

Assertion and Transparent Self-Knowledge

Eric Marcus
Auburn University
marcuea@auburn.edu

John Schwenkler
Florida State University
jschwenkler@fsu.edu

1. Belief is a psychological state or attitude, and thus for a person to know what she believes is for her to know herself to be in a certain state, or to have a certain attitude toward how things are. Doxastic self-knowledge is therefore knowledge of psychological reality. And it is natural to conclude from this that this knowledge must be justified through an empirical process, i.e. ‘a process in which one observes or detects one’s own mental states’ (Gertler 2011a, 255). According to this natural line of thought, just as I can only come to know what you believe by watching your deliberate actions, observing how you respond automatically to certain stimuli, and keeping track of what you assert, so in order for me to know what *I* believe my doxastic states must somehow be made manifest to me.

Our aim in this paper is to challenge this last assumption. We refer to the position in question as the *empirical* conception of doxastic self-knowledge, since according to it knowledge of one’s beliefs, if it is to be knowledge at all, must be a kind of empirical knowledge, or knowledge that is grounded in evidence.¹ On the empirical conception, what accounts for the special character of doxastic self-knowledge is not that this knowledge is non-empirical, but rather the *kind* of empirical grounding it possesses. In particular, self-knowledge is thought to be different from knowledge of other minds because only in one’s own case does one have direct or ‘first-personal’ access to thoughts, feelings, and urges that are manifested only ambiguously in one’s overt behavior. The distinctively first-personal character of this evidence, plus the fact that it is (allegedly) abundant and (again, allegedly) less subject to misinterpretation, is supposed to put a person in a position to know her own beliefs in a way that no one else can. Even if this knowledge is not exhaustive or infallible, the empirical conception is supposed to account for how a person usually knows *more* about her beliefs than anyone else can, and how this self-knowledge is usually *more secure* than anyone else’s knowledge of

¹ What we call the *empirical* conception is equivalent to what Gertler (2011a) calls *empiricism* about self-knowledge. We prefer our terminology because we deny that what Gertler calls ‘rationalism’ is the only alternative to it. We wish to emphasize that in rejecting the empirical conception we do not deny that *what is known* (viz., the thinker’s belief) in an instance of doxastic self-knowledge is an empirical fact. Rather, our claim is that the *process by which* such facts are known is not empirical in the sense just spelled out.

her beliefs will be. Nevertheless, according to the empirical conception, knowledge of one's own beliefs is similar to knowledge of the beliefs of others insofar as both are justified through evidence. It is, on the empirical conception, the difference in the *character* and *extent* of the evidence available to the believer herself that is the ground of first-person authority.²

Our paper will argue against this position by exploring the connection between self-knowledge and assertion. The empirical conception is false, we will argue, because it cannot account for the way that a person is ordinarily able to express her beliefs by asserting their content. Honest assertion is *speaking one's mind*. And this requires knowledge that one is in the state of mind that one speaks from. But speaking one's mind is not speaking *about* one's mind, and so evidence *that* one believes something cannot put one in a position honestly to assert *what* one believes. Rather, it is a condition of honest assertion that a person's grasp of the belief she asserts be grounded non-empirically, i.e. not in *any* sort of evidence at all.

2. Contemporary discussions of assertion often focus on the knowledge norm:

(KA) Assert that *p* only if you know that *p*.

² As is often the case with widespread assumptions, the thesis that doxastic self-knowledge is a kind of empirical knowledge is not usually formulated or defended as such in the philosophical literature. But looking closely one can see how this empirical conception is shared ground among a range of otherwise differing accounts of first-person authority. Gilbert Ryle's behaviorist view in *The Concept of Mind* (1949) and D.M. Armstrong's account of introspection in *Belief, Truth and Knowledge* (1973) are cases in point: while Armstrong grounds self-knowledge in a dedicated faculty through which beliefs and other mental items are brought to consciousness, and Ryle denies that there is any such faculty, both assume that there is *some* way in which self-knowledge results from tracking sources of information (or, as Ryle puts it, 'data') about one's states of mind. More recently, Peter Carruthers joins Ryle in rejecting any appeal to a faculty of direct introspection, but on his *interpretive* account of self-knowledge the reliability of self-attributions is due in part to the fact that 'we almost always have a great deal more evidence available to us when we interpret ourselves than when we interpret others' (2010, 105). And Quassim Cassam has recently advanced an account on which self-knowledge is a species of *inferential* knowledge: 'In the most straightforward case,' he writes, 'you know that you have [an attitude] A insofar as you have access to evidence that you have A and you infer from your evidence that you have A' (2016, 138). Despite their differences, all of these accounts conceive of doxastic self-knowledge as justified by an empirical process in the sense we have defined.

Simplifying somewhat, there are two ways of failing to conform to (KA). One is by asserting something you merely believe but do not know, which is epistemically irresponsible.³ The other is by asserting something that you don't even believe, which is dishonest.⁴ Philosophers discussing the norms of assertion have typically focused on the requirement of epistemic responsibility—for example, they have considered whether acceptable assertion really does require knowledge, or simply justification, and so on (see, e.g., Williamson [2000], Weiner [2005], and Kvanvig [2009]). Our concern, however, is not with this, but solely with what is required for assertion to be honest, and thus with conditions of satisfying the following norm:

(BA) Assert that *p* only if you believe that *p*.

We grant for the sake of argument that as with most (if not all) norms it might sometimes be best, all things considered, to violate (BA). If this is true it does not show that (BA) is not a norm governing honesty, but only that it might sometimes be best to be dishonest, i.e. to assert something that one does not believe—as if, say, the Nazi at the door asks whether there are any Jews in your attic. Thus we put (BA) forward only as an articulation of the norm governing assertion where the requirement to be honest is not trumped by other considerations.⁵ The rule is not intended as a guide to acceptable assertion in general. It captures *part* of our ordinary concept of *one* important feature that assertion often aims at, namely honesty.

It is because assertion is an action, something that a person *does* rather than something that merely happens, that norms for assertion take the form of *rules* that one who asserts can choose whether or not to follow. This is why (BA) is formulated in the imperative mood: it tells a person what *to do* in order to avoid asserting dishonestly. And this is the first step toward seeing the connection between honesty and doxastic self-knowledge. For in order to hold oneself to the rule (BA)—to *obey* this imperative, to *follow* the rule as opposed to merely conforming to it—one who asserts must understand whether or not the conditions it specifies are satisfied. And this shows that honest assertion requires a grasp of one's own beliefs: a person who

³ Goldberg describes such an act as 'reckless'. See Goldberg (2015, ch. 6.)

⁴ Things will be more complicated if knowledge does not require belief (see, e.g., Schwitzgebel and Myers-Shulz [2013]). Still it seems that on such a view there will be some attitude short of knowledge—if not belief, then perhaps something like acceptance in a context—that suffices for honest assertion, and our arguments below will apply in turn to such a requirement.

⁵ On the defeasibility of norms of assertion, see Kvanvig (2009).

was 'self-blind' to her doxastic attitudes would be in no position to follow (BA), since she could not tell whether she believed what she was asserting.

We can give further support to this preliminary conclusion by reflecting on a hypothetical case. Begin by imagining a man to whom it is vitally important that he believe that he has lived an enviable life. However, 'deep down' the man does not believe this, but rather believes that he spent much of his childhood living in fear of his father's anger and disapproval. This belief overwhelms him with shame, and it is precisely this shame that makes it so important for the man to believe that his life has been wonderful. He has for a long time managed to avoid thinking about the horror of his childhood, repressing his uncomfortable beliefs in a way that suits the false self-conception to which he clings—a conception according to which, contrary to fact, he is really very happy with his childhood.⁶ Yet this repression has not changed the man's actual beliefs about his childhood; it is only a way of keeping himself from bringing them to mind. When he reflects on questions that bear on the quality of his life-experience—even the question of whether he had a happy childhood—his answers reflect his distorted self-conception, rather than his repressed beliefs. So although this man believes that he had an unhappy childhood, this belief lies outside his ken.

So far described, the situation of the man we have just imagined constitutes the **Basic Case** whose details we will vary in several ways as our argument proceeds. In our first variant on that case, suppose that this man finds himself listening to a friend bemoan her very unhappy childhood. This testimony might have brought to mind the man's belief that his own childhood was unhappy as well, but instead it prompts him to tell himself how very happy he is with the way his life has gone. Despite this, in an effort to appear sympathetic to his friend the man says to her: 'I too had an unhappy childhood'. Call this the case of the **Would-Be Liar**. The essential thing to see is that in speaking as he does, the *Would-Be Liar asserts what he believes*—since as we described the Basic Case this person does in fact believe that he had an unhappy childhood. Clearly, however, the *Would-Be Liar* does not speak honestly when he asserts this, since his purpose in speaking is to mislead his friend as to what his own life was like. The *Would-Be Liar* has not *followed* the rule (BA), despite having acted in conformity with it. This example shows that an assertion is not honest just because what the speaker asserts is what she in fact believes. In order to speak honestly, one's assertion and one's belief must not correspond merely by accident.

The case of the *Would-Be Liar* brings into focus the question that we will consider in more detail below: What is required, beyond merely having a belief in

⁶ Whether this 'managing' involves intentional action, subpersonal mechanisms, or some combination of the two need not concern us here.

what one asserts, for an assertion to be honest? In particular, what sort of *grasp* of one's belief does honest assertion require?

3. It is tempting to suggest that the reason it is not honest for the Would-Be Liar to say that he had an unhappy childhood, even though he does in fact believe this, is simply that he is ignorant of his belief—that is, the problem is simply that he does not *believe* that he has this belief. Let us suppose that this is so far correct: the Would-Be Liar does not have such a second-order belief. And now let us ask: Would *believing that he believes* that his childhood was unhappy be enough for the Would-Be Liar to assert honestly that this was so?

Not necessarily. Consider the **Lucky Spiritualist**, who as in the Basic Case has a repressed belief that he had an unhappy childhood. The Lucky Spiritualist visits his psychic, who tells the Spiritualist that he (the Spiritualist) believes his childhood was unhappy. And now suppose that *solely for this reason*, the Spiritualist ascribes this belief to himself. Unlike the Would-Be Liar, the Lucky Spiritualist believes that he believes that he had an unhappy childhood. Despite this, it seems clear that if the Lucky Spiritualist were to tell someone that his childhood was unhappy, this still would not be an honest assertion. For the most that the psychic's testimony could put the Spiritualist in a position to assert honestly (if irresponsibly) is exactly what she tells him, namely that he *believes* that he had an unhappy childhood. This is because the psychic tells the Spiritualist nothing about his childhood itself, but only about his state of mind towards it—whereas it is his *childhood* that the man would be describing if he said that his childhood was unhappy. Because of this, even though the Spiritualist meets the condition just stipulated—he not only believes, but also believes that he believes, that he had an unhappy childhood—he still cannot honestly assert this.

Nor will it close the gap if we suppose that the man's belief about his belief is grounded in something more reliable than the testimony of a psychic. Consider the **Faithful Patient**, who is otherwise just like the Lucky Spiritualist, but is told not by his psychic but by his highly skilled therapist that he (the Patient) believes that he had an unhappy childhood. Suppose the Faithful Patient ascribes this belief to himself on the strength of the therapist's testimony. We can imagine that the therapist possesses excellent evidence about the Patient's belief as it is manifested in his unguarded behaviors, so that in virtue of this diagnosis the Faithful Patient *knows* that he believes that his childhood was unhappy. Nevertheless the Faithful Patient cannot, just on the strength of this testimony, honestly assert what he knows himself to believe, *viz., that his childhood was unhappy*. For no matter the quality of the evidence provided by the therapist's diagnosis of the Patient's beliefs about his childhood, if the Patient knows of those beliefs only in *this way* he is not in a position to assert honestly *that things are* as he believes them to be. There is, then,

more to the norm of honesty than the requirement that one assert only something that one knows oneself to believe.

4. At this point it is worth offering a preliminary diagnosis of why it is that the Lucky Spiritualist or Faithful Patient cannot honestly assert that he had an unhappy childhood. To do this, let us draw on a pair of cases described by Jonathan Dancy (2000, 125), which he in turn credits to John Hyman. First there is Zoe, who calls the zookeeper because she believes there is an elephant in her bathtub. Zoe's action makes sense if we see it as the action of someone who believes that an elephant is in her bathtub and then acts, as we will put it, *from the point of view* of this belief—or through its lens, so to speak. Zoe calls the zookeeper in light of what she believes, namely that there is an elephant in her bathtub. And it is this (putative) fact that she takes to justify her action—an action that is an attempt to do something that will get rid of the elephant.

In contrast to Zoe's case consider also that of Silas, who believes that there is an elephant in his bathtub and therefore places a call, not to the zookeeper, but rather to his trusted psychiatrist. Silas differs from Zoe in that the position of the supposed elephant is not at issue in what he does—for it is rather his *belief* about the elephant that he is trying to get rid of. To make sense of Silas we need to appeal to his psychotic belief but not, as with Zoe, as an attitude that characterizes the *point of view from which* he acts. We can mark this difference by saying that Silas calls the psychiatrist *in light of* his belief rather than from its point of view, or through the lens on the world that it provides. Whereas Zoe will justify her action by appeal to the believed fact that there is an elephant in the bathtub, Silas will justify his by the believed fact that he believes this.

The relevance of these cases to those of the Lucky Spiritualist and Faithful Patient should be clear. Like Silas, neither the Lucky Spiritualist and Faithful Patient can reasonably act *from the point of view* of his belief that he had an unhappy childhood. Yet this is just the kind of act that honest assertion is: in asserting that *p* one says something *about the world*, just as in calling the zookeeper Zoe is attempting to do something about the elephant. As we described their situations, the most that the Lucky Spiritualist or Faithful Patient can reasonably do is take steps to do something *about his belief*, in light of the fact *that* he believes this. These steps might include finding out whether the belief itself is true—and then, perhaps, describing that reality in turn. Yet as long as all he goes on is the testimony of a psychic or a therapist about his own attitudes, the man cannot honestly assert that his childhood *was* as he believes it to be, despite the fact that he believes or knows that he believes this.

Our task, then, is to characterize the distinctive form of self-understanding that makes it possible to speak honestly *from the point of view* of a belief, where this involves something more than knowingly saying what one actually believes.

5. Let us ask again: What more is required for honest assertion than the grasp of one's beliefs that is possessed by the Lucky Spiritualist or Faithful Patient—or, for that matter, by someone like Silas? A simple answer is that the Spiritualist and Patient grasp their beliefs only in an alienated or 'third-personal' way, i.e. a way that would put any other person in receipt of the same testimony in the same position as the Spiritualist and Patient themselves. And clearly a third party who overheard, say, the therapist's testimony about the Faithful Patient's belief could not then assert that the Patient had an unhappy childhood. Does the fact that the Faithful Patient is also a mere 'observer' of his attitude explain why he cannot honestly assert it?

There are two ways of developing this simple answer, one of which we accept while finding the other inadequate. The inadequate diagnosis is the one that proceeds from the assumption that we outlined at the start of this paper, that doxastic self-knowledge must be grounded in evidence. So long as one accepts this assumption it will seem that the trouble with the Faithful Patient is simply that the *evidence* that grounds his second-order belief is insufficiently first-personal—that is, the trouble is that his evidence is of the second- or third-personal kind, whereas properly *first-personal evidence* would do the trick. By contrast, our more radical diagnosis is that a truly first-personal grasp of what one believes is not grounded in evidence at all.

To see why we favor this more radical proposal, consider one last variant on our Basic Case, in which the role of the Faithful Patient's therapist is occupied by the subject himself, and where the process through which he comes to know his beliefs arises directly from subjective access to his own 'internal' thoughts, feelings, urges, and so on. Call this person the **Perceptive Introspector**. The Introspector notices how he becomes curiously distracted when someone asks directly about his childhood, and how he always tends to think instead of how well-prepared for success he emerged as a young adult. He also notices that although he is an enthusiastic film buff, when invited to see films such as the *400 Blows* or *Pather Panchali*, films he knows to be about difficult childhoods, he always finds himself unaccountably yearning to do something else. It strikes him that these and similar reactions reveal that he believes himself to have had an unhappy childhood, while wishing to avoid being conscious of this belief. This knowledge is reached in a way that is available to no one else, since no one else but the Introspector can have direct access to his thoughts, feelings, and inclinations. Despite this, the knowledge still will not put the Introspector in a position to assert that he had an unhappy

childhood. For the position of the Perceptive Introspector is really no different from that of Silas, the Faithful Patient, or the Lucky Spiritualist: his relation to his belief about his childhood is such that he cannot honestly speak *from its point of view*—despite the fact that this is a belief he actually holds, and knows himself to hold in a way that no one else could know this. This example reveals that even ‘first-personal’ evidence about one’s beliefs is not sufficient for honest assertion.

We will turn in §7 to consider whether honest assertion might be grounded in some other sort of first-personal evidence than that which the Perceptive Introspector draws on. Before that, in §6 we consider a different strategy for responding to our argument.

6. Thus far we have argued that neither merely believing something, nor believing that one believes it, nor knowing that one believes it, nor knowing this belief on the basis of specifically first-personal evidence, is enough to make possible the honest assertion of a belief. It might seem, however, that the missing element we have thus far failed to discover lies not in the relation between the speaker and her belief, but rather in the relation between the speaker and world. Indeed, we observed above that assertion, i.e. speaking one’s mind, is not (in general⁷) a way of speaking *about* one’s mind. When we express our beliefs through assertion, we do this by speaking about the *objects* of our beliefs, and so the grounds on which we do this must include grounds for taking these objects to be as we say them to be. But evidence *that one believes* something does not (in general⁸) provide any grounds for thinking *that things are* as one believes them to be.

This diagnosis suggests a way of rebutting our challenge to the empirical conception of self-knowledge. Perhaps the problem has been with attempting to understand the honesty of assertion simply through (BA), a norm that says nothing about the *truth* of one’s beliefs—whereas one who asserts honestly must do so out of a concern that she say what is true. Given this, one might conclude that what makes honest assertion impossible in the cases described above is only that these subjects would violate a norm against asserting falsehoods:

(TA) Assert that *p* only if *p*.

The proposal on behalf of the empirical conception would then be this. Honest assertion requires not only following (BA), which presupposes a grasp of one’s beliefs, but also following (TA), which presupposes a grasp of the fact that things are

⁷ The obvious exception is when one asserts certain sorts of things about oneself.

⁸ Here the exception will be any case where there is a special reason to believe that one’s beliefs are correct.

as one believes them. The Faithful Patient and Perceptive Introspector fail to follow (TA)—and it is *this*, rather than their empirical route to doxastic self-knowledge that explains why they cannot assert their beliefs honestly. To speak from the point of view of a belief, as honesty requires, is to speak with the truth of the believed proposition in mind.

We agree that an aim to assert only what is true is among the conditions of honest assertion. We deny, however, that the envisioned appeal to (TA) as a supplement to (BA) offers a way to defend the empirical conception. This is because, as we will argue below, a correct understanding of (TA) as giving a requirement of honest assertion reveals that following (TA), properly understood, *just is* following (BA), properly understood. Furthermore, this internal connection between belief and truth as conditions of honesty cannot be understood except in light of the thesis that, in paradigmatic cases, a person knows her own beliefs in a non-empirical manner.

To begin, notice that as (TA) is a rule, there is once again more to following it than merely conforming to it. A person who just *happens* to assert something that is true—as a variant on the Would-Be Liar perhaps might⁹—has not thereby asserted honestly, since the conformity between her assertion and the truth will have been an accident. Nor, however, is it enough to add that one who asserts honestly must not only speak the truth but also believe what she asserts, since the Would-Be Liar or any of the other characters above, who do believe what they assert, could *happen* also to speak truly without thereby being honest in their assertions. To use a common metaphor, following (TA) means asserting what is true in a way that *aims* at asserting what is true: and this ‘aiming’ will be part of the self-conscious activity of the speaker, who must *understand herself* to be speaking with an aim of being truthful. The task of articulating (TA) as a condition of honesty thus raises all of the same questions that arose in our discussion (BA). It cannot help us to resolve them. And so the proposal in question is inadequate.

This point can also be reached from the opposite direction. Just as it is possible to conform to (TA) without following it, so it is possible to *try* to follow (TA) but fail to do so. Such a person will aim at the truth in her assertion but fail to hit her target; she will assert *by mistake* something that happens to be false. While there is a failure here, there will not have been any lapse in honesty. And the intelligibility of this form of failure seems to turn on what the speaker believes: a

⁹ We have in mind a simple case where someone deliberately asserts the opposite of something she falsely believes. What she says will be true, and so her assertion will be in accordance with (TA), but she will not have *followed* that rule, and will not have asserted honestly. We take no stand on whether such an assertion counts as a lie.

person will *try* to follow (TA) only if what she asserts is something she believes, and this attempt will be a failure if her belief is false. But once again, not just any assertion of something one believes will amount to an honest attempt to follow (TA), as the examples in the previous section make clear: e.g. the Perceptive Introspector will not even have *tried* to follow (TA) if he asserts, merely on the strength of the ‘internal’ evidence revealing his belief, that he had an unhappy childhood. The truth or falsity of his assertion is irrelevant to that assessment.

We are now in a position to see why the envisioned response to our argument cannot succeed. It is true that the Perceptive Introspector, Faithful Patient, and Lucky Spiritualist all fail to satisfy the condition on truthfulness given in (TA). Each of them asserts what he believes, and does so because he understands himself to believe this, but *does not* aim to assert only what is true. But this reveals not that honest assertion has a truth-requirement *in addition* to the requirement to assert only what one believes. It reveals instead that, properly understood, the truth-requirement (TA) and the belief-requirement (BA) are different ways of formulating the very same condition. A person who is in a position to follow the norm of truth as stated in (TA) must know not only how *things* are, but also how *she* is, in respect of the requirement to be truthful. This is brought out in the fact that a person who asserts dishonestly what turns out to be true won’t then conclude that she wasn’t being dishonest after all, while one who asserts honestly what happens to be false may retreat to the judgment that she *believed* what she said, and so at least wasn’t being dishonest. But we have yet to see how this doxastic self-knowledge can be grounded in evidence, including evidence of a specially ‘first-personal’ kind. And the introduction of (TA) does not close that gap.

7. We argued in §2 that in order to assert honestly, a speaker must not only believe what she asserts, but must also grasp the fact that she believes this. The argument of §§3-5 questioned whether this grasp could arise from an empirical process, even one whose nature and scope were distinctively first-personal. Our conclusion, which we work out in more detail in §8 below, is that the self-knowledge that grounds the possibility of honest assertion must be non-empirical. But first we will consider a final reply on behalf of the empirical conception.

The reply we will consider begins with the modest idea that it is dishonest to assert a proposition that does not *seem* to one to be true:

(SA) Assert that *p* only if *p* seems true.

The next step in this reply is the observation that since people tend to believe the things that seem true to them, the fact that something seems to one to be true can provide an empirical route to doxastic self-knowledge—knowledge, that is, that one

believes this seemingly true thing. And the final step is the observation that it *will not seem* to our subject in the Basic Case, or any of our variants on it, as if he had an unhappy childhood. Thus, it will be proposed, the true lesson of these cases is that honesty requires that the asserted proposition seem to the speaker to be true, and that its seeming-to-be-true be part of the speaker's route to knowledge that she believes it.

The details of this strategy will depend in part on how one explicates the relevant sense of 'seems'. One way is phenomenological:

(SA-P) Assert that *p* only if *p* phenomenally seems true.

The concept of a phenomenal seeming—that is, a proprietary phenomenology that attends to actively considering and affirming the truth of a proposition—is contentious,¹⁰ but we will not dwell on that here. We note only that this first interpretation of (SA) can be contrasted with another one, where a person understands her belief not through a bit of phenomenology but rather through an *epistemic* status that links her belief to the world:

(SA-J) Assert that *p* only if you have compelling doxastic justification for believing that *p*.

This latter condition will seem plausible insofar as we observe that people tend to believe the things that they have compelling doxastic justification for believing, which means that satisfying (SA-J) will be evidence that one satisfies the condition in (BA). However, unlike the routes to self-knowledge of the Faithful Patient and Perceptive Introspector, the possession of compelling doxastic justification also links a person to the *truth* of what she believes, perhaps putting her in a position to assert this belief honestly.

Both (SA-P) and (SA-J) are attractive ways to explain the impossibility of honest assertion in our variations on the Basic Case, since as we have granted it will not seem to the Would-Be Liar, Lucky Spiritualist, Faithful Patient, or Perceptive Introspector as if the beliefs they assert are true, nor will they take themselves to be justified in holding those beliefs. Given this, it may seem that (SA-P) or (SA-J), separately or perhaps in combination, identifies the privileged empirical route to self-knowledge that is missing in those cases. But we will argue below that they do not.

Our initial argument for this conclusion is that (SA-P) and (SA-J) do not give *necessary* conditions for honest assertion: that is, it is possible to assert honestly

¹⁰ For doubts about this idea, see Robinson (2005) and Soteriou (2007)

even if the conditions of (SA-P) and (SA-J) are not satisfied. Consider first that, according to (SA-P), one is dishonest if one makes an assertion unaccompanied by a phenomenal seeming-true. Some doubt whether there is any such phenomenal state. But even the non-doubters couldn't plausibly maintain that a phenomenal seeming accompanies *every* occasion of honest assertion. Over the course of just a single conversation, we might make dozens of assertions about matters that we consider settled, and we may not be actively thinking at all about what we are saying. The idea that each such assertion will be dishonest unless accompanied by a subjective appearance of the truth of the asserted proposition is absurd. This shows that it is possible to assert honestly without satisfying the condition in (SA-P).

(SA-J) avoids this difficulty. So long as a subject makes an assertion partly in light of her justification for thinking that *p* is true, she will be following (SA-J). Her justification need not be *phenomenally* present (whatever that might mean); all that matters is that the subject understands it to be there.¹¹ There is, however, still a problem with this condition, namely that one can assert something honestly *without* taking oneself to be justified in believing this. For it's not particularly unusual that one believes something for which one knows one lacks justification. If one asserts such a belief, then others might find fault with this,¹² but their grounds for doing so could not rightly include one's having been dishonest—after all, what one asserted is exactly what one believed, and one asserted it from the point of view of the belief itself. *Honesty* does not require asserting only what one takes oneself to be doxastically justified in believing. Even if the *knowledge* norm for assertion involves (SA-J), the requirements of honesty do not.

We take these defects with (SA-P) and (SA-J) to be incontrovertible and fatal to the reply under consideration. There is, however, a more general problem with this reply that we think is worth reflecting on. For there may linger in the mind of the reader a thought that, even if (SA-P) and (SA-J) do not work, (SA) itself remains unimpeached. Behind this thought lies the idea that (*a*) the doxastic knowledge necessary for following (BA) is something beyond holding the belief itself, (*b*) the extra factor is a piece of evidence, and (*c*) the evidence in question must point to one's belief 'from the inside', so to speak. Condition (*c*) refers to the 'seeming' that according to (SA) is a requirement of honest assertion. Our objections to (SA-P) and (SA-J) may thus seem not to impugn this intuition as such, but only to show that more needs to be done to articulate the relevant sense of 'seems'.

¹¹ Two obvious questions that we will not belabor: On what basis must the subject understand herself to be doxastically justified? And on what basis must she understand herself to believe as she does on the strength of this justification?

¹² Though they might not be right to—for the belief might be 'properly basic'.

The work required to achieve this would be considerable. The phenomenological and epistemic routes considered above are hopeless. Moreover, one can make trouble for (SA) itself. For example, even if a philosopher takes herself to have worked out a convincing proof of, say, compatibilism about free will, she might nonetheless find that the proposition ‘Free will and determinism are compatible’ *seems false*—especially if she has not recently rehearsed the proof for herself. Nonetheless, it is hard to see how there would be anything *dishonest* in asserting that proposition, if in fact she believes it in the face of the recalcitrant seeming. If this is correct, then it is possible to assert something honestly even if it seems to one to be false. But let us put these worries aside and focus on the very idea that the way things seem to a person could be the privileged source of evidence through which she attains non-alienated, distinctively first-personal knowledge of her belief.

The deeper problem with the appeal to (SA) as a defense of the empirical conception is that there is no good reason why, *if* honest assertion is a matter saying what one knows oneself to believe and the evidence constituted by seeming-truth is good enough to confer this knowledge, then evidence of *other* kinds—i.e., the kinds operative in the cases of the Perceptive Introspector and the Faithful Patient—wouldn’t be good enough as well. This concern is especially pressing when we recognize the extent to which how things seem to a person is *not* an infallible indication of what she believes, as the Basic Case and our variants on it all bring out. If we accept that there are cases where a belief is better known ‘third-personally’ than from the inside, then as long as the role of seeming-truth in doxastic self-knowledge is understood in empirical terms there is no reason why it should be impossible honestly to assert a belief that one knows of only through these other kinds of evidence. The distinction between the right and wrong sorts of empirical process is entirely ad hoc, and raises the suspicion that this requirement is a strained attempt to accommodate a very different truth using the wrong set of resources.¹³

We conclude that (SA) will not help in defending the empirical conception against our arguments. It is time to propose an alternative.

¹³ A further problem with this reply is that even if there is some intuitive pull to the thought that honest assertion requires the seeming-truth of the asserted proposition, there is none at all to the idea that this requirement is due to the way that seeming-truth provides a special sort of *evidence* for what one believes. If the proposal in question were correct (as we have argued it is not), it would be correct only in spite of the fact that it flies in the face of the phenomenology of doxastic self-knowledge.

8. We have argued that the empirical conception of doxastic self-knowledge makes it impossible to understand the first-personal grasp of one's beliefs that is required for honest assertion. This is not only because there is no credible account of the empirical process that could permit such a grasp, but also because the idea that *some* empirical processes can ground the required self-knowledge, while other processes that are otherwise equally reliable will not do in this respect, is unmotivated. The conclusion we draw from this is that when honest assertion is possible it must arise from a grasp of one's beliefs that is not grounded in evidence of *any* sort. Following Moran (2001) and others, we can state this conclusion by saying that ordinary self-knowledge of one's beliefs, i.e. the sort of assertion that allows one to assert those beliefs honestly, is *transparent* to belief itself.

The idea that doxastic self-knowledge is transparent finds a classic formulation in a familiar passage from Gareth Evans's *Varieties of Reference*:

... in making a self-ascription of belief, one's eyes are, so to speak, or occasionally literally, directed outward—upon the world. If someone asks me 'Do you think there is going to be a third world war?', I must attend, in answering him, to precisely the same outward phenomena as I would attend to if I were answering the question 'Will there be a third world war?' (Evans [1982, 225])

Against the empirical conception, we propose that it is only when doxastic self-knowledge is *transparent to belief* in the way described by Evans that a person can assert her belief honestly. Only to the extent that a person knows her belief transparently can she follow the rule (BA) by expressing this belief in honest assertion, speaking *from its point of view* just as Zoe calls the zookeeper because she believes that there's a pink elephant in the bathtub. One who speaks in this way is not merely asserting what she believes (as does Would-Be Liar does), is not merely asserting what she believes or knows herself to believe (as do the Lucky Spiritualist, Faithful Patient and Perceptive Introspector). She is asserting that *p* in light of the (putative) fact that *p*. She speaks honestly, because she is speaking *from the point of the view* of the belief she expresses.

Still one might wonder: How is it so much as possible that a person could have non-empirical knowledge of her beliefs? The argument of this paper does not depend on any specific answer to this question, but it does rule out a misinterpretation of the notion of transparency that has gained some currency in

the literature.¹⁴ According to this misinterpretation, transparent self-knowledge arises from a process of the following form:

Step 1: Consider whether p is true.

Step 2: If it is, then conclude that one believes that p .

For the reasons we explained in §7, we agree with many would-be critics of the transparency method that it is incredible that such a process should be the ground of doxastic self-knowledge. That is, if this process is understood so that the judgment (or appearance of truth, or proposition judged) generated in Step 1 functions as *evidence* justifying the doxastic self-ascription in Step 2, the proposal in question is just a version of the view that we have been arguing against. It is not an account of *non-empirical* self-knowledge.

However, as long as the empirical conception of doxastic self-knowledge is taken for granted, it will seem inevitable that the method of transparency should be construed in this way—as a process by which beliefs about one’s own beliefs are justified. But what is Evans offering, if not an account in which a believer’s view ‘on the world’ feeds into a process that results in well-grounded doxastic self-ascriptions? We propose that the key to avoiding this misinterpretation is to see that, for the person who knows her belief transparently, the judgments ‘ p ’ and ‘I believe that p ’ are *not separable* in the way that the empirical conception assumes they must be. To say that these judgments are not separable is to say that, for the self-knowing believer, the questions they address—respectively, whether p and whether she believes that p —do not concern distinct matters, such that her answer to one could be the *basis* of her answer to the other.¹⁵

There are at least two ways that this inseparability can be understood. **Constitutivist** theories hold that we can have non-empirical knowledge of our beliefs because to take oneself to believe something is, at least in the right conditions, also to believe it.¹⁶ As long as the self-ascription through which belief is supposed to be constituted is not grounded empirically, such a view would avoid our objections to the empirical conception. We note, however, an important barrier to making good sense of Constitutivist proposals: they seem to require seeing doxastic self-attribution as entirely groundless, as the product of a brute disposition

¹⁴ For misinterpretation (or creative reconstruction) along these lines, see Fernández (2003); Byrne (2011); Gertler (2011, 188-190); Cassam (2014, 3-5 and *passim*).

¹⁵ Cf. Moran (2001, § 2.6).

¹⁶ Recent Constitutivist proposals include Heal (2001), Schwitzgebel (2011), and Coliva (2012).

or something along these lines.¹⁷ This avoids the pitfalls of the empirical conception but is hardly a credible interpretation of Evans's insight, nor do we find it to possess much independent plausibility of its own.

An alternative to Constitutivism is the **Self-Conscious Conception** of doxastic self-knowledge. This conception reverses the Constitutivist's order of explanation, holding that it is in the nature of belief to be self-conscious, and that because of this a person does not need to rely on introspection or any other empirical process as a source of knowledge of her beliefs. The thought is not that (as for the Constitutivist) the thinker simply finds herself believing that she believes that *p* and, in virtue of this fact, believes that *p*. Rather, according to the Self-Conscious Conception a person is ordinarily in a position to grasp what she believes simply in virtue of believing it, as it is part of what it is to view a proposition with the belief-attitude that one thereby knows oneself to so view it. Developing this approach in detail requires explaining how it could be in the nature of a psychological attitude like belief to be non-empirically self-known—a task beyond the scope of this paper.¹⁸

Nothing in what we have argued here bears directly on the question of which of these accounts (or perhaps some other) of transparent self-knowledge is the right one. Our central contention is that *some* explanation of how there can be non-empirical knowledge of belief is required for an adequate account of honest assertion. Full development of this position is work for another occasion.

¹⁷ Among contemporary Constitutivists, Schwitzgebel (2011) is notable in embracing this consequence.

¹⁸ See Boyle (2011) and Marcus (2016) for work in this vein.

References

- Armstrong, D. M.. 1973. *Belief, Truth and Knowledge*. Cambridge University Press.
- Boyle, Matthew. 2011. "Transparent Self-Knowledge." *Aristotelian Society Supplementary Volume* 85 (1):223-241.
- Byrne, Alex. 2011. "Transparency, Belief, Intention." *Aristotelian Society Supplementary Volume* 85 (1): 201-221.
- Carruthers, Peter. 2010. "Introspection: divided and partly eliminated." *Philosophy and Phenomenological Research* 80 (1): 76-111.
- Cassam, Quassim. 2014. *Self-Knowledge for Humans*. Oxford: Oxford University Press.
- Coliva, Annalisa. 2011. "One variety of self-knowledge: constitutivism as constructivism." In *The Self and Self-Knowledge*, edited by Annalisa Coliva. Oxford: Oxford University Press.
- Dancy, Jonathan. 2000. *Practical Reality*. Oxford: Oxford University Press.
- Evans, Gareth. 1982. *Varieties of Reference*. Oxford: Oxford University Press.
- Fernández, Jordi. 2003. Privileged access naturalized. *The Philosophical Quarterly* 53 (212): 352-372.
- Gertler, Brie 2001a. *Self-Knowledge*. Routledge.
- Gertler, Brie. 2011b. "Self-Knowledge and the Transparency of Belief." In *Self-Knowledge*, edited by Anthony Hatzimoysis. Oxford: Oxford University Press.
- Goldberg, Sanford. 2015. *Assertion: On the Philosophical Significance of Assertoric Speech*. Oxford: Oxford University Press.
- Heal, Jane. 2001. "On first-person authority." *Proceedings of the Aristotelian Society* 102 (1): 1-19.

- Kvanvig, Jonathan L. 2009. "Norms of assertion." In *Assertion: New Philosophical Essays*, edited by Herman Kappelan and Jessica Brown. Oxford: Oxford University Press.
- Marcus, Eric. 2016. "To believe is to know you believe." *dialectica* 70 (3): 375–405.
- Moran, Richard. 2001. *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton: Princeton University Press.
- Robinson, William. 2005. "Thoughts without distinctive non-imagistic phenomenology." *Philosophy and Phenomenological Research* 70 (3): 534-561.
- Ryle, Gilbert. 1949. *The Concept of Mind*, New York: Barnes and Noble.
- Schwitzgebel, Eric. 2011. Knowing your own beliefs. *Canadian Journal of Philosophy* 35: 41-62.
- Schwitzgebel, Eric, and Myers-Schulz, Blake. 2013. "Knowing that *p* without believing that *p*." *Noûs* 47 (2): 371-384.
- Soteriou, Mathew. 2007. "Content and the stream of consciousness." *Philosophical Perspectives* 21 (1): 543–568.
- Weiner, Matthew. 2005. "Must we know what we say?" *The Philosophical Review*, 114: 227–51.
- Williamson, Timothy. 2000. *Knowledge and its Limits*. Oxford: Oxford University Press.