhat complicated inference from that belief, and his introspective belief that he doesn't believe he's in fierce pain, to the conclusion that he's not in fierce pain. And yet, nevertheless, such a person may be certain that he's not in fierce pain; his grounds may make it such that no additional evidence should lower his credence. Children, again, serve as examples here. Since this objection has these two implausible consequences, we should again reject it as ultimately unsuccessful.

10. Conclusion

In this paper, I have developed and defended a novel argument against the Complex View of human persons. Philosophers who wish to maintain the standard materialist account of human persons on which introspection is mechanistic must square their shoulders and accept the unpalatable consequence that introspection is hyperfallible, i.e., that we can never be certain of any introspective belief. For the rest of us: introspection is not hyperfallible, and so introspection is not mechanistic. And surely at least some of our introspective beliefs are immune to defeat. This could not be so on the Complex View. Therefore, certainty teaches us that the Complex View is false.

Acknowledgments

For helpful conversations and valuable feedback, many thanks to Nathan Ballantyne, Adam Pautz, Tim Pickavance, Ted Sider, David Sosa, Michael Tye, and an anonymous referee of this journal.

Notes

[1] See, for example, Chisholm (1991) and, more recently, Barnett (2010). Earlier in his career, Chisholm (1978) took seriously the possibility that we are material simples, though he stopped short of endorsing the view.

[2] A proposition *p* is certain for a subject S just in case S is entitled to believe *p*, come what may. That is, S's grounds for *p* make it such that no additional evidence should lower her credence in *p*. By "*p* is certain for S" I do *not* mean "it is psychologically impossible for S to doubt that *p*." Rather, I mean the normative notion "S cannot rationally doubt *p*." I mean what many have called "absolute" or "Cartesian" certainty. When

a subject has this kind of certainty, her belief is often said to be "Demon-proof" after Descartes' *deus deceptor*.

[3] There are many ways a subject might come to have beliefs about the phenomenal character of her own experiences. Introspection is that way to which the subject has privileged access (normally, at least).

[4] That is, I could gain evidence that would make it reasonable for me to believe that I am experiencing fierce pain, even though I continue to token just this type of experience, which introspectively seems to me not to involve any pain at all.

[5] What I say here is also applicable mutatis mutandis to the "Intrinsic Higher-Order Thought Theory" advocated by Genarro (1996) and Natsoulas (1996).

[6] Tooley (2008), I think, would agree. He says: "assuming that at least some of the basic causal laws of our world are probabilistic, any physical structure is capable of not functioning properly, and so any capacities based on a physical structure could always fail" (Tooley & Plantinga, 2008, p. 97).

[7] Armstrong seems to think it does: "I shall defend the thesis ... that mental states are ... states of the brain. Now if I accept the existence of introspection, as I also do, then I must conceive of both introspection and the objects of introspection as states of the brain. Introspection must be a self-scanning process in the brain. That it is logically possible that such a self-scanning process will yield wrong results is at once clear" (1963, pp. 418–419).

[8] If one is concerned with immaterialist versions of the Complex View, the evidence gained here could be about an Evil Demon rather than a neurosurgeon.

[9] On Shoemaker's view, it may be that the experience of pain is such that, *in certain circumstances*, it necessarily causes the second-order self-attribution *I am aware that there's pain*, and it may be such that this second-order self-attribution is such that, *in the absence of malfunction*, it is caused by the first-order state. But a sufficiently clever neurosurgeon could manipulate one's brain to produce malfunction, to produce a circumstance that is not one of those in which the experience of pain necessarily causes the second-order self-attribution. Thus the neurosurgeon could manipulate my brain such that, though my experience represents that *p*, I form via introspection the belief that "I am not aware that *p*."

[10] When the Inner Sense Model's internal scanner is functioning properly (assuming a good design plan), if my first-order experiential state has the quale P, then the scanner will produce a second-order representation which, while not *itself* having quale P, represents that the first-order state has quale P. However, it is in principle possible to manipulate the mechanism such that, even though my first-order experiential state has the quale P, the second-order representation produced by the scanner represents that the first-order state does *not* have quale P.

[11] However introspection works on Kriegel's view, if the introspective mechanism is functioning properly (assuming a good design plan), if I am visually aware that *p*, my introspective mechanism will produce a representation in virtue of which I am introspectively aware that I am visually aware that *p*. However, this mechanism may be manipulated such that, even though I am visually aware that *p*, it produces a representation in virtue of which I am introspectively aware that I am *not* visually aware that *p*.

[12] Just to be crystal clear, "entails" as it appears in (3) is not equivalent to mere material implication. As we know, material implication is a sad model of genuine entailment.

[13] The contrapositive of (3) is this: For any *p* and *q*, if I can be certain that *q* is false, then either <it's false that *p* entails *q*> or <I can be certain that *p* is false>. Now let *p* represent *that q is false* and let *q* represent *that p is false*. (3*) is now obviously equivalent.

[14] Pollock (1986, pp. 32–33) discusses this problem for constitution theories, i.e., theories which endorse what he calls the "Containment Thesis."

[15] Goldman agrees: ''if one is inattentive, drugged, or otherwise imbalanced, he can misapply concepts to his own experiences and generate false beliefs about them'' (2004, p. 282). Goldman there also cites a case from Pollock, saying ''people typically and inattentively describe ways shadows appear on snow as gray, since they simply assume that shadows are gray, when in fact they appear blue'' (Pollock, 2001, p. 43).

[16] Perhaps a version of the constitution theory survives, restricted to introspective beliefs like ''I am aware of *this*,'' or the kind of beliefs recently discussed by Horgan and Kriegel (2007), or Chalmers' (2003) ''direct phenomenal beliefs.'' But none of these theories furnishes an objection to the main argument, since none of them entails that my introspective belief that ''I am not experiencing fierce pain right now'' is incorrigible. Such theories explain, at most, the introspective certainty of some conceptually stripped-down, relatively content-less introspective beliefs. And so they won't help explain the certainty of an introspective belief like ''I am not experiencing fierce pain right now,'' which has substantially more conceptual content. It features the concept PAIN, for example, which none of the beliefs that concern Horgan and Kriegel or Chalmers do.

[17] Views like this are discussed in Weatherson (2004, p. 379) as well as Horgan and Kriegel (2007). Shoemaker thinks that at least some phenomenal states are ''constitutively self-intimating'' (1990), saying ''it is of the essence of a state's having a certain phenomenal character that this issues in the subject's being introspectively aware of that character, or does so if the subject reflects'' (2001, p. 247) I'm concerned with this type of view in the text. Yet in other places, Shoemaker is careful to add the qualification that self-intimation doesn't occur with broadly logical necessity, but only under normal conditions, or absent malfunction. Such a qualified view would presumably be much less helpful to our objector here, since one's belief that one is properly functioning is far from certain. For that reason, I don't discuss this weaker, more qualified view here in the text.

[18] Or at least she would be, were she to introspect. I thank an anonymous referee for this journal for urging me to consider a constitution theory along these lines, and its implications for my main argument.

References

Armstrong, D. M. (1963). Is introspective knowledge incorrigible? *The Philosophical Review*, *72*, 417–432.

Baier, K. (1962). Smart on sensations. *Australasian Journal of Philosophy*, *40*, 57–68.

Barnett, D. (2010). You are simple. In G. Bealer & R. Koons (Eds.), *The waning of materialism* (pp. 161–174). Oxford: Oxford University Press.

Block, N. (2006). Max Black's objection to mind-body identity. In D. Zimmerman (Ed.), *Oxford studies in metaphysics* (Vol. 2, pp. 3–78). Oxford: Oxford University Press.

Byrne, A. (2005). Introspection. *Philosophical Topics*, *33*, 79–104.

Chalmers, D. (2003). The content and epistemology of phenomenal belief. In Q. Smith & A. Jokic (Eds.), *Consciousness: New philosophical perspectives* (pp. 220–272). Oxford: Oxford University Press.

Chisholm, R. (1978). Is there a mind-body problem? *Philosophical Exchange*, *2*, 24–34.

Chisholm, R. (1991). On the simplicity of the soul. *Philosophical Perspectives*, *5*, 157–181.

Gennaro, R. J. (1996). *Consciousness and self-consciousness*. Philadelphia, PA: John Benjamin Publishers.

Goldman, A. H. (2004). Epistemological foundations: Can experiences justify beliefs? *American Philosophical Quarterly*, *41*, 273–285.

Horgan, T., & Kriegel, U. (2007). Phenomenal epistemology: What is consciousness that we may know it so well? *Philosophical Issues*, *17*, 123–144.

Jackson, F. (1973). Is there a good argument against the incorrigibility thesis? *Australasian Journal of Philosophy*, *51*, 51–62.

Kriegel, U. (2007). The same-order monitoring theory of consciousness. *Synthesis Philosophica*, *44*, 361–384.

Lycan, W. (2003). Dretske's ways of introspecting. In B. Gertler (Ed.), *Privileged access* (pp. 15–30). Burlington, VT: Ashgate.

Natsoulas, T. (1996). The case for intrinsic theory: I. An introduction. *Journal of Mind and Behavior*, *17*, 267–286.

Nichols, S., & Stich, S. (2003). How to read your own mind: A cognitive theory of self-consciousness. In Q. Smith & A. Jokic (Eds.), *Consciousness: New philosophical perspectives* (pp. 157–200). Oxford: Oxford University Press.

Pollock, J. (1986). *Contemporary theories of knowledge*. Lanham, MD: Rowman and Littlefield.

Pollock, J. (2001). Nondoxastic foundationalism. In M. DePaul (Ed.), *Resurrecting old-fashioned foundationalism* (pp. 41–58). Lanham, MD: Rowman and Littlefield.

Rosenthal, D. (2004). Varieties of higher-order theory. In R. J. Gennaro (Ed.), *Higher-order theories of consciousness* (pp. 19–44). Philadelphia, PA: John Benjamin Publishers.

Shoemaker, S. (1990). First-person access. *Philosophical Perspectives*, *4*, 187–214.

Shoemaker, S. (1994). Self-knowledge and ''inner sense''. *Philosophy and Phenomenological Research*, *54*, 249–314.

Shoemaker, S. (2001). Introspection and phenomenal character. *Philosophical Topics*, *28*, 247–273.

Tooley, M., & Plantinga, A. (2008). *Knowledge of God*. Malden, MA: Blackwell.

Weatherson, B. (2004). Luminous margins. *Australasian Journal of Philosophy*, *82*, 373–383.