Review of Philosophy and Psychology
https://doi.org/10.1007/s13164-023-00698-6

Partial First-Person Authority: How We Know Our Own Emotions

Adam J. Andreotta[1]

Accepted: 28 August 2023

Abstract
This paper focuses on the self-knowledge of emotions. I first argue that several of the leading theories of self-knowledge, including the *transparency method* (see, e.g., Byrne 2018) and *neo-expressivism* (see, e.g., Bar-On 2004), have difficulties explaining how we authoritatively know our own emotions (even though they may plausibly account for sensation, belief, intention, and desire). I next consider Barrett's (2017a) empirically informed *theory of constructed emotion*. While I agree with her that we 'give meaning to [our] present sensations' (2017a, p.26), I disagree with her that we construct our emotions, as this has some unwelcome implications. I then draw upon recent data from the science of emotions literature to advance a view I call partial first-person authority. According to this view, first-person authority with respect to our emotions is only partial: we can introspect and authoritatively know our own sensations, and beliefs, in ways others cannot; but we still need to interpret those sensations and beliefs, to know our emotions. Finally, I consider self-interpretational accounts of self-knowledge by Carruthers (2011) and Cassam (2014). I argue that while these accounts are implausible when applied to beliefs, desires, and intentions, they are more plausible when applied to our emotions.

Keywords Self-knowledge · Privileged Access · First-Person Authority · Transparency · Emotions

1 Introduction

It is a commonly held presumption that when it comes to our own mental states, each of us are in an epistemologically privileged position. Even though we can be mistaken, in the normal situation when someone genuinely states that they are tired, professes a belief that house prices are about to rise, or claims that they are anxious about an upcoming seminar presentation, they are taken to be in those mental

✉ Adam J. Andreotta
adamandreotta@outlook.com

[1] Curtin University, Kent St, Bentley, WA 6102, Australia

Published online: 19 October 2023

states.[1] Our mental state self-attributions are not normally challenged by other people either. We seem to have first-person authority (hereafter, 'FPA') with respect to them.

The topic of FPA, sometimes also referred to as 'privileged access' (see, e.g., Alston 1971; Gertler 2003; Byrne 2018), has received much attention in the last few years.[2] Some have attempted to explain FPA in terms of a faculty of inner sense, or internal scanner, that we each possess (see, e.g., Lycan 1996). Others have maintained that FPA is explained by our ability to transparently know our mental states by attending to the features of the world that our mental states are about (see, e.g., Moran 2001; Byrne 2005, 2018; Fernández 2013). Others, still, claim that FPA can be explained on the basis of our ability to self-express our mental states (see, e.g., Bar-On 2004; Finkelstein 2003). And some even reject the thesis that we have FPA, with respect to most of our mental states (see, e.g., Carruthers 2011; Cassam 2014).

In this paper I argue that even if some of the leading theories of self-knowledge can offer plausible accounts of how we know, and have FPA of, some categories of our mental states (such as our beliefs and sensations), they do not map well to our emotions. I advance a view I call *partial first-person authority*, which says that the authoritative way in which we know our emotions is only partial. While we may be able to know our own sensations and beliefs in a first-person authoritative and non-interpretative way, we still need to interpret those mental states, in order to know what emotional states we are in. I call this view partial first-person authority because *some* interpretation occurs when we attempt to know our own emotions, just like when we attempt to know others' emotions.[3] In defending this view, I am rejecting the thesis that self-knowledge of all the different categories of mental states should be explained in the same way, a position exemplified in Byrne's (2018) and Bar-On's (2004) work. An account of how we know and have FPA of our beliefs, for example, need not also apply to our emotions.[4]

I begin, in Section 2, by describing the concept of FPA. In Section 3, I argue that Byrne's (2018) attempt to extend the transparency method to emotions is unsuccessful. In Section 4, I consider Bar-On's (2004) neo-expressivist account. I argue that the view faces a difficulty in explaining emotion self-attribution errors. In Section 5, I offer some critical remarks on Barrett's (2017a) theory of constructed emotion. Finally, in Section 6, I articulate the partial FPA view of emotions.

[1] There are, of course, times where we do not take a subject's first-person self-attribution seriously, such as in pathological cases, or in cases where a person is trying to deceive someone (cf. Parrott 2015, p. 2216).

[2] See Gallois (1996) for a discussion about why it may be useful to use the terms 'first-person authority' and 'privileged access' distinctively.

[3] By the term 'direct access', I mean access which is non-interpretative in nature. For example, I have direct access to my pain, whereas another person must interpret my behaviour or avowals to know I am in pain.

[4] Here I follow Coliva (2016) in rejecting the view that the different types of self-knowledge should be explained in the same way.

2 Emotions, Introspection and First-Person Authority

Emotions are central to our lives. We fear failure, we are disgusted by spoiled food, we are sometimes jealous of others' successes, and feel anxious before public speaking events. We are also easily able to identify these emotions in ourselves. A person who is nervous before a public speaking event may attempt the difficult task of calming their nerves by practicing deep breathing exercises or rehearsing their opening line over and over. The task of knowing that they are nervous, on the other hand, seems trivially easy.

While many of us are adept at knowing what emotions we are currently experiencing, it is much harder to characterise their precise nature. In the literature on the emotions, for example, it has been argued that emotions are evaluative perceptions (Prinz 2004a; Roberts 2003), evaluative feelings (Goldie 2000), evaluative judgements (Solomon 2004a; Nussbaum 2001) and that they are the brain's way of making sense of our bodily sensations (Barrett 2017a).[5] Even though my concern here is the FPA of emotions, and not specifically the nature of emotions, the question of what emotions are is difficult to completely ignore. For example, if it is true that we construct our emotions as Barrett (2017a) suggests, then this has implications for FPA, since this might make our emotional self-attributions particularly authoritative. Similarly, if one were to argue that emotions are reducible to sensations, then self-knowledge accounts of sensation will be applicable to them.[6] My aim here is not to completely ignore the question of what emotions are, but rather give an account of FPA with as few theoretical commitments as possible.

The sense of FPA that I attempt to explain here pertains to a common-sense notion that I hope will strike many as intuitive.[7] The two main components I have in mind are as follows:

> [1A] Epistemological asymmetry: We know our own mental states (e.g., I believe that *P;* I am in pain*;* I intend to φ) in a way that others cannot. We can introspect, or have direct access to, our mental states; whereas others must observe our behaviour or listen to what we say, in order to know our minds.
> [1B] Epistemological security: First-person mental state self-attributions (e.g., I believe that *P;* I am in pain*;* I intend to φ) are more likely to amount to truth, compared to the mental state attributions others make about us.

It is important to note that I am using the term 'introspection' here in a theory-neutral sense that applies, very broadly, to ways of knowing one's own mental states that are asymmetrical to the ways that others come to know them. Some authors may protest that this usage is too broad, and that the term 'introspection' should be

[5] See Solomon (2004b), for collection of contemporary philosophical theories of emotions.

[6] Collier (2011) argues that some have mistakenly attributed this view to Hume. See Collier (2011) and Prinz (2004a pp.11-12), who offer a more nuanced account of Hume's position.

[7] There are other ways of characterising FPA that I will not consider here. For a description of over 30 of these, see Alston (1971) who mentions infallibility, incorrigibility, and self-intimation, among others. For an argument against applying concepts like infallibility and self-intimation to our self-knowledge of emotions, see Jäger (2009).

reserved for internal scanning views. Akeel Bilgrami, for example, says that '[t]he verb 'to introspect' suggests a kind of cognition… which tilts the usage in favor of the perceptualist [inner sense] view' (2006, p. 38). Similarly, Dorit Bar-On characterises introspection as an 'inner gaze, or internal monitoring, or scanning' (2004, p. 103). In contrast to these authors, I will use the term 'introspection' in the broad sense because it gives us a way to discuss the unique way we know our own minds without referring to any specific view in the literature.

Consider my belief that I am in pain. It seems that I can know that I am in pain in a way that others cannot—namely, by introspection (broadly construed), as stated in [1A]. From the first-person perspective, I have direct access to the qualitative feeling of pain, whereas another person must observe and interpret my behaviour. My belief that I am in pain is more likely to be true, as stated [1B], because I can ground it with the sensation of being in pain. Others can, of course, form true beliefs that I am in pain, and they may be able to overturn my belief that I am in pain, but they must rely on my behaviour and speech to do so.

There is, then, an important relationship between introspection and FPA. If you can introspect your pain, in a non-interpretative way, then you can justify your belief that you are in pain in a way that no one else can. Others can form beliefs that you are in pain, of course, but they will lack FPA because they need to interpret your behaviour and speech. Only you can ground your belief that you are in pain with the felt experience of being in pain. Similarly, if you can introspect—again in the theory neutral sense—what you believe, then you can justify your higher-order belief that you have a believe that *P* in a way others cannot. FPA can thus be explained on the basis of introspection. I am not the only one to make this connection between introspection and FPA. Smithies (2012), for example, has advanced a view called introspective justification, which is the 'justification that one has to believe that one is in a certain mental state, which one has just by virtue of being in that mental state' (2012, p. 261). Smithies thinks that his account has implications for FPA, since introspective justification is only possible from the first-person perspective. He claims 'it is a way of knowing about one's mental states that is always available in one's own case' (2012, p. 260). So, if I can justify the belief that I am in pain by introspecting that I am in pain, then I possess a method to justify beliefs about certain mental states that others do not possess.

Introspection, and in turn FPA, is not applicable to all of one's self-attributions, so it remains an open question as to what kinds of self-knowledge conform to FPA. I clearly do not have FPA with respect to my date of birth, my height, or weight. I know these facts in the same way that everyone else does—namely, by observing documents and using measurement tools. Moreover, it is unclear whether I have FPA with respect to my character traits, or moods, since it is unclear whether I can introspect them. The question I address in what follows is to what extent emotions can conform to the FPA framework I have outlined above, and whether existing theories can adequately account for it. If our self-ascriptions of our emotions can be grounded, or justified, by a broadly introspective process—whatever its nature may be—then it is my view that we can have FPA with respect to them.

In what follows, I will criticize three theories, which offer introspective ways of knowing our emotions—and thus potential ways of explaining how we have FPA of

emotions. These include the transparency method, neo-expressivism, and the theory of constructed emotion. I will argue that all three views cannot explain introspective access to emotions and in turn FPA, even if they can explain FPA of other mental states. Since these are not the only views in the self-knowledge literature, I will first give a brief explanation of why I have chosen to focus on them. Consider, alternatively, the acquaintance view, which Gertler (2012) defends. On this view, 'Some introspective knowledge consists in judgments that are directly tied to their truth-maker' (Gertler 2012, p. 99). An acquaintance theorist, such as Gertler, would say that introspective beliefs, such as one's occurrent belief that one is in pain, are connected directly to the facts that make the beliefs true—in this case, the conscious experience of being in pain.

This view is controversial, however, and often taken to have a limited scope by its proponents. Gertler, for example, claims that some acquaintance theorists may 'limit knowledge by acquaintance to sensations' (2011, p.124). If emotions are not simply sensations, as I suggested above, this raises a question about whether the view can explain how we know our emotions. Given the controversy about the scope of the view, I will not consider it in detail here. Acquaintance theorists who do not believe that their view can be extended to emotions may still be interested in what I have to say in Section 6, however. This is because I do not defend any specific 'introspective' view from the literature about how we do come to know our sensations here. Acquaintance theorists looking to account for how we do know our emotions may wish to say that we know our sensations by acquaintance, but must self-interpret those sensations to know our emotions.

Another view I do not consider in detail here is the Inner Sense view. According to this view, we each possess an internal scanner that can provide us with introspective access to our mental states (see Lycan 1996; Goldman 2006). If this theory can be applied to our emotions, then it can potentially explain FPA. There is, however, also controversy about the scope of the view (see Carruthers 2011)—that is, whether it applies to just sensations and propositional attitudes or emotions too. In Section 6, I provide some reasons for doubting that the inner sense view can be extended to emotions. Again, since inner sense theorists may differ about the scope of the view, they may differ in how they view Section 6. If they maintain that we come to know our sensations and propositional attitudes by an inner sense, but not our emotions, then their view will be compatible with what I say there.

3 Emotions and the Transparency Method

One prominent approach for explaining FPA has been given by proponents of the transparency method (hereafter, 'TM'). Contemporary TM theorists (see, e.g., Moran 2001; Byrne 2005, 2018; Fernández 2013; Andreotta 2021a) account for the self-knowledge of our mental states by drawing inspiration from a famous passage by Gareth Evans. In response to someone asking him the question 'Do you think there is going to be a third world war?', Evans claims, 'I must attend, in answering him, to precisely the same outward phenomena as I would attend to if I were answering the question 'Will there be a third world war?' (1982, p.225).

According to Evans (and other transparency theorists) we come to know what we currently believe by directing our focus onto the intentional object (the outward phenomena) of the belief in question. So, if I want to know whether I *believe* that Guyana is in South America I should not look inside, via a faculty of inner sense, for the presence of a pre-existing belief. Instead, I should focus on the content of the belief. In this case, I should judge whether I think that Guyana is in South America. If I judge that it is, then I should self-attribute such a belief. The question about what I believe is *transparent* to the question of what I judge to be the case. FPA can be explained on this view since only I am able to determine what I believe by making a judgement. If other people want to know what I believe, they need to observe my behaviour, or listen to what I say.

While some argue that the view is limited, and only applicable to belief (see, e.g., Nichols and Stich 2003; Finkelstein 2003), other philosophers (see, e.g., Fernández 2013; Byrne 2018) have provided accounts of how TM could be extended beyond belief. In his book *Transparency and Self-Knowledge*, Byrne (2018), for example, offers a uniformed account of TM. He claims that TM can explain how we know, and have FPA of, our sensations, beliefs, desires, intentions, memories, thoughts and importantly, for our purposes, even our emotions.

Byrne does so by appealing to the concept of an epistemic rule, which is a procedure that if followed will yield knowledge. For example, in the case of belief, he offers the following rule: 'BEL If *p*, believe that you believe that *p*' (2018, p.102). The reasoning present in this rule is explained by Byrne via a 'doxastic schema' (2018, p.100), which can be illustrated by the following example:

Evans Wrote the *Varieties of Reference*

I believe that Evans Wrote the *Varieties of Reference*

Concluding that you believe something on the basis of a worldly premise (in this case, that 'Evans wrote the *Variety of Reference*') may seem like a weak inference to make since the schema is clearly invalid, and inductively weak (cf. Byrne 2018, p. 103). Nevertheless, Byrne claims that if one attempts to follow the schema, one will end up with self-knowledge. He says, 'if [BEL] is followed, then the resulting second-order belief is true' (2018, p. 104). In other words, if one forms the belief (second-order) that one has a belief about Evans being the author of the *Varieties of Reference*, via this procedure, then it will be true. The reason for this is that the higher-order belief is formed on the basis of what one judges to be the case.

The rule is supported by considering Moore's Paradox. If I judge that Evans wrote the *Varieties of Reference*, but do not actually believe that he did, then I would be in the same position as someone who judges that it is raining yet does not believe that it is: a seemingly absurd avowal. So, it seems that judging that *P* is a good guide to what one believes about *P*. Epistemological asymmetry is preserved here because only from the first-person perspective can one know what one believes in this way; and

epistemological security is preserved because forming higher-order beliefs on the basis of what you judge would seem to be a secure way of determining what you believe.

Byrne constructs several other epistemic rules which he argues can be applied to different categories of mental states and importantly for our purposes, emotions. Since the topic of emotions is large, he focuses on just one emotion: disgust. The following epistemic rule is offered to explain how we know this emotion.

> DIS If x is disgusting, and produces disgust reactions in you, believe you feel disgust at x (2018, p.178).

Imagine I come across a loaf of mouldy bread in the pantry and want to know whether I find it disgusting. According to DIS, if the mouldy bread is disgusting, and the mouldy bread produces disgust reactions in me, then I should believe I find it disgusting. Byrne adds: 'DIS is strongly practically self-verifying…it can generate unsupported self-knowledge by an inference from a worldly counterpart premise' (2018, p. 178).

DIS initially seems like a promising approach to self-knowledge, and FPA, of disgust. For one reason, it appears to produce Moore's Paradox sentences, just like in the case of belief. If I accept that the bread is disgusting, and that it produces disgusting reactions in me, but I do not believe the bread is disgusting, then I would seem to be saying something 'absurd', just like the person who judges that is it raining yet fails to believe it is. This result would seem to suggest that DIS is a successful rule and that FPA of emotions in general could be explained by adopting similar rules for each different category of emotion (e.g., fear, anxiety, and so on). And like belief, epistemological asymmetry and security both seem preserved: others cannot know my disgust in this way.

Despite this result, I think there are problems for Byrne's approach that prevents TM from adequately explaining how we authoritatively know our own emotions. Firstly, let us focus on what is means to affirm that x is disgusting, which Byrne includes in the antecedent of DIS. What would it mean, for example, to affirm that a loaf of mouldy bread is disgusting? Byrne gives further clarification by suggesting that it is when x is disgusting in its '*presentation* or *appearance*' (2018 p. 179 emphasis in original). So, if the mouldy bread appears disgusting to me, and I am having reactions of disgust, then the antecedent of DIS can be affirmed. Yet this seems to raise a problem: x appearing disgusting to me just seems to be another way of saying that I am disgusted by x. This is a problem because if you need to determine that 'x appears disgusting' in order to determine whether you actually are disgusted by x, then why would you need to use the rule in the first place?

To see why this is a problem for the view, it will be useful to see how DIS is slightly disanalogous to BEL which, recall, involves a connection between judgement and belief. The reason BEL succeeds, in my view, is because judgement and belief are distinct mental events. This is why Moore's paradox arises. It is not illogical to judge that *P* and believe not-*P*, since belief and judgement are distinct, even though it might be absurd to do so. Yet on Byrne's approach, it is hard to see how independency can be preserved: affirming that 'x is disgusting' and feeling that you are 'disgusted by x' seem to be the same thing. It is thus not surprising, then, that a paradox, or contradiction, arises. It is, further, difficult to imagine situations where 'x appearing disgusting' and 'being disgusted by x' come apart. Byrne may reply that they can come apart since 'x appearing disgusting' is a fact about the world, whereas 'being disgusted by x' is a mental state. However, it is hard to see how this could be the case. Disgust refers to what one is disgusted by. There is, after

all, nothing problematic about one person being disgusted by broccoli and another person enjoying it. Such a disagreement does not mean that one party is making a mistake about the nature of vegetables.

Might DIS still be a good rule because it generates true beliefs? I would say no. If this were the only criterion by which to judge the success of transparent rules, then a variety of rules could be created, which would not seem to capture the general idea of TM. Consider the epistemic rule NERV, one might follow in order to know whether one is nervous.

> NERV If x is nerve-racking, and produces nervous reactions in you, believe you feel nervous at x.[8]

Following NERV might lead to true beliefs, as with DIS. Consider I am about to step onto the stage before 200 spectators. From my perspective the situation appears nerve-racking, and I certainly have nervous reactions: my heart is racing, and I may be shaking. Following NERV could lead to the knowledge that I am nervous. However, as with DIS it should be asked why such a rule is needed in the first place. If I know that the prospect of stepping onto the stage is nerve-racking, then I would seem to already know that I am nervous.[9]

There is one component of Byrne's approach that I think is worth developing, however—namely, his focus on attending to one's emotional reactions. As I will go onto describe in Section 6, the self-interpretation of such reactions can play a role in the self-attribution of emotions. As I argue there, however, this can be accounted for without invoking TM.

4 Expressing Emotions

Given the limitations of TM, it is worth considering other approaches to the self-knowledge, and FPA, of emotion. In this section, I consider neo-expressionism, a view defended by Bar-On (2004, 2015).[10] In contrast to Byrne (2018), Bar-On's view is a *non-epistemic* approach, since it does not construe self-knowledge in terms of inward detection, inference, or self-interpretation. FPA, according to a neo-expressivist is not achieved by some '*especially secure epistemic route,* or by deploying an especially secure epistemic *method or procedure'* (Bar-On 2004, p.112). FPA, rather, arises according to Bar-On, 'due to the fact that avowing subjects are uniquely capable of ascribing various current mental conditions to themselves in the course of speaking from those conditions' (2004, p.341). For example, only I can express and, thus show, my anger by expressing it—e.g., by clenching my

[8] This is not a rule that Byrne himself advances. I have constructed it by following the logic of DIS.

[9] While I am pessimistic that TM can be applied to our emotions, my argument against Byrne is limited to his account. It remains possible that a competing account of TM could be given. I have argued elsewhere (Andreotta 2021a), for example, that even though Byrne's account does not work for desire, another approach can.

[10] Finkelstein (2003) and Parrott (2015) also defend views which focus on self-expression. I limit my comments on the view to Bar-On's account in what follows, however.

fists or by raising my voice. It is through such expressive acts that Bar-On claims 'we can speak our minds' (2015, p.144). In order for another person to know that I am fearful or anxious, in contrast, they need to observe my behaviour or listen to what I say. These are classically epistemic routes to knowledge.

Neo-expressivism may initially seem well suited to explain the FPA we have of our emotions since they are commonly associated with various behavioural manifestations, such as body movements, and facial gestures. When I express my disgust at a mouldy loaf of bread, for example, I might say 'Yuck!' and grimace. Doing so not only suggests that I know I am disgusted by the loaf, but it also suggests that I have acquired knowledge of my disgust in a way that another person cannot, thus epistemological asymmetry is preserved. Epistemological security seems to be preserved too, since expressing a mental state this way seems to suggest I know I am in that mental state. Bar-On, of course, recognises that 'expressive failures' can occur from time to time—that is, we can express mental states we are not in. However, she does not think that these errors occur due to epistemic mistakes—that is, by mistaking one mental state for other, or misinterpreting evidence. She thinks that self-deception, wishful thinking, and biases can make us express states that we are not in (Bar-On 2015, p. 145). In what follows, I will focus on this element of her account. I will argue that epistemic errors can occur with respect to emotional self-attributions, and thus question the applicability of neo-expressivism to our emotions.

Let us first consider an example that Bar-On discusses involving a false sensory mental state expression. First, imagine that a person is sitting on the dentist chair waiting to have a tooth pulled. They see the dentist approach, but before any contact is made, they exclaim: "This hurts!" (Bar-On 2004, p. 8). Now, since the doctor has not touched the tooth, such an avowal would seem to be false. Of such a person, Bar-On claims that, 'though she has successfully expressed *pain*, she has not succeeded in expressing *her* pain. She could not have expressed her pain, since there was no pain for her *to* express' (2004, p. 323). While seemingly innocuous, such mismatches are important for Bar-On to address because 'expressing' seems like a success verb (2004, p.325). If self-expressions can sometimes be false, then it is important to ask how such errors arise. Bar-On insists that in such a case, one person does not mistake a pain sensation for another, since that would count as an epistemic error. Bar-On might be right that such a person does not mistake one sensation for another, e.g., mistake a numb feeling for a pain sensation. The error has occurred, mostly likely, because the person has falsely anticipated that they are about to feel pain. However, such an explanation is not so easily applied when it comes to our emotions, as I now argue.

Bar-On (2015) also considers self-interpretational accounts of self-knowledge, such as Carruthers' (2011), who argues that we lack direct access to our beliefs, intentions, and emotions. On Carruthers' view, in order to know what one believes, for example, one needs to self-interpret one's sensations. He claims that knowledge of one's mental states is 'interpretive, relying on sensory, situational, and behavioral cues' (2011, p. 325). Carruthers supports this view by focusing on a series of psychological experiments which purportedly show that subjects do make epistemic mistakes by misinterpreting their sensory, situational, and behavioural cues (referred to by Carruthers as the 'confabulation data'). In response, Bar-On (2015) considers

one specific example involving subjects being induced to write an essay arguing for a conclusion that they do not really hold. Carruthers interprets this data as suggesting that the subjects make such mistakes because they lack any kind of special access to their beliefs. Their errors reveal, in his view, the normal way that we achieve self-knowledge—namely through self-interpretation. In Carruthers' view, the subjects would be making an unconscious inference like the following, "I can see that I have argued for this conclusion, so I must believe it." Bar-On rejects this suggestion, however. She grants for the sake of argument that the subjects do make a false self-attribution of belief but claims that the 'falsity is not due to an epistemic mistake on the subjects' part (namely, failing correctly to recognize or reflectively attend to a state they are in). Instead, the falsity represents an expressive failure, where the failure has identifiable psychological causes' (2015, p. 146). Bar-On is required to give this explanation because if Carruthers is right that our beliefs are known by a process of self-interpretation, then this raises the possibility that our beliefs, and other propositional attitudes, all might be known this way. This would conflict with Bar-On's claim that epistemic errors do not occur (in *first-personal* self-ascriptions, at least), and as a result neo-expressivism. Bar-claims that such as idea is 'absurd' (2015, p. 146–147), and suggests that expressivist failures arise occasionally due to psychological shortcomings.

In response to Bar-On, I wish to partially challenge her claim of absurdity. While I agree with her that Carruthers' position is implausible when applied to beliefs and intentions (and other propositional attitudes), the idea of self-interpretation is much more plausible if we limit it, and apply it partially, to our emotions. I agree with Bar-On that the 'possibility of falsity does not imply that avowals must be vulnerable to *epistemic* errors' (2004, p.221). However, I think we do sometimes see epistemic errors occurring with respect to the self-attribution of emotions, which I take to be a challenge to Bar-On's view. How can it be determined what kinds of errors occur when we misattribute (or confabulate) an emotion? If we could find a case where a person misinterprets their sensations when self-attributing an emotion, then we would seem to have a case where a subject makes a partial epistemic error. An error of this type might show how we learn about our emotions generally. We could then run an *epistemic argument against neo-expressivism*:

> [1] If the neo-expressivist view is true, about our FPA of emotions, then emotional
> self-attribution errors are not due to epistemic mistakes on the subjects' part.
> [2] Emotional self-attribution errors are partially due to epistemic mistakes on the subjects' part.
> [C] The neo-expressivist view is not true, with respect to our FPA of emotions.

Even though [2] is controversial, there are findings from the philosophical and psychological literature that require us to take it seriously.

Consider, first, the philosophical work of Eric Schwitzgebel, who is sceptical of our ability to accurately self-attribute our emotions. He notes that 'We are remarkably poor stewards of our emotional experience. We may *say* we're happy—overwhelmingly we do—but we have little idea what we're talking about' (2008, p. 250 emphasis in original). Schwitzgebel supports this position with reference to his own

experience. For example, he considers a case where his wife tells him that he is angry about doing the dishes, even though he does not, upon serious reflection, see it himself. Schwitzgebel concedes that his wife is right, however, since his wife can 'read his face' better than he can 'introspect' (2008, p. 252). Schwitzgebel does not paint a definitive picture of how his introspective abilities have failed him, but he does describe some of his introspective shortcomings. Regarding his ongoing emotional experience, he claims that he is often unsure what emotion he is currently undergoing (2008, p. 251). Moreover, he also claims it is difficult to answer basic questions about the nature of his emotions. According to his own experience—which Schwitzgebel takes as typical—introspection fails to tell him whether emotional states, like anger or fear, are always felt phenomenally (i.e., as part of one's stream of consciousness) (2008, p. 149); or whether emotions, like joy or fear, are realized in the same way each time (e.g., as a feeling in the head).

The picture that Schwitzgebel paints, then, is one of pessimism with respect to our ability to accurately know our emotions by introspection. However, his reflections also raise important questions about the *way* in which we know our emotions. The introspective limitations that Schwitzgebel describes may not simply show that we possess an unreliable faculty of introspection, broadly speaking. What they may suggest is that self-knowledge of our emotions requires, at least partially, self-interpretation. It may be that we often make mistakes in our emotional self-attributions, the kind Schwitzgebel describes, because we misinterpret our phenomenal experiences (or sensations), which themselves may be accurately introspected. So, rather than take a wholly sceptical conclusion, as Schwitzgebel does, it may be that we still have an epistemic advantage when it comes to our emotional self-attributions—albeit one that is limited to the phenomenal part of our emotional experiences. Schwitzgebel, then, appears to take a glass half empty view, whereas I will take a glass half full view.[11]

Is there more direct evidence, though, to show that the mistakes we make in our emotional self-attributions occur because of *epistemic* errors of the kind I have in mind? Consider, next, several empirical examples which may show that false emotions self-attributions have occurred as a result of mistaken self-interpretations of sensations. These examples by no means settle the question of whether [2] is true, but they do raise important questions about the way in which we acquire knowledge of our emotions.

In a famous experiment from the 1960s, Schachter and Singer (1962) injected subjects with adrenaline and noticed that they would report the associated feelings of the hormone differently depending on the context they were in. Subjects who were around happy people reported the feelings as euphoria; while subjects who were surrounded by angry subjects reported it as anger (Barrett 2017a, p. 34). The experiment is now considered controversial (Barrett 2017c), due to replications issues, but it does raise the interesting possibility that emotional self-attributions could be based on sensory self-interpretations (Reisenzein 2017). If

[11] Here I thank an anonymous peer reviewer for pointing out this minor, though key, difference in perspectives between Schwitzgebel's view and my own.

a subject mistakenly thought they were angry because they misinterpreted their introspection sensations, then that looks like a partial epistemic error.

Another well-known experiment, from the 1970s, which also looked at the connection between sensations, context, and emotions is the famous "love on a bridge" experiment conducted by Dutton and Aron (1974). In these experiments, male subjects were met by a female confederate after crossing a sturdy bridge in one condition; and also, in a condition where they crossed a rickety bridge. The experimenters found the male subjects who crossed the rickety bridge were more likely to ask for the female confederate's phone number, compared to participants who crossed the sturdy bridge. One explanation for these results is that the subjects who crossed the rickety bridge may have interpreted their sensations—e.g., a rapid heartbeat, or shortness of breath—as meaning that they were undergoing a feeling of emotional attraction. It might be that they misinterpreted the sensations they were currently undergoing and formed a false belief about what emotion they were experiencing. Such an explanation is controversial however, as it is possible that crossing the bridge simply caused a change in their emotions, and they were simply expressing (correctly) their mental states (cf. Carruthers 2011, p.142, ft. 13).

A final example is from the psychologist Lisa Feldman Barrett, who gives a personal account of a date she went on:

> As we sat together in a coffee shop, to my surprise, I felt my face flush several times as we spoke. My stomach fluttered and I started having trouble concentrating. Okay, I realized, I was wrong. I am clearly attracted to him. We parted an hour later — after I agreed to go out with him again — and I headed home, intrigued. I walked into my apartment, dropped my keys on the floor, threw up, and spent the next seven days in bed with the flu (2017a, p. 30).

What kind of error does Barrett make here, if one at all? One suggestion, that I consider in more depth below, is that she has misinterpreted her introspected sensations, in the context of a date, and expressed an emotion of attraction that she did not have. If this explanation is correct, then it would seem like a partial epistemic error has occurred, since she has misinterpreted her bodily sensations—namely, the feeling in her stomach. Bar-On could reply that an expressivist failure occurred due to 'self-deception or wishful thinking' (2015, p. 145). This is certainly compatible with the case, but I think it lacks explanatory power here since the mistake Barrett makes seems to have a certain pattern that is suggestive of misinterpretation.

None of these empirical considerations give us conclusive support for [2]. I grant that there are important interpretative questions that still need to be settled. The purpose of raising these examples has been to challenge Bar-On's claim that it would be 'absurd' to think that self-interpretation can always be involved, at least when applied to our emotions. The idea is more plausible if we limit our application of self-interpretation to our emotions and say that our emotional expressions are only partially epistemic. I now develop this idea further over the next two sections.

5 The Theory of Constructed Emotion

In her recent book, *How Emotions are Made,* Barrett (2017a) defends a view called the *theory of constructed emotion*. This view is opposed to the classic view of emotions which maintains that 'Each emotion faculty is assumed to have its own innate 'essence' that distinguishes it from all other emotions' (Barrett 2017b, p.2). Against this view, Barrett claims that there are no fingerprints of emotions, meaning that fear, anger or happiness can be experienced in a variety of ways, and can involve different bodily functions (2017a, pp.13–14). Various cultures, further, may even express a certain emotion category, like anger, in ways that look quite different from the way other cultures express the emotion (cf. Gendron et al. 2014). The second main idea of the book, and the one I focus on here, is that we construct our emotions from our bodily sensations.

This theory, which is supported by recent neuroscientific evidence, is important to consider because it is relevant to our current discussion about FPA.[12] If Barrett is right that we construct our emotions from our bodily sensations, then this might inform us about the nature of FPA: if we need to draw upon our sensations in order to know what emotions we are undergoing, that might suggest that we do sometimes make epistemic errors, since we might misinterpret the sensations we are experiencing and self-attribute, or construct, an emotion we do not have. This would seem to support the view I am proposing which says that the FPA we have with respect to our emotions is only partial.

What does Barrett mean, then, in claiming that we 'construct' our emotions? We can clarify this notion by considering an example. Suppose one is in a doctor's office and feels an ache in one's stomach. According to the theory of constructed emotion, the ache might be experienced as anxiety if one is, for example, waiting for important test results. The theory would also suggest that if the same ache arose in a different context, it could be experienced quite differently. For example, the ache could be experienced as hunger, if one was waiting for one's meal to arrive at a busy restaurant. The same ache, further, could even be experienced by a judge in a courtroom as signifying a feeling of guilt with respect to a defendant (Barrett 2017a, p. 29–30). Barrett claims that an 'emotion is your brain's *creation* of what your bodily sensations mean, in relation to what is going on around you in the world' (2017a, p.30). The same sensation could give rise to a number of different emotions like anxiety, fear, or disgust, given the right context.

This theory would initially seem to support my contention that our emotional self-attributions are liable to epistemic errors if what Barrett means by the term 'construction' is that we self-interpret our sensations in order to know our emotions. If for example, after eating something that causes an ache in my stomach, I believe that I am disgusted at a person who I am not disgusted at. Such an error, further, looks partially like an epistemic one, since I would have constructed the wrong

[12] See Barrett (2017b, p.14, table 2) for a list of the various empirical studies which support this theory.

emotion from my sensations. This does not seem the best way to characterise Barrett's view however, as she says:

> Instances of emotion have no objective fingerprints in the face, body, and brain, so "accuracy" has no scientific meaning. It has a social meaning — we certainly can ask whether two people agree in their perceptions of emotion, or whether a perception is consistent with some norm. But perceptions exist within the perceiver (2017a, p.40).

This position would seem to have implications for the FPA of emotions, since our constructions would seem to settle the matter of what emotional state we are in, thus making it hard to account for error. As Barrett's elaborates: 'We do not recognise, discover, or identify [our emotions]: they are made by us' (2017a, p.40). She claims that emotions are not 'waiting to be revealed' (2017a, p. 40) and that we do not '*recognize* emotions or *identify* emotions' (2017a, p.40).

In response to Barrett, I would accept that while we may sometimes construct our emotions, it is problematic to say that our emotions are *merely* constructions, and that we do not discover them. For one reason, doing so ignores the dispositional element of our emotional lives. As Peter Goldie suggests, if you are a jealous person, then 'you are, at least to some extent, a person who is disposed to be jealous' (Goldie 2000, pp. 12–14). An emotion like jealously has a phenomenological aspect to it, but it also has a behavioural component that can inform others that a person is anxious. What is noteworthy about such dispositions is that they seem to be able to appear without any conscious feelings. Goldie for instances argues (2000, p. 63) that one can have an emotion, such as fear, without being conscious of any thoughts and feelings. This idea seems to be supported by empirical findings, as Jäger (2009, p.135) notes in his discussion of experiments conducted by Mendolia (1999). These experiments show that subjects can believe that they are anxious in cases where they do not have such an emotion; and also, that they can be undergoing anxiety without any awareness of it.

For these reasons, we need to allow that someone can be in an emotional state, not realise it, and then come to know this from another's testimony. Imagine that John learns about his coworker Max's recent promotion and behaves in a way that suggests he is jealous of it. The other coworkers notice John acting uncivilly towards Max, and that his demeaner has changed towards Max since the promotion. If the emotion of jealousy is not 'constructed' by John in this case, does this mean that John does not have the emotion? Further, if a contrary emotion is constructed by John, what to say of the jealousy? Perhaps Barrett would say that the jealously has not been construed by Max until he has some sort of awareness of it, but this would be to ignore his dispositions. If Max really was anxious all along, however, then discovery may indeed be applicable, say when John's coworkers point out his behavior to him and he gains knowledge of his emotion.

A second reason to challenge Barrett's focus on construction, that is related to the first, has to do with correctness conditions. While Barrett claims that emotions have no objective fingerprints in the face or brain, she does claim that 'emotion concepts have social reality' (2017a, p.133). This suggests that even if people express the same emotion in different ways, we would expect there to be some shared concepts,

at least in the same society. Consider, for example, a case that is mentioned by Prinz, who describes a person who claims to be in love with someone but has never shown any signs of it. Of this case, Prinz suggests, 'I think we would regard this person as disingenuous or confused' (2004b, p.50). It is possible, of course, that the person's conception of love is different to other peoples', but if they fail to show any signs at all, it is more likely that such a person is not in love. Once someone brings this evidence to this person's attention, we may even see them change their belief about being in love. This would only be possible if the two parties shared the same concept of love. The central issue with Barrett's view is that it has difficulty accounting for emotions that subjects have but lack awareness of. This is important to account for, as Prinz also points out, because 'we do not say that these emotions disappear when they are unfelt, because the disposition is there all the time' (2004b, p. 50).

The central focus on construction makes it hard to account for 'confabulation' and 'self-deception' too.[13] For if I construct a feeling of anxiety from sensations in a certain context, and that settles the matter, then it is hard to see how I could be wrong about this. If I could be wrong, on the other hand, then I should be able to discover I was wrong. Focusing on self-interpretation allows us to say that we do sometimes make mistakes, since we can misinterpret our sensations in different contexts. None of this requires us to deny that we ever construct our emotions. And I think Barrett is right to focus on our sensations as a guide to knowing our emotions. But I would place the focus here on self-interpretation, rather than construction. I now develop this idea further.

6 Partial First-Person Authority of Emotions

Following Barrett, I contend that our sensations and beliefs play an important role in our emotional self-attributions. In contrast to Barrett, however, I propose that we focus on discovery, in addition to construction, while also recognising the role that *self-interpretation* plays in our emotional self-attributions. This view can be expressed as follows:

> *Partial First-Person Authority (FPA) of Emotions*: Our first-person avowals about what emotions we are currently undergoing are *grounded* in our introspected sensations and beliefs. In order to gain awareness of our emotions we need to self-interpret those (introspected) sensations and beliefs, as well as other propositional attitudes. The FPA we have of our emotions is partially asymmetrical, but not wholly. Self-interpretation is involved *partially* in the process. We have FPA with respect to our sensations and propositional attitudes, but knowledge of emotions requires us to self-interpret those mental states.

According to this view, there is a partial epistemological asymmetry associated with our emotion self-attributions. We have 'direct access' to our sensations and

[13] It is interesting to note that the term 'confabulation' does not appear in Barrett's (2017a) book.

beliefs, since we do not need to draw upon any behavioural evidence to know them, like other people need to do (cf. Bortolotti 2009, p. 210).[14] Evidence drawing is not wholly absent however, as we still need to self-interpret our sensations and beliefs to know our emotions. I call this partial FPA because introspective processes *alone* will not give you knowledge of your emotions: self-interpretation is also required.

I have not, to be sure, accounted for how we do authoritatively know our sensations and beliefs in this paper, and so remain neutral here about how best to do that. Even though I have been critical of TM and neo-expressionism, as ways of explaining FPA of emotions, it remains possible that either of these views can explain how we can know our sensations or beliefs in a non-interpretative way. The acquaintance view or inner sense view, further, may even be able to explain how we have FPA of our sensations or propositional attitudes (e.g., our beliefs and intentions). FPA, then, need not be explained uniformly, and it need not be an all or nothing phenomenon: it can be partial. An acquaintance theorist who thinks that we can come to know our sensations by acquaintance could accept that we know our emotions by self-interpreting those sensations.

As I mentioned in the previous section, focusing on self-interpretation, in addition to construction, allows us to say that we sometimes do *discover* the emotions we are undergoing; and it also suggests the possibility of certain types of errors. Where there is interpretation, there is also the possibility of *misinterpretation*. It is important to account for this phenomenon since emotions have a dispositional component. Focusing predominantly on construction, as Barrett does, makes it difficult to account for cases where a person comes to know that they are undergoing an emotion on the basis of another person's testimonial evidence. If I come to know that I am jealous for example, because someone informs me that I am acting jealously, then the emotion has already existed for some time, as it was my jealous behaviour that caused the person to believe that I am jealous. Such a causal sequence would seem to suggest that the emotion was present before I became aware of it, so discovery seems appropriate here.

What evidence or justification is there for the partial FPA of emotions view? First, I think the view makes good sense of some of the psychological cases we looked at earlier, and also the cases that Barrett mentions. It is important, however, that I clarify how I think this evidence supports the partial FPA view of emotions, as it may seem like I am simply advancing the following two uncontroversial claims:

(1) Knowledge of our emotions is partially based on self-interpretation; and
(2) That self-interpretation is prone to error, i.e., to misinterpretation.

Suppose an inner sense theorist, who thinks that we can introspect our emotions via an internal scanner, considers these two theses. They may state that while we sometimes use self-interpretation to know our emotions, and that we sometimes make misinterpretations, this is not enough to undermine the thesis that we can

[14] Even though most philosophers would accept this, not all do (see, e.g., Carruthers (2011).

know our emotions in a purely introspective way (i.e., via an internal scanner). Such a theorist could, following Carruthers' (2011) terminology, be a 'dual-method' theorist who holds that we can know our emotions via self-interpretation, *and* we can also introspect them.

In my view, the kinds of data that we looked at in Sections 4 and 5, put pressure on such a view. An important question can be raised here: why think that we possess an introspective method in addition to an inferential method for discerning knowledge of our emotions, when a (partially) interpretative method alone (hereafter, the 'single-method') will suffice? If we can know our emotions by self-interpretation, then, it might be that positing the existence of a dual-method would be superfluous, given that a single-method appears to be sufficient.[15]

The partial FPA of emotions view should be seen as a challenge to the dual method position with respect to emotions. If correct, then introspection cannot wholly give us knowledge of our emotions—we always need to partially rely on self-interpretation. If I am right that self-interpretation is always required, then the neo-expressivist view, the transparency method, and the inner sense view, when applied to emotions will be undermined. These views stand in contrast to the claim that we must always partially interpret our sensations to know what emotion we have.

One thing that is notable, with respect to the FPA view of emotions, is that emotional self-attributions appear to be correlated with subjective feelings. Someone who is nervous may also feel their stomach flutter. Further, someone who is presented with sensations, such as a feeling in their stomach, could interpret the feeling as romantic love, even though they are not really undergoing the emotion. This looks like a paradigmatic self-interpretation error, which may be illustrative of the way we learn our emotions. The cases of error are not enlightening because I think introspective views are committed to a kind of infallibility thesis. The cases of error are interesting because I think they help to reveal the normal way we come to know our emotions. When one is on a date, and one becomes aware of a sensation in one's stomach it could be quite natural for one to interpret that feeling as romantic attraction to their date, even if one was not, as a matter of fact, romantically attracted to their date, but say sick. If self-interpretation is the normal way that we come to know our emotions, we would expect to make such mistakes when the context and other cues are misleading.

To say that there is a connection between sensations and emotions is not a new observation, as shown in the works of poets, as Damasio (2018, p. 106) points out, and sometimes portrayed in fiction. For example, in a famous episode of *the Simpsons*, 'Sweet Seymour Skinner's Baadasssss Song' (Oakley et al. 1994), Bart Simpson is sitting at the breakfast table wondering why he is not jumping for joy after finding out that his school principal, and rival, Seymour Skinner was fired because of an incident he was the cause of. He claims that all he has is a: "weird hot feeling in the back of [his] head." His sister Lisa, who is also at the table says,

[15] This argument is clearly inspired by Carruthers' (2011). In contrast to Carruthers, however, I only think there is pattern of error to explain when it comes to our emotions, not also our propositional attitude misattributions. See Andreotta (2021b) for a more detailed discussion of this point.

"That's guilt, you feel guilty because your stunt wound up costing a man his job." Bart replies, "Yeah I guess it is guilt." The viewer momentary accepts that Bart feels guilt, until the very next moment, where we see a small spider on the back of Bart's neck, that is biting him. This raises a question for the viewer about whether Bart really does feel guilt. But it also raises the question about the route Bart has followed to acquire knowledge of his guilt. It may be that this fictional account mirrors a real psychological process where humans self-interpret their sensations to know their emotions.

Why think that such an account is true of human psychology? In two recent studies by Nummenmaa et al. (2014, 2018) it was shown that people locate different areas of the body as feeling a certain way when they undergo basic emotions (e.g., anger, fear, disgust) and non-basic emotions (e.g., anxiety, love, depression, contempt). In the 2014 study, they asked Western Europeans (Finish and Swedish), and East Asian (Taiwanese) participants to point to the specific region on a map of the body where they experienced feelings when they were undergoing certain emotions. Participants used a computer based self-report method called 'emBODY', which presented them with blank silhouettes of a body. After viewing stories, movies or facial expressions, they were instructed to note the emotions they were undergoing, and also colour in the parts of the silhouettes that they felt represented the places in the body that those feelings were occurring. Participants who felt anger, for example, noted the presence of associated feelings coming from the head, stomach, and shoulder area. Happiness, on the other hand, was reported as having a feeling that was present in the whole body. And envy was associated with feelings that were in the head region and upper chest. In the follow up study, Nummenmaa et al. (2018), tested for an even greater number of emotions, and other states (100 in total). In addition to anger and fear, they looked at guilt, sadness, and disgust, amongst many others. As with their earlier study, it was found that various emotional experiences were paired with different feelings in the body, indicated by the subjects in similar regions on the bodily maps. Guilt tended to be felt higher up on the body, whereas nervousness around the middle of the body. The authors concluded that these bodily sensations 'could be at the core of the emotional experience' (2014, 650).

These results are relevant to our current discussion because they suggest that our awareness of our emotional experiences involve bodily sensations. This is just what one would expect if self-knowledge of one's emotions required one to self-interpret one's bodily sensations. And since the experimenters also found that certain areas of the body were active during various emotion experiences, this would also seem to suggest that we might sometimes mistake one emotion for another since certain emotions may have similar bodily feelings. I may for example see that I am at a doctor's office and interpret a certain feeling as anxiety; but if I see that I am in a crowded lunch bar, I may self-interpret the same feeling as hunger. The best explanation of all of this, in my view, is that self-interpretation of our sensations is our normal route to self-knowledge of emotions.

I am not the first to focus on self-interpretation with respect to the way we know our mental states. Carruthers (2011) and Cassam (2014) both argue that self-knowledge of most of our mental states is self-interpretative in nature. They both deny that

we have any introspective way of knowing our propositional attitudes, for example.[16] Carruthers, for instance, argues that in order to know whether I believe there is going to be a third world war, I would need to draw upon sensory evidence. He supports this claim by looking at the data from split-brain studies, choice blindness experiments, and priming studies. Problematically for Carruthers' view, his interpretation of these data is controversial (see Andreotta 2021b) and not widely accepted; and further, there is notable lack of sensory data present when we acquire knowledge of our beliefs, as well as other propositional attitudes. Nummenmaa et al. (2014, 2018), for instance, do not note of any sensations that are paired with our experiences of believing or intending to do something. Even though some of our propositional attitudes are associated with certain feelings, it is implausible to suggest that we always need to interpret our sensations to know these states. One critic of Carruthers' work, for example, Georges Rey, has pointed out, '[d]esire, wonder, doubt, pretence, curiosity, for example, don't seem to be linked to any specific sensations' (2013, p. 274). Self-interpretation as the only way we can know our propositional attitudes, thus, seems implausible. If we focus on partial self-interpretation and limit this approach to our knowledge of our emotions, on the other hand, then we are left with a much more plausible and empirically supportable position.

Let us take an example of Cassam's, involving emotions, to see why this is the case. In a discussion of his inferentialist position, Cassam examines a case of self-knowledge involving the emotion love from Marcel Proust's *Remembrance of Things Past*. The case involves the character Marcel's realisation that he loves Albertine, which comes about after her absence causes him to experience anguish (or suffering). Cassam suggests that Marcel can come to know that he loves Albertine by focusing on the suffering he is undergoing. Cassam suggests that the suffering is thus the *basis* of Marcel's self-knowledge but stresses that it does not itself constitute Marcel's love. This is because love might only be one of several possible explanations of Marcel's suffering. Cassam explains further:

> The inference is mediated by an interpretation of his suffering that is grounded in his understanding of the relationship between this kind of suffering and romantic love. His *route* to self-knowledge here is inference, whereas the *basis* of his self-knowledge is suffering (2014, p. 181 emphasis in original).

Cassam as an inferentialist would deny, then, that we can know our emotions, such as love, by an introspective, non-interpretative method.

One feature of this self-knowledge process that Cassam does not spend much time discussing, however, is the phenomenological set of sensations that make up the anguish or suffering that Marcel can draw upon. In terms of FPA, this is significant, because the suffering would involve felt experiences which can be introspected. Marcel can, further, attend to his introspected memories, beliefs, and intentions. Introspecting these mental states can give him an important epistemic advantage over other people trying to determine what emotion he is undergoing,

[16] Carruthers does allow that we can have non-interpretive access to our sensory-based attitudes, such as one's belief that one is seeing red.

even if self-interpretation is involved in part of the process. So to say, as Cassam does, that Marcel must rely on self-interpretation to know that he is in love is partially true; but not the whole story. Marcel can ground his self-attribution of love with his introspected experience of suffering, which is epistemically significant.

The FPA view of emotions, then, can be thought of as a middle ground position, between those who think we can have direct introspective access to our emotions on the one hand and those such as Cassam, who thinks that it is all self-interpretation, on the other. We do not need to abandon the idea of asymmetry completely, as Cassam seems to suggest, even if we do need to recognise that self-interpretation is involved in the self-attribution process. This may require us to abandon the kinds of epistemic confidence we feel we have when it comes to emotion self-attributions, compared to the confidence we have when we self-attribute a belief or a sensation; but again it is important to recognise that a kind of epistemic advantage is still there. Even in cases where we might misattribute emotions, we may get other mental states right. Consider Damasio, again, who says 'precise feelings that comparable situations evoke may well be tuned by cultures' (2018, p. 109). In one example, he suggests that the nervousness of students before an exam can be experienced by German students as butterflies in the stomach, whereas, in Chinese students it can be experienced as a headache. Both students may have FPA with respect to the sensations they are experiencing, but may interpret those sensations in very different ways, given the context, culture, background beliefs and so on. Some students may interpret those sensations 'correctly' and some may not. We need not say, along with Cassam, that it is 'inferential all the way' down (2014, p.161). In the case of emotions, our FPA is partial because some of the evidence we draw on to know them is available only to us; and it is stronger than the evidence we draw on regarding other people's emotions, which the Cassam-style view denies.

One may object to the partial FPA of emotions view by saying that self-interpretation occurs 'internally', and thus is an introspective process; so we can still have FPA about our emotions. In a recent paper on the introspection of emotions, De Vlieger and Giustina (2022) give an account of how we introspect our emotions. They advance a three-stage model, which begins with *primitive introspection*. This involves becoming aware of the non-classificatory 'information about the phenomenology of the introspected experience' (2022, p. 561). The second stage, *reflective introspection*, is relevant to our focus here. They claim, 'It consists in classifying the introspected experience as an instance of a known experience type' (2022, p. 563). This coheres with what I have suggested here: the partial FPA of emotions view suggests that we need to self-interpret our sensations to know them. De Vlieger and Giustina add that 'At this stage, the subject gives or attempts an *interpretation* of the introspected experience: they try to figure out what kind of experience' (2022, p. 563 emphasis added) it is.

Although I agree with De Vlieger and Giustina that we need to interpret our sensations (as well as other mental states such as memories and beliefs) to know our emotions, I disagree that this process is best thought of as 'reflective introspection'. Let us consider one of their examples to see why this is the case. Consider Caroline, who feels her heart pumping, notices a smile coming across her face, and experiences an urge to jump up into the air. According to De Vlieger and Giustina,

Caroline can know that she feels the emotion joy by self-interpreting these experiences—what they refer to as reflective introspection.[17] Now, while Caroline clearly has epistemically significant evidence to draw on to self-ascribe her emotion—namely, her felt sensations—she still needs to do interpretive work, just like other people who are attempting to attribute an emotion to her. So, I would not call this *interpretative* process 'introspective' simply because it occurs internally. Caroline's epistemic authority is thus only partial.

7 Conclusion

Some of the existing theories of self-knowledge and FPA do not map well to our emotions, even if they can explain the self-knowledge and FPA we can have of other mental states, such as our sensations, beliefs, and intentions. While the thesis that we know all of our beliefs or intentions by self-interpretation is implausible, there is still room to account for self-interpretation, in a partial way, when accounting for the knowledge we can have of our emotions. Self-knowledge, therefore, need not be explained uniformly as Byrne (2018) and Bar-On (2004) maintain.[18] The special epistemological advantage we have to our emotions is grounded in our ability to know our own sensations and beliefs in a way others cannot. But interpretation still has a partial role to play, however, just like when we acquire knowledge of others' emotions. We do not, I have argued, possess a wholly introspective method for coming to know our emotions.

I provided support for the partial FPA view of emotions by looking at the empirical literature. If people really do draw from their bodily sensations in order to know what emotions they are undergoing, we would expect there to be certain sensations associated with our emotional self-attributions. This is indeed what the empirical literature seems to suggest. I also accept, however, that more research is required to make the support for this view even stronger. Further research into the way in which we make errors in our self-attributions of emotions will hopefully not only inform us about the way we know our emotions, but also potentially inform us about why some people are more accurate than others in the self-reporting of their emotions.

Acknowledgements This paper was presented to audiences at the 2021 Australasian Association of Philosophy Conference, which was held online; and at Curtin University, Bentley, in June 2021. I thank everyone who took part in those enlightening discussions.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions

Data Availability No new data were created or analysed in this study. Data sharing is not applicable to this article.

[17] De Vlieger and Giustina are not specifically concerned with first-person authority.

[18] This conclusion about variety aligns with Coliva's (2016), who argues that self-knowledge of all the different categories of mental states arises from a variety of methods.

Declarations

Conflict of Interest None.

References

Alston, W. 1971. Varieties of Privileged Access. *American Philosophical Quarterly* 8 (3): 223–241.
Andreotta, A. J. 2021a. 'Extending the transparency method beyond belief: A Solution to the Generality Problem', *Acta Analytica* 36 (2): 191–212.https://doi.org/10.1007/s12136-020-00447-9
Andreotta, A. J. 2021b. 'Confabulation does not undermine introspection for propositional attitudes', *Synthese* 198 (5): 4851–4872.https://doi.org/10.1007/s11229-019-02373-9
Bar-On, D. 2004. *Speaking My Mind: Expression and Self-Knowledge*. Oxford: Clarendon Press.
Bar-On, D. 2015. Transparency, expression, and self-knowledge. *Philosophical Explorations* 18 (2): 134–152.
Barrett, L.F. 2017a. *How Emotions are Made: The Secret Life of The Brain*. New York, NY: Houghton-Mifflin-Harcourt.
Barrett, L.F. 2017b. The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience* 12 (1): 1–23. https://doi.org/10.1093/scan/nsw154.
Barrett, L. F. 2017c. 'Extended endnotes for How Emotions are Made.' Retrieved from: https://how-emotions-aremade.com/notes/Schachter_and_Singer_(1962). Accessed 14 Sept 2023.
Bilgrami, A. 2006. *Self-Knowledge and Resentment*. Cambridge, MA: Harvard University Press.
Bortolotti, L. 2009. *Delusions and Other Irrational Beliefs*. New York: Oxford University Press.
Byrne, A. 2005. Introspection. *Philosophical Topics* 33: 79–104.
Byrne, A. 2018. *Transparency and Self-knowledge*. Oxford: Oxford University Press.
Cassam, Q. 2014. *Self-knowledge for Humans*. Oxford: Oxford University Press.
Carruthers, P. 2011. *The Opacity of Mind: An Integrative Theory of Self-Knowledge*. Oxford: Oxford University Press.
Coliva, A. 2016. *The Varieties of Self-Knowledge*. London: Palgrave Macmillan.
Collier, M. 2011. Hume's Science of Emotions: Feeling Theory without Tears. *Hume Studies* 37: 3–18.
Damasio, A. 2018. *The Strange Order of Things: Life, Feeling, and the Making of Cultures*. New York: Pantheon.
De Vlieger, B., and A. Giustina. 2022. Introspection of Emotions. *Pacific Philosophical Quarterly* 103 (3): 551–580.
Dutton, D.G., and A.P. Aron. 1974. Some evidence for heightened sexual attraction under conditions of high anxiety. *Journal of Personality and Social Psychology* 30 (4): 510–517. https://doi.org/10.1037/h0037031.
Evans, G. 1982. In: *The Varieties of Reference*, ed. J. McDowell. Oxford: Oxford University Press.
Fernández, J. 2013. *Transparent Minds: A Study of Self-Knowledge*. Oxford: Oxford.
Finkelstein, D.H. 2003. *Expression and the Inner*. Cambridge, MA: Harvard University Press.
Gallois, A. 1996. *The World Without, the Mind Within: An Essay on First-Person Authority*. Cambridge: Cambridge University Press.
Gertler, B., ed. 2003. *Privileged Access: Philosophical Accounts of Self-Knowledge*. Aldershot: Ashgate Publishing.
Gertler, B. 2011. *Self-Knowledge*. London: Routledge.

Gertler, B. 2012. 'Renewed Acquaintance'. In: *Introspection and Consciousness*, eds. D. Stoljar. Oxford: Oxford University Press.

Gendron, M., D. Roberson, J.M. van der Vyver, and L.F. Barrett. 2014. Perceptions of Emotion from Facial Expressions Are Not Culturally Universal: Evidence from a Remote Culture. *Emotion* 14 (2): 251–262.

Goldie, P. 2000. *The Emotions: A philosophical perspective*. Oxford: Oxford University Press.

Goldman, A. 2006. *Simulating Minds*. Oxford: Oxford University Press.

Jäger, C. 2009. Affective Ignorance. *Erkenntnis* 71 (1): 123–139.

Lycan, W.G. 1996. *Consciousness and Experience*. Cambridge, MA: MIT Press.

Mendolia, M. 1999. Repressors' appraisals of emotional stimuli in threatening and nonthreatening positive emotional contexts. *Journal of Research in Personality.* 33: 1–26.

Moran, R. 2001. *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton: Princeton University Press.

Nichols, S., and S. Stich. 2003. *Mindreading: An Integrated Account of Pretence, Awareness, and Understanding Other Minds*. Oxford: Oxford University Press.

Nussbaum, M.C. 2001. *Upheavals of Thought: The Intelligence of Emotions*. Cambridge: Cambridge University Press.

Nummenmaa, L., E. Glerean, R. Hari, and J.K. Hietanen. 2014. Bodily maps of emotions. *Proceedings of the National Academy of Sciences of the United States of America* 111: 646–651. https://doi.org/10.1073/pnas.1321664111.

Nummenmaa, L., R. Hari, J.K. Hietanen, and E. Glerean. 2018. Maps of subjective feelings. *Proceedings of the National Academy of Sciences of the United States of America* 115: 9198–9203. https://doi.org/10.1073/pnas.1807390115.

Oakley, B. (Writer), Weinstein, J. (Writer) & Anderson, B. (Director). 1994. 'Sweet Seymour Skinner's Baadasssss Song' (Season 5, Episode 19) in D. Mirkin, J. L. Brooks, M. Groening, & S. Simon (Executive Producers), *The Simpsons*. Gracie Films, Twentieth Century Fox Film Corporation.

Parrott, M. 2015. Expressing first-person authority. *Philosophical Studies* 172: 2215–2237. https://doi.org/10.1007/s11098-014-0406-9.

Prinz, J. 2004a. *Gut Reactions: A Perceptual Theory of Emotion*. Oxford: Oxford University Press.

Prinz, J. 2004b. 'Embodies Emotions' In: *Thinking about Feeling: Contemporary Philosophers on Emotions*, edited by Robert C. Solomon. Oxford: Oxford University Press.

Reisenzein, R. 2017. Varieties of Cognition-Arousal Theory. *Emotion Review* 9 (1): 17–26. https://doi.org/10.1177/1754073916639665.

Rey, G. 2013. We Are Not All 'Self-Blind': A Defense of a Modest Introspectionism. *Mind & Language* 28: 259–285. https://doi.org/10.1111/mila.12018.

Roberts, R.C. 2003. *Emotions: An Essay in Aid of Moral Psychology*. New York: Cambridge University Press. https://doi.org/10.1017/CBO9780511610202.

Schachter, S., and J. Singer. 1962. Cognitive, Social, and Physiological Determinants of Emotional State. *Psychological Review* 69 (5): 379–399.

Schwitzgebel, E. 2008. The Unreliability of Naïve Introspection. *Philosophical Review* 117: 245–273.

Smithies, D. 2012. 'A Simple Theory of Introspection'. In: *Introspection and Consciousness*, eds. D. Smithies, D. Stoljar. Oxford: Oxford University Press.

Solomon, R. C. (2004a) 'Emotions, Thoughts, and Feelings'. In: *Thinking about Feeling: Contemporary Philosophers on Emotions,* edited by Robert C. Solomon. Oxford: Oxford University Press.

Solomon, R.C., ed. 2004b. *Thinking about Feeling: Contemporary Philosophers on Emotions*. Oxford: Oxford University Press.