

# Self-Knowledge: Expression without Expressivism

Lucy Campbell 

Department of Philosophy, University of Warwick

**Correspondence:** Lucy Campbell

Email: [lucy.campbell@warwick.ac.uk](mailto:lucy.campbell@warwick.ac.uk)

## Funding information

Leverhulme Trust, Grant/Award Number: ECF-2018-075

## 1 | INTRODUCTION<sup>1</sup>

Few would say that we are infallible with respect to our own states of mind. Still, our ordinary ways of thinking represent people as having an especially intimate relationship to their own attitudes, sensations, emotions, and so on. This special intimacy is suggested *inter alia* by the default assumption that when sincerely made, present-tense mental state self-ascriptions – statements like “My feet are aching”, “I want to leave the party”, “I think Justin is bored” – will be true. Following familiar usage, I will call such statements ‘self-ascriptions’, and will label this feature of them their ‘first-person authority’.

It is natural to understand first-person authority as bound up with the fact that we are in an especially good position to *know about* our own states of mind, with the idea being that sincere self-ascriptions express this knowledge. Call this the ‘epistemic approach’ to understanding first-person authority.

But there is also a prominent line of thought which rejects the epistemic approach. Expressivists about self-ascriptions think that it fails to capture a second feature of these statements, what I will call (following Bar-On (2004, 6 ff.)) their ‘epistemic asymmetry’ with other kinds of assertion.

---

<sup>1</sup>Versions of this material were presented in 2019–20 at Warwick, Stirling, and Cardiff. For comments and criticisms at these meetings, not all of which I have been able to address here, I thank Dorit Bar-On, Tom Crowther, Naomi Eilan, Alexander Greenberg, David Hunter, Anneli Jefferson, Guy Longworth, Alisa Mandrigin, Lucy O’Brien, Orestis Palermos, Johannes Roessler, Wesley Sauret, Jack Shardlow, Peter Sullivan, Alessandra Tanesini, and Crispin Wright. For comments on earlier drafts, thanks to Sophie Archer, Alexander Greenberg, Eylem zaltun, and Johannes Roessler, and in particular to an anonymous reviewer for this journal, who helped me to significantly improve the paper. Thanks also to Maria Alvarez, Nathan Hauthaler, Jane Heal, and Mark Textor, for discussion of these and related topics over the years. This paper was written during a Leverhulme Early Career Fellowship, and I am grateful to the Leverhulme Trust for this support.

---

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2020 The Authors. *Philosophy and Phenomenological Research* published by Wiley Periodicals LLC on behalf of Philosophy and Phenomenological Research Inc

Self-ascriptions appear to be epistemically unlike other kinds of assertion in that it is typically inappropriate to ask a self-ascriber for the reasons or evidence on the basis on which she knows about her own state of mind. Nor is it clear what epistemic method or process might underwrite one's ability to self-ascribe mental states. In response to such considerations, Expressivists offer a non-epistemic explanation of first-person authority, in terms of the idea that self-ascriptions are 'avowals', utterances whose primary function is to express the state they self-ascribe, rather than a second-order judgment to the effect that one is in this state.

In my view, Expressivists are right to take seriously self-ascriptions' epistemic asymmetry with other forms of assertion, and right also that doing so requires understanding self-ascriptions as avowals. But they are wrong to think that doing so requires a non-epistemic approach to first-person authority, and also wrong to find such an approach unproblematic.

After explaining how Expressivism is motivated by the need to reconcile self-ascriptions' first-person authority with their epistemic asymmetry (section 2), I argue that we should reject Expressivism because it doesn't adequately capture the epistemology of self-ascription (section 3). This discussion motivates developing a version of the epistemic approach, albeit one which like Expressivism, puts self-ascriptions' epistemic asymmetry centre stage. I do so by showing how viewing self-ascriptions as avowals can itself explain why they manifest genuinely epistemically ungrounded self-knowledge (sections 4-5). The account works by locating the phenomena of avowal, self-knowledge, and self-ascription within a 'forward-looking' epistemology, on which propositional knowledge is viewed as a certain kind of rational ability.

## 2 | FIRST-PERSON AUTHORITY, EPISTEMIC ASYMMETRY, AND EXPRESSIVISM

Here I detail the epistemic approach to understanding first-person authority (2.1), and explain why it might seem unable to accommodate their epistemic asymmetry (2.2). I then describe the Expressivist's non-epistemic response to these considerations (2.3).

### 2.1 | First-person authority and the epistemic approach

Contrast self-ascriptions with ascriptions of mental states to others. Seeing Justin across the room, shifting position uncomfortably, I say "Justin's bored". But Justin isn't bored; far from it. He's absorbed in what Debbie's telling him about pipistrelles – although it's true his new shoes are making his feet ache. We can make mistakes about our own states of mind (of which more in 3.2), but we don't seem able to make this kind of mistake. Justin might think himself newly fascinated by pipistrelles when in fact he's newly fascinated by Debbie, but he doesn't seem able to think himself bored whilst fascinated, or mistake his aching feet for boredom.

Within a certain boundary at least,<sup>2</sup> our default assumption is that someone level-headed and sincere who stably and confidently self-ascribes some mental state will do so correctly.<sup>3</sup> It is the cases

<sup>2</sup>This qualification captures the fact that some of our mental states can remain first-personally unknown, centrally states about which we are self-deceived, or the kinds of states which are the objects of psychoanalytic enquiry. I say more about such cases in 5.1.

<sup>3</sup>Note that the 'default' assumption I describe here is different (although not unrelated) to that found in Wright (1998, 41).

where a person gets it wrong about her own state of mind which need a special explanation, not those where she gets it right. As it is sometimes put, if you want to know about S's states of mind, S is the best person to ask. These are all aspects of self-ascriptions' first-person authority.

Usually, thinking of someone as an authority on some topic goes hand-in-hand with thinking of them as particularly *knowledgeable* about it. The person whose word we take on trust on questions of particle physics is the particle physicist: she's the expert. In court, the testimony of an eyewitness is privileged over certain other kinds of evidence: she should know, she saw it happen. Luke and Peter tell each other everything, so if we want to understand Peter's odd behaviour, Luke is a good person to ask. When sincere, the particle physicist's claims about particle physics, the eye-witness's claims about what she saw, and Luke's claims about Peter's behaviour are, compared to claims made on these topics by the rest of us, particularly likely to be true. Why? Because the particle physicist, the eyewitness, and Peter's friend Luke, *know a lot about* their respective subject-matters.

Many sincere self-ascriptions are recognised as being even more authoritative than the statements of experts, and it is natural to understand this first-person authority as reflecting the fact that we are each in a special epistemic position with respect to our own minds: S is especially likely to accurately self-ascribe her own states of mind, because she is especially knowledgeable about them. This natural thought embodies the epistemic approach to understanding self-ascriptions, and to the relationship to one's own mental states which these ordinarily manifest.

## 2.2 | Epistemic asymmetry

But self-ascriptions have a second distinctive feature, which can seem hard to square with the epistemic approach. For epistemically speaking, self-ascriptions are distinctly unlike other forms of assertion. It is from this perspective that the Expressivist thinks we need a *non*-epistemic account of first-person authority.

Rather generally, it is often reasonable to query a speaker's epistemic credentials when she asserts something. Sometimes we do this from a sceptical perspective ("Do you *know* that, or are you just guessing?"; "Have you got any evidence?"; "You reckon? Prove it!"), but sometimes we are just curious about how someone found something out, or what evidence she's going by ("How do you know?"; "What led you to that conclusion?"). Either way, ordinary assertions about contingent matters of fact are typically open to appropriate epistemic questioning of some kind or another.

But these kinds of epistemic query or challenge seem out of place or inappropriate in relation to ordinary self-ascriptions. If I say that Abi is feeling sick, you might ask me how I know this – and I might answer, saying (e.g.) "I slipped some deathcaps in her stroganoff" or "Look at her!" But I can't sensibly answer in these ways if asked how I know that *I* am feeling sick. Again, if I say that I am feeling sick you might ask me why this is so (why I am *feeling sick*), but it is hard to make sense of seeking a substantive answer to the question of why I *believe* that I am feeling sick. Answers like "Because I *feel* sick" or "Because I *am* feeling sick" simply repeat my initial self-ascription; they do not explain it.

The same goes in other cases: "How do you know that your feet ache?" is vaguely idiotic unless perhaps heard as a facetious way of downplaying one's interlocutor's pain ("You think your feet ache *now* – come back to me when you've tried dancing in these heels."); asking someone who says she wants to go skydiving *why* she thinks this, is hard to hear as anything other than a convoluted way of asking after her reason *for going skydiving*, or of dismissing her desire as beyond one's own comprehension. Yet read literally, the form of words employed makes this a question about her *beliefs*, not her *desires*. "What's your evidence for thinking that Justin is bored?" is perfectly reasonable; "What's your evidence for thinking *that you believe that* Justin is bored?" is hard to interpret full-stop.

In all of these cases it is (at least) hard to hear these questions as straightforward requests to be provided with the speaker's epistemic credentials. Self-ascriptions in this way clearly contrast with expert claims like those considered in 2.1. Asking the particle physicist how she knows that the Higgs Boson has been discovered at CERN, for instance, is perfectly reasonable. The question need not be a sceptical or 'pointed' one, one which carries the suggestion that she might not know (Austin 1961, 78; see also Roessler 2013, 3). It might only express an open-minded interest in understanding the structure of the scientific evidence.

One common response to the observation that self-ascriptions are in this way epistemically asymmetrical with other forms of assertion, is to deny that in order for a person to have formed their belief on some epistemic basis, or as the result of some epistemic method, this basis or method must be reflectively accessible to her.<sup>4</sup> But this appeal to epistemic externalism misses the point.

To see why, it is instructive to compare my self-ascription "I want to leave the party" to (the externalist's friend) the chicken sexer's claim, "this chick is female". Say we buy the externalist's claim that the chicken sexer *knows* that the chick she is holding is female despite being unable to say how her knowledge is epistemically grounded, or (if this is different) how she came by it. Still, it is perfectly appropriate to ask the chicken-sexer how she can tell, or why she believes, that the chick is female. (Granted, such questions become inappropriate once we discover that she can't answer them, but this is a separate issue.) The chicken-sexer's ability to make accurate statements about the sex of chicks whilst having no idea how or on what basis she does so strikes us as mysterious, indeed rather extraordinary.<sup>5</sup> The case of an ordinary self-ascriber is not like this. From the common sense perspective, there is nothing mysterious or extraordinary about my ability to truly self-ascribe aches and pains, or a desire to leave the party, without being able to cite any epistemic method or basis. On the contrary, the strange thing would be for an interlocutor to expect me to cite any such method or basis.

In short, epistemic externalism allows one to hold onto the theoretical possibility of knowing that something is the case without being able to explain on what basis or by what method one knows it. But it does not explain why, in the case of self-knowledge, it is out of place so much as to enquire into the knower's epistemic basis or method. Plausibly, the ordinary inapplicability of such questions to my self-ascription "I want to leave the party" shows not that the epistemic basis for this self-ascription is inaccessible to me, but that it rests on no epistemic basis at all.

### 2.3 | Expressivism and the non-epistemic approach

The foregoing suggests that the epistemic asymmetry between self-ascriptions and other kinds of assertion is rather intractable. The Expressivist's move is to reject the epistemic approach to understanding first-person authority wholesale: if self-ascriptions are first-person authoritative because they express a (special) form of *knowledge*, then it should be appropriate to epistemically query the self-ascriber. But it isn't. So we should reject the epistemic approach. Or so goes the thought.

The Expressivist therefore opts for a *non-epistemic* explanation of self-ascriptions' first-person authority. Key to this strategy is the idea that self-ascriptions are a special kind of speech-act which does not primarily function to describe or report one's being in the self-ascribed state, or to express a

<sup>4</sup>There are a number of other responses in the literature which I do not discuss here. My aim is not to provide a survey, but to impart an initial sense of the apparent difficulty faced by the epistemic approach.

<sup>5</sup>In fact this may not accurately describe the practise of real-world chicken sexers, who appear to have a rather detailed sense of how their judgments are based (see Biederman and Shiffrar, 1987). Thanks to Johannes Roessler for this pointer (p.c.).

belief or judgment to the effect that one is in it, but instead to directly express the self-ascribed state itself. Expressivists term this special kind of speech-act an *avowal*. In avowing “My feet ache”, for instance, I ‘give voice to’, ‘vent’, or ‘speak from’ my pain itself.<sup>6</sup> Despite their surface similarities, “My feet ache” (or “I want to leave the party”, or etc.) as it would ordinarily be said, and “Justin’s feet ache” (“Justin wants to leave the party”, or etc.), as it would ordinarily be said, do very different jobs. The main job of the latter is to express my view on Justin’s state of mind. The main job of the former is to express not *my view on* my state of mind, but my state of mind itself; to express my pain (my desire to leave the party, or etc.).

On this view, it is not because they express self-knowledge, but because they are avowals, that self-ascriptions are first-person authoritative. The special relationship a person bears to her own mental states is that of being able to express them, and self-ascription is one way of doing so. This is why S herself is the best person to ask about whether she is in some state M: she can directly express M in an avowal. It is because avowals are direct expressions of one’s mental states that there is not the same scope for inaccuracy in avowing as there is in ascribing mental states to others. The strangeness and uncommonness of being both sincere and inaccurate in self-ascribing a mental state can be understood on the model of other kinds of non-epistemic expressive mistake, such as saying “Ow!” on feeling a mild and pleasant sensation, or “Hurrah!” when one means “Oh no!”

Expressivists explain self-ascriptions’ epistemic asymmetry in a related way. Rather generally, a person doesn’t need any epistemic basis for taking herself to be in some state in order to *express* it – to express her pain by groaning, say. When it is an avowal of pain, my self-ascription “My feet ache” needs no more evidential support than does a pained grimace. Nor should we expect my capacity to self-ascribe my pain to depend on any method for discerning it, or process of coming to know about it. On this view, it is because there are simply no answers to epistemic queries about self-ascriptions that the queries themselves turn out to be inappropriate.

### 3 | AGAINST EXPRESSIVISM

As we have seen, the Expressivist denies that first-person authority is to be *explained in terms of* a self-ascriber’s self-knowledge. But this is consistent with claiming that a self-ascriber will typically have self-knowledge of the state she self-ascribes, and even that this is expressed in her self-ascription. In 3.1 I describe three forms of Expressivism, distinguished by their attitudes to these additional claims. In 3.2 and 3.3 I argue that none of these positions is strongly motivated. Understanding what goes wrong for these different forms of Expressivism will motivate the approach to understanding self-ascription which I develop in sections 4 and 5.

#### 3.1 | Three forms of Expressivism

I call the three forms of Expressivism I want to distinguish ‘Eliminativist’, ‘Deflationist’, and ‘Parallelist’ Expressivism.

What I call Eliminativist Expressivism is often attributed to the later Wittgenstein, and is held by some contemporary Wittgensteinians (Kenny 1973, ch. 10; Hacker 1986, ch. X). As we

---

<sup>6</sup>For different ways of unpacking these metaphors see Finkelstein (2003) and Bar-On (e.g. 2004). I say more about avowals and expression in section 4.

have already seen, it is central to any Expressivist view that we can speak truly about our own states of mind.<sup>7</sup> What the Eliminativist denies is that this capacity is in any way an epistemic one. The Eliminativist does not claim that we are ignorant about our own states of mind, but more fundamentally that epistemic concepts simply don't apply here. Our capacity to speak truly about our own minds is expressive *rather than* epistemic, so that describing a person either as knowing or as being ignorant about her own belief, pain, etc., would be equally, strictly speaking, nonsense.<sup>8</sup>

Deflationist Expressivists take their mark from a weaker reading of Wittgenstein (Finkelstein 2003 §6.4; see also Kemmerling, 2012). Unlike the Eliminativist, the Deflationist allows that one can “meaningfully [and] truly” (Finkelstein 2003, 151) describe someone who avows some mental state as expressing knowledge that she is in it. But she also denies that this ‘knowledge’ is *knowledge*, in an ordinary or substantive sense. For the Deflationist, prefacing a statement like “I want to leave the party” or “My feet are aching” with “I know” makes no substantive difference to its meaning or truth-conditions.

Only those who I call Parallelist Expressivists wholeheartedly accept that avowals like “My feet ache” or “I want to leave the party” express a genuine form of knowledge (see Bar-On e.g. (2004)). The Parallelist, like all Expressivists, rejects the epistemic approach to understanding first-person authority, explaining the latter in terms of expression and avowal. But she also holds that we genuinely know facts about our own states of mind, and that this knowledge is expressed in our self-ascriptions – as it were, in parallel to the expression of the self-ascribed state. For the Parallelist, self-ascriptions are first-person authoritative, and genuinely knowledgeable, but not first-person authoritative *because* they are genuinely knowledgeable.

### 3.2 | Rejecting Eliminativist and Deflationist Expressivism

I start with a specific worry about Deflationist Expressivism. As noted above, the Deflationist holds (a) that it can be both *meaningful and true* to say that someone *knows* that (e.g.) her feet are aching, but (b) that doing so adds nothing substantive to saying that her feet *are* aching. Holding (a) distinguishes the Deflationist from the Eliminativist, whereas holding (b) distinguishes her from the Parallelist. But intuitively (a) and (b) aren't a happy combination. If attributions of self-knowledge are *meaningful*, how can they *add nothing to the meaning* of simple attributions of the relevant (i.e. putatively self-known) states? Deflationism threatens to collapse into either the epistemically weaker Eliminativism, or the epistemically stronger Parallelism.

But I won't dwell on this. Because even if Deflationism were stable, I do not think that it is well-motivated, and for the same reason that I think that Eliminativism is not well-motivated. In short, I am not convinced by the considerations given for denying that our relationships to the states we authoritatively self-ascribe are genuine or substantive knowledge-relationships.

<sup>7</sup>A much stronger ‘Expressivism’ holds that self-ascriptions are *semantically* akin to non-linguistic expressive behaviours, strictly speaking lacking meaning and truth-conditions. I don't consider this idea here. It clearly can't accommodate first-person authority understood as the idea that self-ascriptions are *true*, if sincere.

<sup>8</sup>This is a simplification: the Eliminativist can allow that we can know about our own states of mind in special cases – as when I infer that I must be jealous based on my erratic behaviour and unjustified feelings of rage. Self-ascriptions made on such a basis would express knowledge akin to the knowledge we have of the minds of others; they would not be first-person authoritative or epistemically asymmetrical with other forms of assertion.

The most often leant-on argument against thinking that we properly speaking *know about* our own states of mind, derives from Wittgenstein.<sup>9</sup> The following from *Philosophical Investigations* is a particularly commonly cited passage, although there are many others scattered throughout Wittgenstein's later works:

246. If we are using the word “to know” as it is normally used (and how else are we to use it?), then other people very often know when I am in pain. — Yes, but all the same not with the certainty with which I know it myself! — It can't be said of me at all (except perhaps as a joke) that I *know* I am in pain. What is it supposed to mean — except perhaps that I *am* in pain?

Other people cannot be said to learn of my sensations *only* from my behaviour, — for I cannot be said to learn of them. I *have* them.

The truth is: it makes sense to say about other people that they doubt whether I am in pain; but not to say it about myself.

(Wittgenstein 2003)

Such remarks are taken to suggest the following kind of argument: it only makes sense to apply the ordinary concept *knowledge* in contexts where it would also make sense to apply certain other internally related epistemic concepts, concepts like *learning* and *doubting*. But it doesn't make sense to think of someone as having *learned*, or as potentially *doubting*, that they are in a certain mental state. So either it doesn't make sense at all to say that a person *knows* that she's in some mental state (so concludes the Eliminativist Expressivist), or it can make sense to say that someone knows that she's in some mental state, but only if we use using 'knows' in a non-standard or non-substantive sense (so concludes the Deflationist Expressivist).

Let's grant for the sake of argument that it ordinarily wouldn't make sense to say that someone has *learned* that they are in some mental state, or that they might *doubt whether* they are in some other. Still, this doesn't get us very far. For there a number of other concepts which are internally related to *knowing* which can sensibly and intelligibly be applied to a person's relationship to facts about her own mind. This seems true even in relation to facts about one's own pains and sensations, states often represented as especially unamenable to being objects of genuine first-person knowledge. So I focus on these.

First, it's possible to *make mistakes or get it right* about whether one is in some mental state.<sup>10</sup> It is relatively familiar to mistake a sensation in one's toe for a burning pain when in fact one has merely brushed it up against the cold bath tap. Such cases also show that, second, one can *realise or fail to realise* that one is in some state – in this case, I *come to realise* that I am having a sensation of coldness in my toe rather than burning pain. Consider too the experience of waking up ill or in pain. At least

<sup>9</sup>Christopher Coepe (1973, 247) interprets Wittgenstein as subscribing to Eliminativism, diagnosing this as stemming from an acceptance of traditional thinking about knowledge as true belief 'plus an account'. PMS Hacker (1986, 277) thinks Wittgenstein is an Eliminativist too, but his diagnosis depends on viewing Wittgenstein as rejecting a “traditional epistemology” – in this case one which treats our own minds as especially knowable. For an interpretation of Wittgenstein which is in some respects (broadly) in line with the account I develop in section 4, see Sankowski (1978).

<sup>10</sup>There are interesting differences between the kinds of mistake which seem possible in relation to different kinds of mental state, e.g. I can be self-deceived about my own beliefs and desires, but seemingly not about my own pains. A full understanding of self-knowledge would include a detailed study and explanation of these differences.

sometimes, this is an experience of becoming aware of a pre-existing sensation, rather than of developing (e.g.) a headache or nauseous feeling as one wakes up.

Third, it's possible to *be unsure whether or sure that* you're in some mental state, for example at its onset being unsure whether a certain sensation is a pain or an itch, and then after a few seconds becoming sure that it's (e.g.) a pain. Again, I don't deny that there are cases in which what started as an itch becomes a pain (insect bites and stings often seem to be like this – turning from itch to pain as the poison takes hold), or cases where a sensation is indeterminate. The point again is that not all cases are like this. A particularly anxious or hypochondriac person may well interpret as painful something which they later recognise to be merely itchy. This kind of thing strikes me as relatively common (although in need of a special explanation). Fourth, it seems possible to *forget* you're in some state (“I was so happy to see her, I forgot all about how much my feet hurt!”), and to remember being in it (“She was so dull and self-absorbed I remembered how much my feet hurt, and decided to leave.”). Fifth, we *inform* others of facts that we know – and no less of facts about our own states of mind. Sixth, we *infer* things from facts that we know, and again no less from facts about our own states of mind<sup>11</sup> - and so on.

In summary, even if we grant that some epistemic concepts internally related to that of *knowledge* (e.g. *learning*; *doubting*) do not apply to our typical relationships to facts about our own mental states, it seems equally true that other such concepts do so apply. I don't want to deny that there are differences between the precise ways in which epistemic concepts such as *mistake*, *realisation*, *being sure*, etc., might characterise our relationships to facts about our own mental states, by contrast with the ways in which they characterise our relationships to facts about the mental states of others, or wholly non-mental facts. But this would only get us the conclusion that self-knowledge is in certain respects *different* from knowledge of facts about other minds and about worldly goings-on. I think that this conclusion is true, and a key aim of the account I provide in section 4 is to accommodate and explain it. But the conclusion falls short of motivating either Eliminativist or Deflationist Expressivism.

### 3.3 | Rejecting Parallelist Expressivism

The Parallelist Expressivist accepts that self-ascriptions ordinarily manifest a substantive form of knowledge. But Parallelism also suffers a fatal lack of motivation. There is a dilemma for the Parallelist, which arises when we ask whether or not self-knowledge is supposed, like other forms of knowledge, to require epistemic support by evidence, reason, or method. For simplicity, I discuss the issue in relation to evidence.

Consider my knowledge that I want to leave the party. Does this rest on evidence for thinking that I want to leave? If she says that it does, the Parallelist will be left without an explanation of its epistemic asymmetry with other kinds of (putatively) knowledge-expressing assertion. The central Expressivist idea, recall, is that a non-epistemic account of self-ascriptions' first-person authority can explain their epistemic asymmetry. But if the Parallelist conceives of my self-ascription “I want to leave the party” as expressing evidentially grounded self-knowledge that I want to leave the party, in addition to expressing my desire to leave, then the question of why it should be inappropriate for you to ask me to cite my evidence simply reappears. The explanation of my self-ascription's epistemic asymmetry with other assertions will have gone missing.

The Parallelist is therefore likely to suggest that the self-knowledge expressed in a self-ascription is evidentially *groundless*. The fact that my self-ascription expresses knowledge that I want to leave

<sup>11</sup>On these last two points see, respectively, Roessler (2015, §2) and Hyman (2015, §7.4).



the party needn't threaten its epistemic asymmetry with other forms of assertion if the knowledge it expresses is itself epistemically asymmetrical with that expressed in other forms of assertion, in virtue of being, unlike them, evidentially ungrounded.

But note that now the Parallelist has no reason to offer a *non-epistemic* account of first-person authority. The motivation for doing so was – again – to accommodate self-ascriptions' epistemic asymmetry with other forms of assertion. But if the self-knowledge expressed in a self-ascription is already unlike the knowledge expressed in other assertions, precisely in being evidentially ungrounded, then an explanation of self-ascriptions' epistemic asymmetry comes along for free, and there is no further explanatory work left for a *non-epistemic* account of first-person authority to do. Since the adoption of a non-epistemic account of first-person authority (in terms of the phenomena of expression and avowal) is the *sine qua non* of Expressivism, Parallelist Expressivism is again left unmotivated.<sup>12</sup>

### 3.4 | Taking stock

As I explained in 2.2, it can seem difficult for an epistemic approach to understanding first-person authority to accommodate self-ascriptions' epistemic asymmetry with other forms of assertion. It is from this perspective that the kind of non-epistemic approach favoured by the Expressivist was supposed to look attractive.

But on closer inspection, no form of Expressivism appears able adequately to capture self-ascriptions' typical epistemic profile. The argument for denying that self-ascriptions express a substantive form of knowledge is not compelling, so I think we should reject both Eliminativist and Deflationist Expressivism. The Parallelist allows that self-ascriptions express genuine self-knowledge, but in so doing undermines her Expressivism. On one horn of the dilemma in 3.3 she is unable to explain self-ascriptions' epistemic asymmetry; on the other she removes the need for a non-epistemic account of their first-person authority. Yet the central claim of Expressivism is that we need a non-epistemic account of self-ascriptions' first-person authority in order to explain their epistemic asymmetry.

I want to develop an account of self-knowledge and self-ascription which is broadly along the lines of that described on the second horn of the Parallelist's dilemma. This was unavailable to the Expressivist because it is a version of the epistemic approach. But if we are prepared to allow that self-knowledge is genuinely epistemically ungrounded, I think that the epistemic approach turns out to be benign. I develop these ideas in the rest of the paper.

## 4 | SELF-KNOWLEDGE: EXPRESSION WITHOUT EXPRESSIVISM

According to the epistemic approach, self-ascriptions are first-person authoritative because they are particularly likely to express knowledge of one's own mind. But if the epistemic approach is to work, it needs to do justice to the epistemic asymmetry between self-ascriptions and other forms of assertion. And as long as we are careful not to make certain problematic assumptions

<sup>12</sup>Dorit Bar-On provides various possible positive accounts of self-knowledge (2004, ch. 9). One of these (see esp. p. 390) has some affinity with the account I will develop later on, but as an Expressivist Bar-On seems unable to adopt it on pain of landing on the second horn of the Parallelist's dilemma described above.

about the nature of knowledge in general, and self-knowledge in particular, I think that it is not only natural but also theoretically unproblematic to view first-person authority in epistemic terms.

The main problematic assumption that we need to avoid is that knowing that *p* is rather generally a matter of having a well-grounded (and true) belief that *p*. This very general picture includes epistemologies which otherwise differ on some very fundamental questions, including (*inter alia*) what true belief must be grounded in to count as knowledge (evidence? reasons? epistemic process or method?); whether grounding is epistemically *internal or external*, and whether or not knowledge is *reducible* to well-grounded true belief.<sup>13</sup> So the picture of knowledge as well-grounded true belief is very widely held. But it is far from mandatory. Once this is recognised, we can start to consider how it might be rejected, and how rejecting it might make room for a better understanding of the distinctive character of both self-knowledge and self-ascription. This is my plan for the rest of the paper.

#### 4.1 | The idea in outline

My explanation of why self-knowledge is not epistemically grounded (and so of self-ascriptions' epistemic asymmetry) rests on the idea that simply being in a mental state is *in principle sufficient* for knowing that one is in it. 'In principle sufficient', here, means *inter alia*, 'sufficient without the possession of additional reasons or evidence for thinking that one is in it' (I will say more about this idea in 5.1). A person has self-knowledge as we have been understanding it – epistemically ungrounded knowledge that one is in some state of mind – when she knows about her own state just in virtue of being in it.<sup>14</sup> This explains both self-ascriptions' first-person authority and their epistemic asymmetry, as follows.

The idea explains first-person authority in the kind of epistemic terms which the Expressivist urged us – I have argued wrongly – to reject. It allows us to say that sincere self-ascriptions are strongly or by default presumed to be true, and that the best person to ask about *S*'s state of mind is *S* herself, because being in some state is in principle sufficient for knowing that one is in it, so that a person's statements about her own states of mind are strongly or by default presumed to be *knowledgeable* (and because knowledge is factive).

The idea explains epistemic asymmetry, because most circumstances are such that their obtaining is *not* – even in principle – sufficient for anyone to know that they obtain. Knowing about many circumstances does require that a person has some kind of evidence or reason for thinking that it obtains, or some way of finding out about it. Thus when someone asserts that it's going to rain tomorrow, or that Justin is bored, or that the next Government will be socialist, understanding

<sup>13</sup>To make things more digestible I will occasionally talk below as if the view we need to avoid is internalist and evidentialist, but the general thrust of my discussion does depend on this restriction.

<sup>14</sup>For readers familiar with the self-knowledge literature, this puts my account within the family of 'Constitutivist' views. A general worry about some varieties of Constitutivism concerns the extent to which they really explain self-knowledge. Some Constitutivist accounts concentrate on explaining first-person authority, but don't aim to say much about self-knowledge *per se* (as in e.g. Heal, 2002). If the considerations in 3.2 are on the right lines, then such accounts need supplementation. Others aim explicitly to explain self-knowledge, but do not obviously say enough about why a subject who (e.g.) believes that *p* thereby meets the general criteria for *knowing* that she believes that *p* as opposed to merely *truly believing* that she believes that *p*. I raise this worry in relation to Matt Boyle's (2011) Constitutivist account of doxastic self-knowledge in (Campbell, 2018). The account of self-knowledge which I offer here develops and extends the suggestion I make about doxastic self-knowledge at the end of that discussion.

their assertion as expressing knowledge will involve understanding how it is supported by evidence or reasons, or attained via some recognised epistemic method or process (even if in a given case it may for pragmatic reasons be strange to ask for these – see again 2.2). But this is not the case with assertions understood to express self-knowledge. Hence the epistemic asymmetry between the two kinds of assertion.

On the view I will develop, the epistemic asymmetry of self-ascriptions derives from a more fundamental kind of epistemic asymmetry, between two kinds of *knowledge*, one epistemically grounded, the other not. On such a view there is no tension between giving an epistemic account of first-person authority and explaining epistemic asymmetry. On the contrary, both epistemic asymmetry and first-person authority are explained in epistemic terms – in terms of the *kind of knowledge* which self-knowledge is.

But why is it that simply being in some state of mind should in principle be sufficient for knowing that one is in it? Only by answering this question will we be able to understand why it is that a person can know about her own states of mind without reasons, evidence, or method, when these are required for knowing about other kinds of fact. So only by answering this question will we be able to fully understand self-ascriptions' first-person authority and epistemic asymmetry in the way outlined above. It is in answering this question that I think we should appeal to the phenomena of mental state expression and avowal. Unlike the Expressivist, who sought to provide a non-epistemic account of self-ascription, I think that we can understand the distinctive epistemology of self-knowledge in terms of our capacity to express our first-order states of mind.

The explanation I will give depends fundamentally on rejecting the well-grounded true belief account of knowledge and adopting what I will call a 'forward-looking' account, on which knowing is having a certain kind of rational capacity or ability in relation to a fact. I describe this in 4.2, before developing my account of self-knowledge and self-ascription within this framework in the rest of the section.

## 4.2 | Re-thinking knowledge in general

Taking seriously the idea that self-knowledge is not epistemically grounded requires situating it in a general epistemic framework in which reasons, evidence, or epistemic process or method are not essential to propositional knowledge *per se*. Of course we must not deny that knowledge *ever* requires reasons, evidence (or *etc.* – in what follows, I will leave this 'or *etc.*' implicit). I can't know that it is raining on the other side of the world, or that the next Government will be socialist, without good reason for thinking so. What we need, then, is a general account of propositional knowledge which makes room for differently structured knowledge species, some but not all of which are epistemically grounded.

In this context, we should look to what I will call *forward-looking* approaches to understanding propositional knowledge.<sup>15</sup> Contrasting with the familiar model of the essence of knowledge *per se* as epistemically well-grounded true belief, forward-looking accounts understand knowledge in terms of

---

<sup>15</sup>Proponents of forward-looking accounts of knowledge often emphasise that knowledge takes as objects facts rather than propositions. Hyman, for example, prefers to label this knowledge 'factual' (see e.g. Hyman 2015, 159). I agree that the objects of knowledge are facts rather than propositions, but I don't like calling this knowledge 'factual' because I find that this confusingly invites a contrast with 'fictional' knowledge, as opposed, for example, to objectual knowledge. I maintain the label 'propositional' for the relevant form of knowledge, using this to refer to *propositional structure* (which I take it is possessed by facts) rather than metaphysical status.

the rational capacities possessed by knowers.<sup>16</sup> Ryle thought of propositional knowledge as a *capacity to get things right* (Ryle 2000, 128). Recently, more detailed versions of the idea have been developed: John Hyman has argued that knowing that *p* is being *able to be guided by the fact that p* or to *use the fact that p as a reason* (2015, ch. 7), David Hunter (n.d.) suggests that knowing that *p* is *being in a position to use one's rational abilities or capacities* in relation to the fact that *p*, and Stephen Hetherington (2011) argues that knowing that *p* is *knowing how* to do various *p*-related things. A forward-looking approach can clearly accommodate both grounded and groundless species of knowledge. The idea would be having the relevant capacities in relation to some facts (e.g. that the next Government will be socialist) requires having reason to think that they obtain, whereas having these capacities in relation to other facts (e.g. that one's feet are aching or that one wants to leave the party) does not, and can be explained in some other way. I will give an explanation of this latter kind in relation to self-knowledge in 4.3 and 4.4. First, I describe the version of the forward-looking account within which I will be working.

It would take a separate paper to adequately determine the details of the best version of the forward-looking approach, and to defend this against objections. Nor can I be entirely neutral between the various options, since the differences between them run deep, and any neutral account would be overly thin on content. Instead my dialectical strategy is to pick a version of the forward-looking approach which is *prima facie* workable and clear enough in its details, and will enable me to make progress. I will not argue here for this account by contrast either with traditional epistemologies, or other forward-looking accounts, and leave these tasks for future work.

The forward-looking account I will adopt views knowing that *p* as being able to *rationaly respond to or operate with*, the fact that *p* (henceforth simply 'rationally respond').<sup>17</sup> My knowing that today is a Friday, for example, is my being able to rationally respond to the fact that today is a Friday – which I might do in a variety of ways, for example by feeling glad that it is nearly the weekend, or believing that yesterday was Thursday, or intending to go to the pub tonight, and then later on by going there. It will be useful in what follows to have a label for this version of the forward-looking approach. I will call it the Rational Response account of knowledge. Three clarifications about the Rational Response account are in order.

First, when I talk about 'rationally responding' to a fact, the relevant sense of 'rational' is that which contrasts with 'a-rational' as opposed to that which contrasts with 'irrational'. Rationally responding in the relevant sense *includes* 'irrationally' responding; it includes, for example, my responding to the fact that today is a Friday by tapping my head whenever I hear or say the word 'thirteen', or see it written down, or notice a collection of thirteen objects. Clearly irrational in one sense, this nevertheless counts as 'rationally responding' to the fact that today is a Friday in the sense relevant to the Rational Response account of knowledge – in behaving as I do, my response to the fact that today is Friday is broadly in the space of reasons, even though it may be crazy to respond in this way to the fact (and even if I recognise this).

<sup>16</sup>Note that despite its central reliance on notions of epistemic capacity or ability, Virtue Epistemology (as in e.g. Sosa, 2007) is not a forward-looking approach because it understands knowledge as the product of the *exercise* of (and so downstream from) a knower's capacities.

<sup>17</sup>This is closest to Hyman's view, but weaker, since Hyman's 'using a fact as a reason' to my ear sounds more demanding than my 'rationally responding to a fact'. For a collection of objections and responses to Hyman's account see (Hyman 2015, 7.3). Many of these carry over to the Rational Response account, and I consider one of them directly in 5.5. My weaker account avoids at least one problem for Hyman's, however: it seems overly intellectualist to say that a cat which chases a squirrel up a tree can properly be thought to 'use' the fact that the squirrel is up the tree 'as a reason' for climbing it, but not to say that her climbing the tree 'is a rational response' to this fact. I take it that the cat knows that the squirrel is up the tree, and that we need to accommodate this.

Second, there is no particular way in which I need to be able to rationally respond to the fact that today is a Friday in order to know that today is a Friday. What these ways are will depend *inter alia* on what else I know, on my motivations, dispositions, habits, and further abilities and capacities. I know that the bins get collected on Fridays, whereas you who have only just moved to the neighbourhood incorrectly think that they are collected on a Monday. Thus I am able to rationally respond to the fact that today is a Friday by putting the bins out first thing; you are not. But this doesn't impugn your knowledge that today is a Friday, because you are able to rationally respond to the fact in other ways. You might be able to rationally respond to the fact that today is a Friday by doing sixty press-ups, or playing *Flight of the Bumblebee* (such is your training routine). I am not capable of such feats, so can't rationally respond to the fact in these ways. Starker differences still hold between the respective ways in which different animals can rationally respond to facts – a distinctive class of ways in which we humans can rationally respond to facts is opened up by our linguistic capacities: we can *tell one another about* them, for instance. (More on this in 4.4.)

Finally, it is crucial to mark the distinction between being able to rationally respond to a fact and actually rationally responding to it. This is just an instance of the more general distinction between having a capacity or ability and exercising it. Seemingly, any account of knowledge needs to allow that a person knows a vast number of facts at any one point, but cannot be making use of all of this knowledge at that point. I don't lose my knowledge that Birmingham is the UK's second largest city when I am asleep, for example. The Rational Response account views knowing that *p* as a matter of being *able* to rationally respond to some fact, not as actually rationally responding to it. Actually rationally responding to some fact is, in turn, *manifesting, exercising, or expressing* one's knowledge.

Let's turn back to self-knowledge. According to the Rational Response account, knowing that one is in some mental state is a matter of being able to rationally respond to the fact that one is in it. Just as my knowing that today is a Friday is a matter of my being able to rationally respond to the fact that today is a Friday, my knowing that my feet ache is a matter of my being able to rationally respond to the fact that my feet ache – I might do so by e.g. taking my shoes off, or vowing never to wear them again. And my knowing that I want to leave the party is a matter of my being able to rationally respond to the fact that I want to leave the party – I might do so by telling my host I am going to go, or by faking a bomb scare.<sup>18</sup>

### 4.3 | Groundless self-knowledge

I said above that self-knowledge is groundless because being in some state of mind is itself in principle sufficient for knowing that one is in it. In the context of the Rational Response account of knowledge, this now comes to the claim that being in some state of mind is in principle sufficient for *being able to rationally respond to the fact that one is in it*. In this section I will complete my account of self-knowledge by explaining why this should be. This feeds into an explanation of self-ascription and avowal which I give in 4.4.

<sup>18</sup>This is another place where my weakening of Hyman's account improves the account. Hyman cannot allow that leaving the party because I want to is a direct manifestation of my knowledge that I want to leave the party. On his view, I manifest such knowledge by leaving the party only if I use the fact that I want to leave the party as a reason for leaving the party. But typically, when I  $\varphi$  'because' I want to  $\varphi$ , my wanting to  $\varphi$  is not my reason for  $\varphi$ -ing. Nonetheless, I think that  $\varphi$ -ing because one wants to can often, in mature humans, manifest knowledge that one wants to, and is a rational response to wanting to (there are exceptions in the other direction too: sometimes desires just *cause* me to do things, as when I find myself snatching a chip off your plate). This is captured by a conception of knowledge as the ability to rationally respond to a fact, but not by a conception of knowledge as the ability to use a fact as a reason.

There are three parts to my explanation of why being in some state of mind is in principle sufficient for being able to rationally respond to the fact that one is in it. The first is that there is an essential connection between a state's being properly speaking *mental* and the potential for it to be expressed or manifested in the behaviour and broader mentality of the creature whose state it is. The second is that in creatures with certain kinds of rational, agential, and communicative capacities, expressing one's state of mind can itself take the form of a rational response to the fact that one is in it. And the third is that engaging in this form of expression does not in any way depend on the possession of evidence or other epistemic reason for believing that one is in the state, or on any method or process of finding out about the state. In the context of the Rational Response account of knowledge, these three components constitute an explanation of why simply being in some state of mind is in principle sufficient for having epistemically ungrounded knowledge that one is in it.

My starting point draws an essential connection between expressibility and mentality. It is important to be clear about the strength of this claim. I do not claim that any given mental state will in fact be expressed in behaviour or in other aspects of the creature's mentality, or even that a creature in state *M* will be *disposed* to behave or think in *M*-like ways. My claim is only that a mental state (*qua* mental) is in principle *expressible* in these ways. So the idea is not the Behaviourist one that being in some mental state just is being disposed to behave in a certain way. (My claim is not a Behaviourist one for a second reason: it allows that states of mind might be expressed purely in covert aspects of the creature's mentality.) The link between mentality and expressibility therefore isn't implausibly strong. But it does have a kind of strength to it: importantly, the relationship which I think there is between mentality and expressibility is not merely contingent. It is *internal* to a state's being mental that it is in principle expressible, I am claiming.<sup>19</sup>

Second, mental state expression is not itself *essentially* a rational phenomenon (in either of the two senses distinguished above). Mental states can express themselves in one's behaviour in *brute* ways, sometimes even beyond one's ken or against one's will. Consider Freud's Little Hans, whose phobia of horses Freud explains as the product of an Oedipal wish-complex (Freud 1977, 169–303). Taking Freud's explanation seriously means thinking of Hans' Oedipal wishes as expressed in his fearful feelings and thoughts about, and behaviour around, horses and similar animals. But the expression of Hans' wishes in these ways does not itself draw on or co-opt his rational capacities.<sup>20</sup> Other cases in which expression is a largely a-rational affair include automatic or brute behavioural manifestations of sensory, perceptual, or emotional states such as surprise, pain, aggression, anger, or hilarity; the expression of desire (fear) for (of) an object in desirous (fearful) feelings towards that object, or in wishful (fearful) thinking. Much mental state expression in animals will be of this kind: a mouse squeaking in shock; a dog jumping excitedly on hearing the word 'walk', or cowering on hearing the word 'vet' ... and so on.

<sup>19</sup>It ought to be clear that by 'mental' phenomena I mean to include only personal-level phenomena. Sub-personal goings-on are not expressible, but nor are they 'mental' in the relevant sense. They are also not self-knowable.

One way of marking the distinction between personal level – what I here identify as properly mental – phenomena, and sub-personal phenomena, is as a distinction between phenomena which are included in our ordinary or everyday understanding of the mentality of creatures, and phenomena which have been posited or discovered only due to work in empirical disciplines. The former category includes states which are in a certain sense 'unconscious', since our everyday understanding of the mind itself allows that people can (e.g.) be driven by motivations of which they are unaware. This makes sense of the fact that while in a sense 'unconscious', such states are still properly attributable *to the person*. Such states thus contrast with sub-personal states, for example representational states of the visual system.

<sup>20</sup>I say 'does not *itself* draw on ...' here because what Hans does as a result of his phobia may involve rational capacities. He might go to great lengths to avoid places where he'll come across horses, for example. But the immediate manifestation of his wish in his phobia does not itself co-opt his rational capacities.

But the fact that mental state expression can be entirely a-rational in this way does not mean that it never exploits or engages a creature's rational capacities. In creatures with such capacities, mental state expression can take on a distinctively rational form. Consider an example. My partner has yet again unfathomably left a heavy object in an exposed region of the living room floor. Predictably, I stub my toe. My pain may well express itself in the brute kind of way described above – I simply yelp or gasp, in what Candace Vogler nicely calls an “eruption” of pain or annoyance.<sup>21</sup> But, especially if my partner is in ear- or eye-shot, my behaviour might take on an other-regarding, and at least minimally communicative, shape or dimension. I yelp *pointedly*, at him or *in his direction*.

My brute yelping and my pointed yelping equally express my pain.<sup>22</sup> To characterise a behaviour as an *expression* (or as expressive) of some psychological condition is to characterise it as *coming from* that condition – from my pain in this case – and this is what both yelps do. But characteristic of my pointed yelping in the second case, by contrast with my brute yelping in the first, is not only where it has come from, but also – so to speak – where it is going. It is directed at my partner, perhaps at his moving the offending object, or apologising, or seeing that he has behaved carelessly – but at the very least at his recognising that I am in pain.

And this makes it clear, I think, that my pointed yelping is not only a reaction to or manifestation of my pain, but is a response to the fact that I am in pain, which is broadly in the space of reasons, and normatively governed. If asked for a full explanation of why I was staring so hard at my partner and frowning while I yelped as I did, part of my answer would cite factors which made rational sense of this: I was in pain, and wanted him to know as much. If my yelping went unheard or unregistered *as expressive of pain*, it would have failed in its communicative aim. With communication comes the possibility of warning, sharing, scolding, and so on. And these are acts in the space of reasons, at least potentially done in order to achieve further effects, or *qua* fitting responses to the circumstances. These features of my pointed yelping show that while it is not a particularly rationally sophisticated response (it is not thought-through, for example, or done carefully or skilfully), it is still a *rational* response, in the sense I have described (opposed to ‘a-rational’ rather than ‘irrational’), because it at least minimally co-opts my broadly rational capacities: it is intentional, and communicative.

According to the Rational Response account of knowledge, this means that my pointed yelping is not only an expression of *my pain*, but is in addition an expression of *knowledge that I am in pain*. The Rational Response account says that knowing that I am in pain is being able to rationally respond to the fact that I am in pain. What I have just been bringing out is that my pointed yelping is itself an exercise of this ability. The general picture of self-knowledge suggested by considering this particular example is as follows: being in some state of mind is in principle sufficient for knowing that one is in it because a) states of mind *qua* mental are essentially in principle expressible, and b) for creatures like us, expressing a mental state can itself take the form of rationally responding to the fact that one is in it – this is the case when the expression of a mental state is structured by the exercise of our general rational, agential, and communicative capacities.

It is important to be explicit about why, on this picture, self-knowledge should be ungrounded in (e.g.) evidence. Not only is it clear that my pointed yelping is (*inter alia*) a rational response to the fact that I am in pain, but it is also clear that my ability to respond to the fact in this way does not itself depend on my possession of reason or evidence in support of the proposition that I am in pain, or any process or method of discovery. It is true that my behaviour (as it were) jumps up an epistemic

---

<sup>21</sup>Vogler's phrase is “eruption of feeling” (2002, 234), and refers specifically to expressions of emotion.

<sup>22</sup>The latter also expresses my anger or exasperation. I leave this to one side to keep the focus simple.

level – brute yelps (etc.) express my pain alone whereas pointed yelps (etc.) express both my pain and knowledge that I am in pain. But the possibility of my pointedly yelping no more depends on my having evidence to the effect that I am in pain than does that of my brutally yelping.

Furthermore, even if I were to be provided with evidence for thinking that I was in pain – if I were to be handed a helpful confirmatory brain-scan – this would make no difference whatsoever to my capacity to express my pain in a pointed yelp. This means that it would make no difference whatsoever to my ability to make this kind of rational response to the fact that I am in pain, and so that it would make no difference to my possession of self-knowledge that I am in pain.

The account I am developing thus helps us to see both that, and why, the possession of an epistemic basis is completely beside the point in relation to self-knowledge, even when available. On this picture, the epistemic asymmetry between self-knowledge and knowledge of other kinds runs very deep indeed.

#### 4.4 | Knowledgeable self-ascription and avowal

So far, my account of self-knowledge has focused on the fact that creatures like us have various general rational, agential, and communicative capacities, which can be co-opted in the expression of one's mental states, rendering these expressions also rational responses to the fact that one is in the expressed states. But some of these very general capacities are, of course, linguistic. And one of the forms taken by a mature human's rational response to the fact that she is in some state is a linguistic utterance, where this class includes explicit self-ascription.<sup>23</sup> Imagine, then, that after or instead of yelping, I say "My toe really hurts!"

For the first time in our story my linguistic capacities are brought into play in the form taken by my rational response to the fact that I am in pain. In this respect my behaviour is more sophisticated than was my pointed yelping, because it draws on a more sophisticated set of capacities, including now linguistic ones. But for all this my behaviour still expresses my pain. My self-ascription "My toe hurts" is – just like my pointed yelp – an expression of my pain which is also a rational response to the fact that I am in pain. To say this is to classify my self-ascription as an avowal. But whereas the Expressivist replaced the idea that self-ascriptions express self-knowledge with the idea that they are avowals, it is clear how on the view I am developing, my self-ascription expresses self-knowledge *because* it is an avowal.

In this case, my linguistic capacities make a difference to the way in which my pain is expressed. But they do not make a difference to the epistemology of things. The fact that I express my pain now in a way which co-opts my linguistic capacities does not suddenly introduce a need for the self-knowledge I thereby manifest to be based on evidence – where when I expressed self-knowledge non-linguistically (in a pointed yelp), this was evidence-free. The ability to rationally respond to the fact that I am in pain does not depend on my possessing evidence for thinking that I am in pain, whether I exercise this ability by pointedly yelping, or by saying "My toe hurts". What it does depend on is my being able to exercise my general rational, agential, and communicative capacities in structuring the

<sup>23</sup>Because I have explained how self-knowledge is expressed in non-linguistic behaviours (pointed yelping) first, and only then moved on to explaining how it is expressed in explicit self-ascription, I may seem committed to a view on which self-knowledge is independent of the possession of language. But this needn't be the case. It will not be the case if the rational and communicative capacities exercised in pointed yelping can themselves be possessed only by someone who also has linguistic capacities. A detailed investigation of the relationship between the capacity for self-knowledge and for language goes beyond the scope of this paper.



way my pain is expressed. The case of avowal is simply the special case in which the rational capacities whose exercise structures the expression of a state include linguistic capacities, exercised in the form of an explicit self-ascription.

My pointed yelping was understood above to express groundless self-knowledge in virtue of being an exercise of the ability to rationally respond to the fact that I am in pain, one which itself does not (indeed *could* not) depend on evidence for thinking that I am in pain. And the same can be said about my avowal. Avowals are like other forms of expressive behaviour which, in virtue of co-opting one's broadly rational, agential and communicative capacities, are also rational responses to the fact that one is in the state. The latter can take more or less sophisticated forms (asserting "I'm in pain" or pointedly yelping), but insofar as they *express* the relevant states, there is simply no room for them to be based on evidence or any other kind of reason.

I have been trying to bring out the variety of ways in which a person's behaviour can be expressive of some mental state of hers. At each level of our example, a more sophisticated set of capacities is co-opted in the expression of my pain. In all but the first of these levels (brute yelping), I am rationally responding to the fact that I am in pain. In the context of the Rational Response conception of knowledge, this means that in all but the first of these levels, I am not only expressing my pain, but am also expressing *knowledge that I am in pain*. Having groundless knowledge that I am in pain is a matter of *being able to* express my pain in a way that is also a rational response to the fact that I am in it. And as I have been trying to make plausible, someone with the relevant general capacities will in principle have this ability in relation to a given state of mind just in virtue of being in it – because insofar as she has these general rational capacities, she will be able to exercise them in the rational expression of this state.

I complained in section 3 that Expressivists have either not wanted or have not been able to explain why avowals express a substantive form of knowledge. On the view I have described, any expression of a state of mind which is also a rational response to being it is will manifest self-knowledge. Avowals are the special case in which this rational response takes the form of an explicit linguistic self-ascription.

## 4.5 | Summing up

Let me sum up the discussion of this section. We saw earlier that we should try to hang on to the epistemic approach to understanding self-ascriptions' first-person authority, but that the approach will only work if it can also accommodate their epistemic asymmetry with other forms of assertion. I suggested that in order to do so we need to develop a view on which self-knowledge is genuinely epistemically ungrounded. I explained that we can do all of this if we accept that simply being in some state of mind is in principle sufficient for knowing that one is in it, but emphasised that this strategy will only work to the extent that we understand how this can be the case, given that the mere obtaining of facts is not generally sufficient for anyone to know about them.

I then argued that we can make sense of self-knowledge by adopting a forward-looking general account of knowledge. Specifically, according to what I called the Rational Response account of knowledge, the question of why being in some state of mind is in principle sufficient for knowing that one is in it becomes the question of why simply being in that state is in principle sufficient for being able to rationally respond to the fact that one is in it. What I tried to make plausible in 4.3 is that in suitably capacitated creatures, states of mind can be expressed in a form which is also a rational response to the fact that one is in them, and that the ability to engage in such rational responses does not (indeed *could not*) itself depend on an epistemic basis.

Finally, we saw that expressions which are rational responses can take linguistic as well as non-linguistic forms. One way in which we can exercise our linguistic capacities in the expression of our states of mind is by explicitly self-ascribing them. Because such avowals are expressions of the first-order states, they are not (again, could not be) epistemically grounded. But insofar as they are rational responses to the fact that one is in the first-order state, they manifest a genuine form of knowledge.

## 5 | CLARIFICATIONS, OBJECTIONS, REPLIES

The picture I have sketched requires some refinement, and there are some objections I would like to respond to. I turn to this now. I tackle what might seem the most substantial worry for my account – that it presupposes what it sets out to explain – last, in 5.5.

### 5.1 | Being in some mental state as ‘in principle’ sufficient for knowing that one is in it

What is the force of the ‘in principle’ in my claim that being in some mental state is in principle sufficient – for creatures with certain kinds of rational, communicative, and agential capacities – for knowing (being able to rationally respond to the fact) that one is in it?

The qualification of course makes room for cases in which a person fails to know that she is in some state. There are several kinds of case. I considered some when objecting to Eliminativist and Deflationist Expressivism in 3.2, but there are others. I may be self-deceived about certain of my intentional states. My attention might be taken over by a sudden circumstance (**A BEAR!!!!**), during which period I wouldn’t be able to say who I believed the next Government would be if I tried.<sup>24</sup> I have only the vaguest idea, and certainly no detailed knowledge, of the kinds of unconscious features of my psychology responsible for my various neuroses ... and so on. Some of these cases seem to be ones in which I *lack* self-knowledge; others cases in which my self-knowledge is inaccessible (perhaps wholly, or temporarily, or unless and until I am prepared to let myself access it). On an account of knowledge as the ability to rationally respond to a fact, this distinction can be understood as the distinction between lacking the knowledge ability and having it while it is for some reason masked.

It would take more work than I can do here to adequately detail the kinds of factors which can undermine or mask the knowledge ability in relation to facts about a person’s own states of mind. But the fact that there is more to say here does not undermine the claim that in a range of familiar cases, being in the state in question is sufficient for being able to rationally respond to the fact that one is in it.

Importantly, part of my idea involves holding that the explanatorily basic case will be that in which a person in possession of the ordinary gamut of human rational, communicative, and agential capacities has, and is able to exercise, these in relation to a broad range of her states of mind. What will need explaining, for such a creature, will be the cases in which these capacities cannot be exercised in relation to a given state, not the cases in which they can.

And I think this is a general feature of explanation in terms of ability or capacity. Compare: if we know that I have the capacity to pick up objects, then faced with a particular object, we don’t need to add further explanation in order to understand why I am able to pick *this object* up, if I am. We would

<sup>24</sup>I owe this “**A BEAR!**” example to Paul Snowdon’s wonderful undergraduate lectures on the later Wittgenstein, which I attended in London circa. 2008.

need to engage in further explanation if, having the general capacity to pick up objects, and faced with a particular object, I am *unable* to pick it up: the explanation might be that the object is too large or heavy (here I lack the particular ability to pick *this thing* up even though I have the general ability to pick things up), or that I've got my hands tied behind my back (here I *have* the particular ability to pick this thing up, but this ability is masked).

## 5.2 | Self-knowledge is prior to self-ascription

According to the view I have described, some of the behaviours in which we express our states of mind also express knowledge that one is in those states. The behaviours in question are ones which are both expressions of, and rational responses to the fact that one is in, the relevant state. I am not claiming that my knowledge that I want to leave the party, or that my feet ache, depend on expressing these states, whether in an avowal or any other overt or covert behaviour. I am identifying my knowledge that I am in these states with the *ability* or *capacity* to rationally respond to the fact that I am in them, not with my exercise of this capacity.

And I think that in this respect my account improves on an otherwise closely related approach developed by Johannes Roessler, according to which one might know about one's states of mind "in expressing them" (e.g. 2015, 157).<sup>25</sup> Rather generally, I think that we think of knowledge as prior to its expression, manifestation, or exercise, and that this is hard to maintain on an account like Roessler's.

This does not mean that a person never comes to know about, for example, her belief on some matter in or by expressing or self-ascribing that belief. But often – I want to say – one self-ascribes some state of mind for the first time without thereby *coming* to know that one is in it. In such cases I know about that state of mind prior to, and independently of, self-ascribing it, and my self-ascription expresses or manifests my knowledge. This natural idea is better understood within an account of self-knowledge of mental states in terms of the subject's ability to express them (my view) as opposed to an account of self-knowledge in terms of the subject's exercise of such an ability (Roessler's).

## 5.3 | Directly rationally responding to the fact that one is in M

Self-ascriptions are not *per se* expressive of self-knowledge, either pre-theoretically, or on my account. Again, not all knowledge of one's own states of minds is self-knowledge – we have been reserving that label for knowledge which appears especially easy to come by, ungrounded in evidence or reasons, and unbacked by any clear method or process of discovery.

On my account, a self-ascription will be expressive of *knowledge* that one is in the self-ascribed state when it is a rational response to the fact that one is in that state (that is, when it manifests the ability to rationally respond to the fact that one is in that state). A self-ascription is expressive of *self-knowledge* that one is in the self-ascribed state when it is a *direct* rational response to the fact that one is in that state – that is, when it is a rational response to the fact that one is in that state which is not

<sup>25</sup>In other respects the account I have given is in agreement with Roessler's, in particular in diagnosing traditional difficulties in understanding self-knowledge as stemming from the problematic identification of knowledge and well-grounded true belief, and in developing an account which rejects this assumption. Roessler's view and mine are equally "akin to [Expressivism] in giving a central role to the notion of expression", and in "appeal[ing] to the sense in which self-ascriptions may be taken to express the speaker's first-order mental states [in] providing, not an alternative to, but a way to develop, an "epistemic" explanation of first-person authority" (Roessler 2015, 155). See also (Roessler, 2013).

mediated by the possession of evidence for thinking that one is in that state, or the taking of a view on whether one is in that state, or by anything that we might call *finding out* or *working out* that one is in it.

We can think of avowals as just such *direct* rational responses. Avowing “My feet ache” is in this sense continuous with preparing a foot-bath. This kind of direct response to the fact that one is in some state – a response which one can make without having any *evidence for thinking that* one is in the state in question – is available only in relation to one’s own states of mind. But this does not mean that the kind of third-person knowledge I can have of someone else’s mind isn’t also available in relation to some of my own states. Sometimes *only* this kind of knowledge might be available. In such cases, I can self-ascribe a state and do so knowledgeably, but this will not be *self-knowledge*. In such cases, I can rationally respond to the fact that I am in the state, but only indirectly – by forming a view about my own mind in response to the evidence or as a result of some process of finding out about it.

#### 5.4 | Are direct rational responses really ‘expressions’?

An objection might be made to the characterisation of my account as one on which self-knowledge is explained in terms of the phenomenon of mental state expression. I have classified avowals as coming in the class of direct rational responses to the fact that one is in some state of mind. But is a rational response to the fact that I am in pain (pointedly yelping, getting painkillers, or avowing my pain) really an *expression* of pain in anything like the sense that a brute yelp is an expression of pain? Is a rational response to the fact that I want to leave the party really an *expression* of this desire in anything like the sense in which a brute expression of desire, such as experiencing an intense feeling of claustrophobia when the host indicates she’s up for a late night, is an expression of it?<sup>26</sup>

My answer to this is: yes and no. I have not claimed, and do not need to claim, that getting painkillers or avowing “I’m in pain” are expressions of pain in the very same sense that brutally yelping is an expression of pain. On the contrary, one is a brute manifestation of my pain, the other a rational response to the fact that I am in pain. I do however think that one way in which we use words like ‘express’ or ‘manifest’ is to mark what I above called direct rational responses to the fact one is in some state, and to distinguish these from the kind of indirect response one might have to such a fact when this is mediated by ‘finding out’ or responding to evidence. If Hans comes to know (by being told) of his Oedipal wishes, he might be able to knowingly self-ascribe them, thus rationally responding to the fact that he has them. But he will not thereby be *expressing* or *manifesting* them in any sense. Compare his fear of horses, which is something he is able to rationally respond to directly – again, without going via any process of finding out, or attention to reasons or evidence for its existence. When he says “Papa, I’m scared”, and flinches when a carriage approaches on the road, we do, I think, view his self-ascription as expressive of his fear even though it is not a *brute* expression like his flinching.

If a reader insists on retaining the label ‘expression’ for what I have called ‘brute expressions’, then I won’t spend too much time arguing about it. The crux of my account does not lie in the fact that we call these rational responses ‘expressions’, but in the fact that appropriately capacitated creatures can go in for them, in principle just in virtue of being in a state of mind, and *without need for any judgmental or epistemic intermediary* (what I have called ‘directness’). In the context of the Rational Response conception of knowledge, we can take or leave the label ‘expression’ for these rational responses.

I prefer to take the label, because I think it is perfectly apt for the relevant phenomenon. In addition, I have wanted to mark the fact that although I have rejected Expressivism, because I have rejected the

<sup>26</sup>Thanks to Crispin Wright (p.c.) for pressing this worry.

non-epistemic approach to understanding self-ascriptions, my account nevertheless owes a lot to the Expressivist tradition.<sup>27</sup>

## 5.5 | A worry about circularity

The final objection I will consider is the most serious. If it is a genuine problem then it threatens to undermine my entire approach rather than only requiring me to clarify certain details, or to adjust my labelling.

The objection is that the fact that a person is able to (directly) rationally respond to the fact that she is in some state *depends on* or *is explained by* the fact that she knows that she is in it, so that we don't explain (e.g.) my knowledge that I want to leave the party by pointing to the fact that I am able to rationally respond to this fact. The worry is that my account is circular, so ultimately unexplanatory.<sup>28</sup>

To clarify the concern, the worry can't simply be that knowing that my feet ache or that I want to leave the party is *necessary* for rationally responding to the fact that my feet ache, or that I want to leave the party. Far from being a problem for my view, this is one of its central commitments, given that it identifies knowledge with the ability to rationally respond to a fact. Rather, the worry seems to depend on the idea that my knowledge that I want to leave the party is somehow more explanatorily fundamental than my ability to rationally respond to this fact, so that in explaining the former in terms of the latter, I am getting things the wrong way round.

But even if knowledge is in some sense more explanatorily fundamental (and I question this below), I don't think this poses a problem for the explanation of self-knowledge that I have offered. This is because neither the general explanation of knowledge in terms of the ability to rationally respond to a fact, nor the specific explanation of self-knowledge in terms of the ability to (directly) rationally respond to a fact about one's own state of mind, are supposed to be reductive explanations. Positively, knowledge is best thought of as being *elucidated* as the ability to rationally respond to a fact. An elucidation of knowledge is a description of the *character* of the knowledge-relationship which is not at the same time a description of the *foundations* of this relationship. There is no commitment here to any abstract claim about explanatory fundamentality in either direction. In the context of our discussion it has been useful to explain – or perhaps better, to understand – knowledge as the ability to rationally respond to a fact, rather than *vice versa*. But the fact that I have understood knowledge in terms of this ability rather than the ability in terms of knowledge does not entail that the ability is 'explanatorily fundamental' relative to knowledge, in any way which goes beyond our specific explanatory needs in the current investigation.

In the context of my investigation in this paper, what we needed to understand – as I set things up in 4.1 – was how being in some mental state could in principle be sufficient for knowing that one is in it, independently of any epistemic method or basis. I have tried to explain these characteristic features of self-knowledge by saying something about the structure of knowledge in general, and then explaining why this structure should be manifested by the self-knower in a way which does not depend on her having an epistemically well-grounded belief about her own mental state. The explanation trades on the idea that the structure of knowledge is that of the ability to rationally respond to a fact. It does not

---

<sup>27</sup>As developed, my account owes a lot to the Expressivist idea that expressibility is (something like) a or the 'mark of the mental'. Of course if we give up on the label 'expression' for the kind of direct rational response I have described, we would also have to give up calling the mark of the mental 'expression'. But we would not have to give up on the substance of the idea which underlies it.

<sup>28</sup>Hyman discusses this kind of objection as it arises in relation to his conception of knowledge (in general) as the ability to be guided by the facts (see Hyman 2015, 181–84). If I understand him, our responses are broadly in line.

require either assuming that we can reduce knowledge to the ability to rationally respond to a fact, or denying that knowledge is more explanatorily fundamental than this ability.

But although I don't think my account of self-knowledge requires denying that knowledge is more explanatorily fundamental than the ability to rationally respond to a fact, I also think it's worth questioning the presupposition that knowledge *is* more explanatorily fundamental. I suspect that the idea of different levels of explanatory fundamentality which are not relative to a given explanatory context might very generally be a questionable one. And even if we can grant a context-free sense for the notion of explanatory fundamentality, we might still question the claim that *knowledge* is more explanatorily fundamental than *being able to rationally respond to a fact*. Can we really make sense of someone as *knowing* something without being able to *make use of* this knowledge in the sense of being able to respond to the fact it is knowledge of? I suspect not. And if not, then a conception of the ability to rationally respond to a fact is already implicit in our conception of knowledge, so that it might be better to think of these phenomena as explanatorily on a par.

I think that these issues of methodology and explanation are important and complex, and doubtless they require more discussion than I have had room for here. But I hope I have said enough to give a sense of why I don't think that there is any straightforward circularity worry about the account of self-knowledge I have offered.

## 6 | CONCLUSION

In this paper I have argued against Expressivism as an approach to understanding mental state self-ascriptions, and proposed my own positive account of self-ascription, avowal, and self-knowledge.

The feature of Expressivism which caused trouble was its fundamental insistence on giving a non-epistemic account of self-ascriptions' first-person authority. I argued that such a view was both unattractive and unmotivated. But I think that the Expressivist is right to think that properly understanding self-ascriptions requires taking seriously their epistemic asymmetry with other forms of assertion, and also right to think that doing so requires emphasising their status as avowals; as directly expressing the states they self-ascribe, rather than the self-ascriber's point of view on the state she is in.

The account I have developed is one on which self-ascriptions' status as avowals explains why they manifest self-knowledge, understood as itself genuinely epistemically groundless. Both self-ascriptions' first-person authority and their epistemic asymmetry are thus explained in terms of self-knowledge: it is because self-knowledge is a distinctive, genuinely ungrounded, form of knowledge, that self-ascriptions contrast with other forms of assertion in these idiosyncratic ways.

So understanding self-ascriptions requires understanding groundless self-knowledge. The view I have developed of the latter is one on which being in some state of mind is in principle sufficient for knowing that one is in it, because it is in principle sufficient for being able to rationally respond to the fact that one is in it – at least for creatures with the rational, agential, and communicative capacities which we humans have. When these capacities are co-opted in the expression of a state – whether in a relatively unsophisticated way exemplified by a pointed yelp directed at a careless partner, or more sophisticatedly in an overt linguistic self-ascription of pain – the expression of a state of mind is also a rational response to the fact that one is in it. The ability to go in for these kinds of rationally structured expressions of our states of mind does not require the possession of evidence or reason to think that one is in the state, nor any method or process of finding out about it. Indeed, I have argued that there is simply no work for such phenomena to do in grounding self-knowledge.

I have thus tried to show that properly understanding self-knowledge and self-ascription requires paying serious attention to mental state expression, but also that it requires doing away with Expressivism.

**ORCID**

Lucy Campbell  <https://orcid.org/0000-0001-8778-524X>

**REFERENCES**

- Austin, J. L. (1961). Other Minds. In J. Urmson, & G. Warnock (Eds.), *Philosophical Papers*, edition. (pp. 76–116). Oxford: Oxford University Press.
- Bar-On, D. (2004). *Speaking My Mind: Expression and Self-Knowledge*. Oxford: Oxford University Press.
- Biederman, I., & Margaret M. Shiffrar (1987). Sexing Day-Old Chicks: A Case Study and Expert Systems Analysis of a Difficult Perceptual Learning Task. *Journal of Experimental Psychology*, 1, 640–45. <https://doi.org/10.1037/0278-7393.13.4.640>
- Boyle, M. (2011). Transparent Self-Knowledge. *Aristotelian Society Supplementary Volume*, 85, 223–241. <https://doi.org/10.1111/j.1467-8349.2011.00204.x>
- Campbell, L. (2018). Self-Knowledge, Belief, Ability (and Agency?). *Philosophical Explorations*, 21, 333–49. <https://doi.org/10.1080/13869795.2018.1426779>
- Coope, C. (1973). Wittgenstein's Theory of Knowledge. *Royal Institute of Philosophy Supplement*, 7, 246–67. [https://doi.org/10.1007/978-1-349-15546-0\\_15](https://doi.org/10.1007/978-1-349-15546-0_15)
- Finkelstein, D. H. (2003). *Expression and the Inner*. Cambridge, MA: Harvard University Press.
- Freud, S. (1977). *Case Histories I: "Dora" and "Little Hans"*. Edited by Angela Richards. Translated by Alix Strachey and James Strachey. The Pelican Freud Library 8. London: Penguin Books.
- Hacker, P. M. S. (1986). *Insight and Illusion: Themes in the Philosophy of Wittgenstein*, Revised edition. Oxford: Oxford University Press.
- Heal, J. (2002). The Presidential Address: On First-Person Authority. *Proceedings of the Aristotelian Society*, 102, 1–19. <https://doi.org/10.1111/j.0066-7372.2003.00040.x>
- Hetherington, S. (2011). *How to Know: A Practicalist Conception of Knowledge*. Chichester: Wiley-Blackwell.
- Hunter, D. n.d. On Being a Believer (ms.). Toronto.
- Hyman, J. (2015). *Action, Knowledge, and Will*. Oxford: Oxford University Press.
- Kemmerling, A. (2012). First Person Authority without Glamorous Self-Knowledge. *Epistemology: Contexts, Values, Disagreement - Papers of the 34th International Wittgenstein Symposium*, 401–29. <https://doi.org/10.1515/9783110329018.399>
- Kenny, A. (1973). *Wittgenstein*. London: Penguin Books.
- Roessler, J. (2013). The Silence of Self-Knowledge. *Philosophical Explorations*, 16, 1–17. <https://doi.org/10.1080/13869795.2013.744084>
- Roessler, J. (2015). Self-Knowledge and Communication. *Philosophical Explorations*, 18, 153–68. <https://doi.org/10.1080/13869795.2015.1032326>
- Ryle, G. (2000). *The Concept of Mind*. London: Penguin Books Ltd.
- Sankowski, E. (1978). Wittgenstein on Self-Knowledge. *Mind*, 87, 256–61. <https://doi.org/10.1093/mind/LXXXV.2.256>
- Sosa, E. (2007). *A Virtue Epistemology: Apt Belief and Reflective Knowledge*, Vol. I. Oxford: Oxford University Press.
- Vogler, C. (2002). *Reasonably Vicious*. Cambridge, MA: Harvard University Press.
- Wittgenstein, L. (2003). *Philosophical Investigations: The German Text, with a Revised English Translation*. Malden, MA: Blackwell Pub.
- Wright, C. (1998). Self-Knowledge: The Wittgensteinian Legacy. In C. Wright, B. C. Smith, & C. Macdonald (Eds.), *Knowing Our Own Minds* (pp. 13–45). Mind Association Occasional Series. Oxford: Clarendon Press.

**How to cite this article:** Campbell L. Self-Knowledge: Expression without Expressivism. *Philos Phenomenol Res.* 2020;00:1–23. <https://doi.org/10.1111/phpr.12739>