

## **Why Do We Need an Expressivist Account of Self-Knowledge and What Do We Need from It?**

This paper focuses on the expressivist approach to knowledge of our own mental states. In short, I examine the motivations for expressivism, and then consider the question of what a satisfactory expressivist account of self-knowledge should look like.

I begin by examining motivations for expressivism through considering the traditional introspectionist view on self-knowledge, which suggests that the first-person is privileged in an epistemic sense. Introspectionism seems to be able to account for the distinctive features of self-knowledge: immediacy, authority, and salience, and captures our intuition regarding direct access to our own states. However, it can also quickly lead us to the notorious problem of other minds: if we have access only to other people's public behaviours but not to their mental states, we cannot rule out the epistemic possibility of solipsism.

Introspectionism is hence considered problematic, alongside any other theories that could lead to the problem of other minds. The next question is then this: how can we do any better? I suggest that we should start by noting the following point: if the introspectionist suggests that their picture appeals to our folk intuition of the mental, we should also realise that their story only captures one side of our folk conception of the mind. In particular, in over-emphasising the asymmetry between knowledge of our own psychological states and knowledge of other people's, the introspectionist overlooks how our mental lives can *manifest* themselves in outward expressions instead.

This suggests the following motivations for an alternative account of self-knowledge. Namely, to provide a theory which does not face the threat of the problem of other minds, which does not over-emphasise the asymmetry between the first- and third-person perspectives, and accounts

for our folk intuition that we do have a (different) kind of direct access to other people's mental lives.

I then suggest that Wittgensteinian expressivism meets the above desiderata. Later Wittgenstein criticises the Cartesian introspectionist picture, arguing that if mental concepts are learnt only through associating words with private objects, what is 'inside' falls out of our language-game. Instead, he insightfully points out that our concepts of mental states must be partly constituted by the outward expressions of them, and considers avowals (e.g., 'I am in pain') as replacements for natural expressions. 'Simple expressivism' is then the view which treats avowals merely as replacements for natural expressions that are not truth-evaluable. However, this view fails to account for semantic continuity of avowals and implies that the subject stands in no epistemic relation to the content of the avowal, thus failing to account for self-knowledge.

I then assess Bar-On's 'neo-expressivism', which aims to get around these problems of simple expressivism and argues that avowals can express self-knowledge. Bar-On distinguishes between avowals as *acts* (which 'a-express' mental states) and as *products* (which 's-express' propositions) to solve the problem of semantic continuity, and suggests that when one avows that I am in pain, she also expresses her belief or judgment that she is in pain, to solve the epistemic problem. Bar-On concludes that the belief expressed by the avowal is then knowledgeable, due to its being epistemically grounded in the mental state it expresses.

However, I argue that both simple and neo-expressivism face a crucial, shared difficulty: the problem of unexpressed mental states. We frequently have self-knowledge of mental states that we do not express publicly, yet expressivism seems to require expression for self-knowledge. If a stoic spy does not express her pain publicly, the current expressivist proposal implies she lacks self-knowledge of her state. Bar-On attempts to solve this by appealing to 'thought avowals'—articulate thought tokens produced in 'inner speech'. She argues it is natural to think of such inner acts as expressing one's annoyance or outrage.

I reject Bar-On's 'thought avowal' solution on three grounds. First, it is questionable whether such 'inner thinking' can be seen as an expressive behaviour or perform the same expressive role as public avowals. Second, relying on knowledge of a second-order 'thought token' to know a first-order state invites an infinite regress. Third, and most crucially, appealing to inner expressions accessible only to the first-person loses the motivation we started from: avoiding the problem of epistemically privileged access. If self-knowledge relies on inner tokens that is not publicly accessible, the problem of other minds can immediately be reintroduced.

Finally, I briefly sketch a new 'dispositional account' of self-knowledge, which inherits the central ideas of the Wittgensteinian expressivist thought but at the same time overcomes this main problem of both simple and Bar-On's neo-expressivism. I argue that what matters for the Wittgensteinian picture is not that we *actually* express the states, but that we are *disposed* to do so. There is an internal connection between the mental states and expressions of them even when we withdraw from publicly expressing them. Hence, according to the dispositional view, we can have knowledge of these unexpressed states via awareness of the disposition to express.

This view essentially offers a reduction of the problem of self-knowledge of mental states to that of self-knowledge of agency and actions in general. Following Anscombe, I posit that we generally have awareness of our voluntary behaviours and our disposed actions under normal circumstances. This account avoids the problem of other minds because it appeals to disposed expressions which *could* be publicly available if circumstances merited, and rejects some of the key assumptions that lead to common versions of the problem of other minds. By shifting the focus from actual expression to the disposition to express, this account preserves the insights of expressivism while robustly explaining our knowledge of unexpressed states.