**Wittgenstein, self-knowledge and nature**

Annalisa Coliva

University of California Irvine

1. **Introduction**

The aim of this chapter is to explore Wittgenstein’s suggestive remarks about self-knowledge—that is, our knowledge of our own mental states—and their connection with naturalism. As is well known, Wittgenstein was critical of naturalizing the epistemology of mind in behaviourist or in more physicalist-oriented terms. Yet, he placed great emphasis on the role of instinctive and natural manifestations of our own mental states. These natural manifestations were necessary, in his opinion, to acquire the relevant psychological concepts and thereby get in a position to give new expression to our mental states through language. Furthermore, he thought that these verbal manifestations would become “second nature” to us. So much so, that just as it would not make sense to regard a spontaneous cry out of fear as based on having that emotion in view and on giving expression to it through the cry, he thought that at least some of our psychological “avowals” should not be considered as judgements based on, and justified by having the relevant mental states in view. Rather, he thought they should be seen as immediate and spontaneous, though culturally ingrained, expressions of the mental states that elicited them. *A fortiori*, in his view, psychological avowals properly so regarded should not be considered as the result of inference to the best explanation starting from the observation of one’s overt behaviour.

Wittgenstein put his own variety of naturalism at the service of *dissolving*—rather than solving—the very problem of self-*knowledge*. To put it crudely, in Wittgenstein’s perspective there is no real epistemology of the mind, when it comes to (at least some of) our psychological self-ascriptions. For, despite their surface grammar, they are not linguistic manifestations of true and justified, or reliably formed, beliefs about our own mental states, reached through the observation of one’s own mental states or by means of inference to the best explanation starting with the observation of one’s overt behavior. Rather, their function is merely expressive of one’s on-going mental states and their occurrence is just as instinctive and not observationally or inferentially mediated as a cry or a sigh.

For these reasons, Wittgenstein is rightly considered the father of both contemporary expressivism and of constitutivism regarding psychological avowals. In particular, his remarks can be seen at the origin of the idea—central to contemporary expressivism—that the main function of (at least some of) our psychological avowals is expressive rather than descriptive. Contemporary expressivists, however, have also tried to move away from Wittgenstein’s anti-epistemological outlook.[[1]](#footnote-1) Contemporary constitutivists, by contrast, are united in rejecting the idea that, at least in some central cases, our psychological self-ascriptions are underwritten by an appropriate epistemic relation linking the subject and her first-order mental states, such that the former can actually be taken to manifest true and appropriately justified beliefs about one’s own mental states. Constitutivists, however, are critical of the idea that the relevant self-ascriptions serve merely an expressive function and that they replace forms of more instinctive behaviour.[[2]](#footnote-2)

In my own work on self-knowledge, I have tried to combine these two perspectives, by endorsing the anti-epistemological outlook characteristic of constitutivism and by retaining a role for the idea that psychological concepts are primarily acquired as ways of expressing, rather than naming, an on-going mental state. Still, in my view, these concepts can and indeed are used to make judgements, whose characteristic traits—such as transparency, authority and groundlessness—[[3]](#footnote-3)are not guaranteed by any peculiar epistemic relation holding between a subject and her own mental states. Rather, they are features that hold a priori, once these judgments are seen as falling within the scope of appropriately qualified C-conditions.[[4]](#footnote-4)

These developments will not be the subject of the present chapter. What is important to note, however, is that Wittgenstein was well aware of the limitations of his own expressivist position. He was very careful—indeed much more careful than several contemporary expressivist theorists—not to over-generalize the expressivist treatment. For he was very mindful of the fact that exactly the *same* twists of phrase could sometimes be used to express judgments about our own first-order mental states and that, in those cases, they would manifest a subject’s own beliefs regarding her first-order mental states, reached through a variety of epistemic methods, all open to error, at least in principle. Hence, just as much as he can rightly be seen as the ancestor of contemporary expressivism and constitutivism with respect to self-knowledge, he should actually be seen as the father of contemporary *pluralism* regarding self-knowledge as well.[[5]](#footnote-5)

1. **Against introspectionism and behaviourism: The case of sensations**

In this section, I review Wittgenstein’s criticism of introspectionism and behaviourism in order better to situate his positive proposal with respect to self-knowledge. These remarks will also help us see what kind of naturalisms Wittgenstein’s positive proposal eventually stands opposed to.

Let us therefore start with his attack on introspectionsim. As understood here, introspectionism is the view according to which mental states are luminously presented to us—that is, their nature and, when applicable, their content—would be something they would wear on their sleeves and that we would know by perceiving them (or, more mildly, by attending to them), while paying attention to their distinctive elements. Just to fix idea, consider pain. It seems extremely natural to think that that sensation has its own distinctive phenomenology (or quale) and that in order to self-ascribe it, we need only to pay attention to its occurrence, recognise it for what it is and self-ascribe it, through an exercise of the relevant concepts. These, in turn, would have to be acquired, at least in the case of the concept of pain, by creating a mental word, demonstrative concept, or file—depending on your favorite account of atomic concepts—in response to the occurrence of that very sensation with its distinctive phenomenology.

It is quite clear that although Wittgenstein criticized the idea of a private *language* in the *Philosophical Investigations*, his remarks could be applied to the case of the *concept* of pain, understood along the lines just sketched. The key analogy is given by the fact that unless one confined that concept to a single occurrence of a demonstrative—viz. that—used in connection with a specific occurrence of a given mental state—Mi—,[[6]](#footnote-6) that concept would have to have either a predicative or a complex demonstrative structure—it should be either pain or that pain—and should be applicable in potentially many occasions. The first encounter with its referent on one occasion should therefore be able to set up a rule for the future employment of that concept. The very point Wittgenstein makes against the idea of a private language could then be raised against the notion of a private concept of pain. For a procedure of private ostensive definition of a word like ‘pain’ or of a concept like pain or that pain cannot set up a standard against which future applications could be deemed correct or incorrect: “whatever is going to seem right to me is right. And that only means that here we can’t talk about ‘right’” (PI 258). That is to say, if ‘pain’/pain/that pain is the name I give to a sensation of mine, or else it is a demonstrative or predicative concept whose content is given by that sensation, its use over time counts as correct whenever I would think so, even when I may use it to refer or think to a different kind of sensation or to nothing at all. This, however, simply means that the distinction between correct and incorrect uses of that term/concept has vanished. If that distinction can no longer be drawn, though, there is no rule-governed use of the term. Thus, the word has no meaning, given Wittgenstein’s normative conception of language. *Mutatis mutandis*, the private ostensive definition would only seem to give rise to a rule for the future employment of the concept. Hence, the putative concept pain (or the concept that pain) would just be an impression of concept.[[7]](#footnote-7)

Furthermore, the idea that we are acquainted with our sensations, which are the referents of our psychological terms (or concepts) such as “pain”, would lead to solipsism. That is to say, to the idea that one can know only one’s own mind and merely surmise what goes on in other people’s minds, if indeed there are other people at all. For one cannot know what another person is referring to when she uses those very terms, not even by analogy with one’s own case, since their referents are foreclosed to one. Indeed, one cannot even know whether another person is referring to anything at all, by using those terms. Hence, one cannot know if it is person—that is, someone capable of having sensations (and other mental states)—rather than a creature who resembles us in its looks and behavior but who is not really enjoying any mental state.

However, if Wittgenstein is clearly against introspectionism for the reasons just rehearsed, he is equally critical of behaviorism, despite the fact that his expressivism might be taken—and indeed, it has been taken—to be a covert form of it. Crude behaviorism has it that there are no mental states, just outward behavior. Although, as we shall presently see, for Wittgenstein the language of sensations supervenes on, and sometimes replaces their natural and instinctive manifestations, he is careful to stress that this is not tantamount to denying the existence of sensations and further mental states. Surely, moreover, he was all against the idea that we would know of our sensations through inference to the best explanation starting from the observation of our own overt behaviour. What he did object to is simply the way in which, following introspectionism, we are led to think of the language of sensations after the name-object referential model, which immediately leads to the endorsement—albeit implicitly—of the idea of a private language. As he wrote in a series of telling passages:

Suppose everyone had a box with something in it: we call it a “beetle”. No one can look into anyone else’s box, and everyone says he knows what a beetle is only by looking at *his* beetle.—Here it would be quite possible for everyone to have something different in his box. One might even imagine such a thing constantly changing.—But suppose the word “beetle” had a use in these people’s language?—If so it would not be used as a name of a thing. The thing in the box has no place in the language game at all; not even as a *something*: for the box might even be empty.—No, one can “divide through” by the thing in the box; it cancels out, whatever it is. That is to say: if we construe the grammar of the expression of sensation on the model of “object and designation” the object drops out of consideration as irrelevant. (PI 293)

“Are you not really a behaviorist in disguise? Aren’t you at bottom really saying that everything except human behavior is a fiction?”—If I do speak of a fiction, then it is of a *grammatical* fiction. (PI 307)

The origin of this “grammatical” fiction, for Wittgenstein, resides in the fact that we talk of mental processes and states, leave their nature undecided, yet fail to understand them because the analogy with states and processes “falls to pieces”.

So we have to deny the yet uncomprehended process in the yet unexplored medium. And now it looks as if we had denied mental processes. And naturally we don’t want to deny them (PI 308).

Again:

“So you are saying that the word ‘pain’ really means crying?”—On the contrary: the verbal expression of pain replaces crying and does not describe it”. (PI 244)

Now, if the language of sensations and mental concepts do not work—in fact, cannot work—on the basis of a name-object referential model, how do they work? Wittgenstein raises and answers this very question in PI 244:

How do words *refer* to sensations?—There doesn’t seem to be any problem here; don’t we talk about sensations every day, and give them names? But how is the connection between the name and the thing named set up? This question is the same as: how does a human being learn the meaning of the names of sensations?—of the word ‘pain’ for example. Here is one possibility: words are connected with the primitive, the natural, expressions of the sensation and used in their place. A child has hurt himself and he cries; and then adults talk to him and teach him exclamations and, later, sentences. They teach the child new pain-behavior.

Hence, according to Wittgenstein, subjects do have sensations, which have primitive and natural expressions. For instance, in the case of pain it will be crying and moaning. Adults teach children to accompany and sometimes replace that primitive behavior with a linguistic one, which becomes more and more articulated, up to making use of an entire sentence like “I am in pain”, after having gone through other less complex linguistic formulations, like “Ouch!”[[8]](#footnote-8) and “Pain!”. Yet, the *function* of that eventual sentence is not to describe one’s inner state of mind, nor, for that matter, to describe one’s primitive pain-behavior. Rather, it is to express and give voice to one’s sensation, just as a cry is the immediate expression of one’s on-going pain. “I am in pain”, therefore, is not the linguistic manifestation of one’s belief to be in pain, reached through an inner procedure in which one’s present sensation is recognized as pain and named that way. “I am in pain”, for Wittgenstein, is just the refined linguistic, yet by no means less immediate, expression of pain.

Several consequences follow from this alternative picture of the workings our psychological language. First, it can be apprehended only via one’s instinctive behavioral manifestations. Hence, for Wittgenstein:

“But doesn’t what you say come to this: that there is no pain, for example, without *pain-behavior*?”—It comes to this: only of a living human being and what resembles (behaves like) a living human being can one say: it has sensations; it sees; is blind; hears; is deaf; is conscious or unconscious”. (PI 281)

An “inner process” stands in need of outward criteria. (PI 580)

The human body is the best picture of the human soul. (PI II, iv, p. 178e)

The idea is that only of a living human body (and more generally of a living animal body that behaves similarly to us), *contra* a human corpse, or a zombie, can we correctly say that it does have sensations and feelings and can thus teach a subject to express them in linguistic terms.

Secondly, “I am in pain” has an expressive function like “Ouch!” or “Damn it!”. That is, even if its surface grammar is that of a descriptive sentence, which can be meaningfully denied and embedded in suppositions—“Suppose I am in pain” —and therefore in conditional statements—“If I am in pain, I go to see the doctor”—, its function is not descriptive. Hence, either it is not up to semantic evaluation or else, if it is, it is only in a minimal sense.

Wittgenstein is often considered one of the inspirational sources of minimalism about truth. The idea, that is, that there is no more to truth than its disquotational function— “Grass is green” is true iff grass is green—and similar platitudes—if “Grass is green” is true, “Grass is not green” is not true. Surely, Wittgenstein flirts with minimalism from time to time, but it is not clear that he really endorsed it, at least in connection with psychological avowals.[[9]](#footnote-9) Rather, as we shall presently see, he is more drawn towards a radical form of semantic contextualism.[[10]](#footnote-10)

Thirdly, not all uses of the very same sentence are on a par, for him. For, as we have just seen, “I am in pain” is very often an avowal, hence an immediate linguistic expression of one’s on-going pain. Yet, that very sentence can be used differently when it is embedded in negation or in wider, especially suppositional contexts. However, given Wittgenstein’s equation of meaning and use (PI 43),[[11]](#footnote-11) that is tantamount to saying that, despite the identical linguistic form, different occurrences of “I am in pain” can have different meanings. Indeed, we explain their function, that is, their use, differently. Since, for Wittgenstein, “the meaning of a word is what is explained by the explanation of the meaning” (PI 560), that means that the very same sentence can actually have an altogether different meaning depending on the context of its utterance, not just a different function. In particular, Wittgenstein distinguishes between the avowing, merely expressive function, and the descriptive one, which occurs when, in the context of a conditional statement, for instance, we suppose being in pain and describe what we would or would not do, just as we would do in the case of another person.

As we shall see in the following, for Wittgenstein it is important to realize the existence of these semantic asymmetries we are often oblivious to, given our tendency to think of language as working in a uniform manner. In particular, it is important to realize that even psychological *self*-ascriptions can depend on assuming a *third-personal* stance with respect to oneself. Still, different occurrences of “I am in pain” in an avowing and in a descriptive mode are related to one another and are not like “I went to the bank” when used to refer to one’s going to the financial institute or to the bank of the river. In both cases, the criteria for using that sentence depend on characteristic forms of behavior. In the one case, the sentence replaces the instinctive behavior elicited by an on-going painful sensation; in the other, it is used after a process of inference to the best explanation, based on one’s reflection on one’s behavior over time, which is not the one elicited by an on-going painful sensation, but still produced by distressing feelings or moods. Moreover, when it comes to “I am in pain” and to “S is in pain”, it should be kept in mind that one’s avowals of pain replace an instinctive kind of behavior, which is the same kind of behavior that allows us to say of some other person that she is in pain, even though we could never be in a position to *avow* her pain. Nevertheless, despite the fact that the similarity in human reactions could be used to build some bridge between the use in the first-person and in the third-person present of our psychological vocabulary, Wittgenstein is very much attracted to a strong form of semantic contextualism,[[12]](#footnote-12) which tends to deny the uniformity of meaning of “pain” across changes of person and tense. So there is a family resemblance between “pain” when avowed and “pain” when used to describe one’s own dispositions or someone else’s pain, but no identity in meaning.

1. **Wittgenstein’s anti-epistemological account**

There is a fourth element to Wittgenstein’s overall account that needs to be taken into consideration carefully. Namely, his idea that the use of “I know” in connection with one’s own psychological avowals is problematic. Here are some famous quotes:

In what sense are my sensations *private*?—Well, only I can know whether I am really in pain; another person can only surmise it.—In one way this is wrong, and in another nonsense. 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 behavior, for *I* cannot be said to learn 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. (PI 246; cf. PI II, xi, p. 221e)

I can know what someone else is thinking, not what I am thinking. It is correct to say “I know what you are thinking”, and wrong to say “I know what I am thinking”. (A whole cloud of philosophy condensed in a drop of grammar). (PI II, xi, p. 222e)

Hence, according to Wittgenstein, other people can know my own mental states. Thus, the typically solipsist claim “Only I can know my own pain” is false. Others do too, even though not by analogy with what happens in their own case, but through the kind of pain behavior that I exhibit. In fact, “I know I am in pain” is nonsensical, when “I am in pain” is an avowal, and if “I know” is taken to express an epistemic relation between a subject and a proposition, or even a fact. Why so? To become clear about that, it is useful to recall the criteria for the correct and meaningful use of “I know” Wittgenstein presents in *On Certainty*.[[13]](#footnote-13) Here is a quick summary of them:

1. One must have reasons for one’s knowledge attribution;
2. These reasons must be stronger than what they are supposed to ground;
3. There must be a method to find out whether what one claims to know is the case;
4. The knowledge claim must be relevant;
5. It must make sense to say “I do not know”.

All these criteria would be violated in the case of “I know I am in pain”. For what could one offer as a reason for “I know I am in pain” apart from just repeating that one is in pain? (1) would thus be flouted. One might think that the reason is the sensation itself. Surely, Wittgenstein would have objected to that, since, in his view, sensations as such cannot be reasons for their ascriptions. As he famously wrote (PI 289):

“When I say ‘I am in pain’ I am at any rate justified *before* *myself*”.—What does that mean? Does it mean: “If someone else could know what I am calling ‘pain’, he would admit that I was using the word correctly”?

Clearly, the question must be answered in the negative, for him, and in fact, he goes on to say:

To use a word without a justification does not mean to use it without right [*zu Unrecht*].[[14]](#footnote-14)

Now, in the recent literature on Wittgenstein’s epistemology, this remark is often at the basis of epistemic interpretations, which tend to attribute to Wittgenstein an early endorsement of non-evidential warrants or, as they are usually called nowadays, “entitlements”.[[15]](#footnote-15) As a matter of fact, however, to have a right to use the expression “I am in pain”, in this context, is not to be interpreted in an epistemic way. What Wittgenstein is saying, rather, is that even if “I am in pain” is not justified by one’s on-going pain, to use that sentence to avow and therefore to express or exhibit one’s pain is perfectly correct. In other words, the use of that sentence is legitimate even if it is not backed by *any* epistemic warrant or guarantee.

Let us suppose, however, that one’s sensations were one’s reasons for the relevant self-ascriptions. In that case, they would certainly not be any stronger than the very knowledge claims they are supposed to ground (*contra* (2)). As to condition (3), clearly one does not find out whether one is in pain (at least when “pain” is meant as meaning an occurrent sensation and not a dispositional state one enjoys and self-attributes through observation and inference to the best explanation). One is or has pain and gives voice to it. Furthermore, a claim like “*I know* I am in pain” would—in the normal run of cases—be totally irrelevant (thus flouting (4)) and indeed odd in the course of a conversation, up to the point that, if no proper context for its occurrence could be provided, one would start doubting whether a subject is in her right mind, or would start thinking that perhaps she is trying to make a joke, etc. Finally, in the normal run of cases, it does not make sense to suppose that one might not know whether one is in pain (*contra* (5)). Of course there can be cases of confused sensations—an itch which borders pain—but the phenomenon of vagueness cannot support the idea that when we experience sensations we may not know that we do.[[16]](#footnote-16) As to the possibility of unconscious sensations, they would be a conceptual solecism in this connection. Thus, for Wittgenstein, the use of “I know” in connection with one’s avowals would not make sense.

Still there is a *grammatical* sense of “I know” which would be appropriate in connection with one’s psychological avowals. Again, the existence of such a grammatical use of “I know”, for Wittgenstein, is clear from his remarks in *On Certainty*. Two aspects of the grammatical use of “I know” are particularly relevant with respect to avowals. Namely: (1’) the fact that “I know” would actually mean “A doubt is excluded” or “I cannot be wrong” (OC 59) and (2’) that the “I” is unimportant (OC 58). Here are the relevant passages in the *Philosophical Investigations*:

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. (PI 246)

If anyone said “I do not know if what I have got is a pain or something else”, we should think something like this, he does not know what the English word “pain” means; and we should explain it to him.—How? Perhaps by means of gestures, or by pricking him with a pin and saying: “See, that’s what pain is!” This explanation, like any other, he might understand right, wrong, or not at all. And he will show which he does by his use of the word, in this as in other cases.

That expression of doubt has no place in the language-game; but if we cut out human behavior, which is the expression of sensation, it looks as if I might *legitimately* begin to doubt afresh. My temptation to say that one might take a sensation for something other than what it is arises from this: if I assume the abrogation of the normal language-game with the expression of a sensation, I need a criterion of identity for the sensation [itself]; and then the possibility of error also exists. (PI 288)

And with respect to (2’), the idea that “I” is not a referential expression when used in psychological avowals, which is present in Wittgenstein since his remarks in the *Blue Book* (pp. 66-67), makes its appearance again in PI 404-411. Here are some passages:

In saying this [“I am in pain”] I don’t name any person. Just as I don’t name any person when I *groan* with pain. Though someone else sees who is in pain from the groaning. (PI 404)

It would be possible to imagine someone groaning out: “Someone is in pain—I don’t know who!”—and our then hurrying to help him, the one who groaned. (PI 407)

“But you aren’t in doubt whether it is you or someone else who has the pain!”—The proposition “I don’t know whether I or someone else is in pain” would be a logical product, and one of its factors would be: “I don’t know whether I am in pain or not”—and that is not a significant proposition. (PI 408)

“I” is not the name of a person, nor “here” of a place, and “this is not a name”. But they are connected with names. (PI 410).

To say that “I know” in connection with one’s sensations could have a grammatical use, at most, means to say that there is no inner epistemology. Furthermore, it means to say that it is a characteristic trait of our linguistic practice that subjects’ avowals are not challenged, in the relevant circumstances, unless there are reasons to doubt of their sincerity. As Wittgenstein writes, although in the context of explaining away, as we may put it, the epistemology of self-ascriptions of intentions:

“Only you can know if you had that intention.” One might tell someone this when one was explaining the meaning of the word ‘intention’ to him. For then it means: *that* is how we use it. (And here “know” means that the expression of uncertainty is senseless). (PI 247)

It is important to dwell a bit longer on what it means for a remark to be grammatical, according to Wittgenstein. The notion of grammar, which is subject to continuous development over the years, from the immediately post-*Tractarian* works up to Wittgenstein’s last collection of notes, namely *On Certainty*, is connected to the ones of analyticity, aprioricity and of conceivability. Early examples of grammatical sentences, for Wittgenstein, are (i) “An object cannot be of two different colours all over its surface at the same time” or (ii) “Patience is played alone”. The idea, on the one hand, is that these are meaning constitutive propositions. Thus, it is constitutive of what we mean by “patience” that it is played alone. If we changed the rules of playing patience and allowed for multiple players, the very meaning of that word would alter. Similarly for what we call “(physical) object”. Now grammar is “autonomous” for Wittgenstein. This means that it is not answerable to reality. This is clear from Wittgenstein’s treatment of grammatical propositions such as (i): we do not hold (i) because objects cannot physically be of two different colours all over their surface at the same time. Nor is it because the structure of our sensory experience is of that kind and excludes the possibility of perceiving an object as having two colours all over its surface. The order of explanation goes the other way round, in his opinion: given our concepts, reality and experience are constrained in the way (i) prescribes. Similarly, it is not because imagination fails us that we cannot conceive of an object as being of two colours all over its surface at once (cf. PI 251). Rather, if by “object” we mean what we do, we would not know where to start from in order to conceive of a physical object as having two different colours all over its surface at the same time. Hence, to say that “I know I am in pain” can at most be a grammatical proposition means to say that it makes explicit a fact about how we use the vocabulary of sensations, or equivalently for Wittgenstein, a structural aspect of our conceptual scheme. Hence, in the normal run of cases, subjects’ pronouncements over their on-going sensations are taken at face value. Moreover, of someone who sincerely expressed a doubt about whether she is or is not having an on-going sensation, we would not understand what she might actually mean (bar vagueness). The impression of those words as having nonetheless a meaning would be due to the projection of meaning onto them from their ordinary contexts of use, in which they do have meaning.

Still, in philosophy, according to Wittgenstein, we constantly run the risk of mistaking grammatical propositions for empirical ones, yet as having the remarkable property of indubitability. That is how the ideas of metaphysical necessity and of infallibility arise, in his view. When taken that way, however, Wittgenstein’s reaction is to declare them nonsensical; when understood as being grammatical, instead, his reaction is to find them obvious. As he writes:

“This body has an extension”. To this we might reply: “Nonsense!”—but are inclined to reply “Of course!”—Why is this? (PI 252)

The answer to this question is precisely that, underneath their superficial identity, different tokens of the same sentence may actually be employed in such different a way as to amount to nonsense in one case, or to grammatical claims in the other case. The mark of the latter is that we find the statement obvious, just as is obvious to anyone who knows the meaning of “patience” or of “physical object” that patience is played alone and that physical objects don’t disappear of their own making if not perceived, and that we categorize plenty of things as falling under that category. “This hand is a physical object”, then, is obvious to anyone who speaks our language and who has been brought up within our community.

1. **Propositional attitudes: expressivism and genuine self-knowledge**

Let us now turn to Wittgenstein’s account of propositional attitudes and of our knowledge of them. According to him, we should pay attention to the variety of mental states we can enjoy. In PI 574 he writes: “The concepts of believing, expecting, hoping are less distantly related to one another than they are to the concept of thinking”. The difference depends on the fact that thinking is occurrent and has salient phenomenological aspects to it. Believing, expecting and hoping, in contrast, have a dispositional element to them, and there is an internal relation between these attitudes and their contents. That is to say, it is constitutive of these attitudes that they have a certain content and the content they have individuates each of their tokens as the particular belief, desire or intention it is. This means, for instance, that if some external cause stops me from desiring that P, it does not mean that my original desire was a desire that had that cause as its object. To exemplify, if someone punches me and causes me to stop wanting to have an ice-cream it does not mean that I wanted to be punched. This relates also to Wittgenstein’s discussion of actions (PI 611-648, cf. BlB 23-24). Intentional actions have reasons and reasons, for him, cannot be their causes. In general, causes may be “detached” from actions without turning them into different ones, whereas reasons cannot. Intentions, in contrast, are internally related to their contents. They are not the causes of the actions that fulfil them (PI 632, cf. BlB 23-24). In his view, moreover, intentional actions and willing do not stop short of the action willed (PI 622). Although there may be cases in which trying and making an effort to bring about the action are relevant. Nor are they individuated by specific (bodily) feelings (PI 625), although they may be accompanied by them.

When it comes to our knowledge of intentions and actions (as well as of beliefs and desires), it depends on the circumstances. In the usual run of cases, we express, avow or voice our own intentions. In that case, there is no real epistemology of the mental, for the reasons we rehearsed in connection with (putative) knowledge of our own sensations. That is, the criteria for the empirical and therefore meaningful use of “I know” are flouted. To go over them once more: we do not have independent reasons for “I know I ψ that P” (where ψ is a propositional attitude verb like intending, desiring, believing, hoping, willing, etc.). That is, reasons other than the very mental state itself (*contra* (1)). A fortiori, even if one conceded that the mental state could be a reason for our knowledgeable self-ascriptions, it would not be any stronger a reason than what it is supposed to ground (*contra* (2)). We do not find out our intentions, beliefs, etc. We do have them and give immediate expression to them by saying or thinking “I intend to φ” or “I /believe/desire that P”. We do not discover them either through introspection or inference to the best explanation given our behavior or feelings, which may, at least on occasion, accompany the occurrence of these mental states (*contra* (3)). To say “I know I ψ that P” in ordinary circumstances would not count as making a relevant contribution to the conversation. In particular, asserting it would not add anything to one’s simple self-ascription of the relevant mental state (*contra* (4)). Finally, in the relevant circumstances, it would not make sense to say or judge “I do not know whether I ψ that P” (*contra* (5)). For that kind of judgement could only meaningfully occur in a context in which I haven’t yet formed the (specific) intention (or the belief or the desire).

Now, if the expressivist story has to get purchase in relation to propositional attitudes like intending, believing and desiring, it will have to be the case that the relevant self-ascriptions are learned as ways of accompanying and possibly replacing pieces of instinctive behavior. The question arises, however, whether there is such a distinctive kind of instinctive behavior, which would manifest our intending, desiring or believing. While with pain it seems clear that there is a characteristic pre-linguistic manifestation, with propositional attitudes it is not that obvious. Here is Wittgenstein’s reply:

What is the natural expression of an intention?—Look at a cat when it stalks a bird; or a beast when it wants to escape. (PI 647)

The idea then is that also when it comes to wanting, intending and desiring there are characteristic primitive reactions. For instance, a child wants to have a glass of water, tries to reach for it, the adult sees her doing so and says “You want a glass of water. I’ll give it to you”. Again, a child puts her coat on and stands by the door. The adult says “You want to go out. We’ll go in a moment”. Little by little, the child is taught to accompany her primitive behavior with pieces of linguistic vocabulary, up to the point where she herself is in a position to say “I want a glass of water/to go out”.

The case of belief, however, is certainly more complicated. For there do not seem to be specific primitive reactions which would signal one’s believing that today it is sunny, or that one’s favourite toy is soft and tender. In particular, even if the child reaches for her toy and holds it close to her face, we do not normally teach her to replace that behavior with a linguistic one that makes explicit the attitude of belief. That is to say, we do not say “You believe that your teddy-bear (say) is tender”, even if, to us, the child’s behavior manifests that attitude. We simply voice the embedded content of the attitude. It is perhaps for this reason that when Wittgenstein explicitly talks about self-ascriptions of belief in the present tense, he takes them to be equivalent to the assertion of their embedded content, just in a more tentative way.

In PI II, x, while addressing Moore’s paradox—“I believe it is raining, but it isn’t”—Wittgenstein observes, first, that “I believe that” is “transparent”.[[17]](#footnote-17) For “The expression ‘I believe that this is the case’ is used like the assertion ‘This is the case’” (p. 190e). Perhaps, sometimes, as a “hesitant assertion [of P]” (p. 192e). This leads him to hold that “to believe” has a different meaning when used in the first person present, as opposed to in the first person past, or within a suppositional context, or in the third person. Here are some relevant passages, whose significance is better appreciated by keeping in mind Wittgenstein’s equation of meaning and use (PI 43, 560):

Moore’s paradox can be put like this: the expression “I believe that this is the case” is used like the assertion “This is the case”; and yet the *hypothesis* that I believe this is the case is not used like the hypothesis that this is the case. (PI, II, x, p. 190e)

Similarly: the statement “I believe it is going to rain” has a meaning like, that is to say a use like, “It’s going to rain”, but the meaning of “I believed then that it was going to rain”, is not like that of “It did rain then”. (ibid.)

Of course, we do not find this claim intuitive: semantic continuity across changes of tense and person seems obvious to us. Here is Wittgenstein’s response:

“But surely ‘I believed’ must tell just the same thing in the past as ‘I believe’ in the present!”—Surely √-1 must mean just the same in relation to -1, as √1 means in relation to 1! This means nothing at all. (ibid.)

Hence, the idea seems to be that since √1 admits of two possible solutions—1 and -1—both of which belong to real numbers, while √-1 does not admit of a solution within the domain of reals, “√” means something different in the two cases. Although the solutions of the application of the function are different depending on whether it is applied to 1 or -1, and indeed in order to provide a solution to √-1 we need to extend numbers to imaginary ones, while no such an extension is needed in order to provide a solution to √1, that does not mean that the function itself is ambiguous. Thus, the analogy provided by Wittgenstein to support a strong form of contextualism about meaning is dubious. Indeed, while one may happily agree that the *function* of a sentence may vary contextually—from playing an expressivist role, to playing a descriptive one, and that the criteria of its legitimate employment may vary accordingly—there is no need to follow Wittgenstein all the way down up to saying that the *meaning* of the words occurring in it would change as well.

Notwithstanding Wittgenstein’s strong form of contextualism about linguistic meaning, there is something else worth noticing. Namely, the relevance to an appropriate treatment of self-knowledge of his paying close attention to the different functions the same words, conceding that they may retain the same linguistic meaning, can have depending on context. He writes:

This is how I think of it: Believing is a state of mind. It has duration; and that independently of the duration of its expression in a sentence, for example. So it is a kind of disposition of the believing person. This is shown me in the case of someone else by his behavior; and by his words. And under this head, by the expression “I believe …” as well as by the simple assertion.—What about my own case: how do I myself recognize my own disposition?—Here it will have been necessary for me to take notice of myself as others do, to listen to myself talking, to be able to draw conclusions from what I say! (PI II, x, p. 191e-192e)

Does it make sense to ask “How do you know that you believe?”—and is the answer: “I know it by introspection”? In *some* cases it will be possible to say some such thing, in most not. (PI 587)

What Wittgenstein seems to suggest here is that “S believes that P” is used to ascribe a dispositional state of mind and that “I believe that P” may, on occasion, although quite rarely, be used to describe one’s own dispositions and be applied on the basis of the same criteria that govern third-personal ascriptions of belief. In the vast majority of cases, however, “my own relation to my words is wholly different from other people’s. That different development of that verb would have been possible, if only I could say ‘I seem to believe’” (p. 192e, cf. RPP II, 3). So, in some unusual cases we would know of our beliefs as dispositions in a third-personal way—that is, by inferring from the observation of our behavior, as well as of our words, to their likely mental cause. In those cases, Wittgenstein notices, “it would also be possible for someone to say ‘It is raining and I don’t believe it’, or ‘It seems to me that my ego believes this, but it isn’t true’”. Hence, in these cases, Moore’s paradox would actually disappear.[[18]](#footnote-18) Still, in the vast majority of cases, “I believe that P” is used in an expressive way and, for him, is tantamount to asserting “P”, albeit in a tentative voice, so to speak. This, as I have argued elsewhere,[[19]](#footnote-19) is deeply wrong, even though it is right to remark upon the variety of uses of the very same words. Other telling passages are the following ones:

We say “I am expecting him”, when we believe that he will come, though his coming does not *occupy our thoughts* (…). But we also say “I am expecting him” when it is supposed to mean: I am eagerly waiting for him. We could imagine a language in which different verbs were constantly used in these cases. And similarly more than one verb where we speak of “believing”, “hoping”, and so on. Perhaps the concepts of such a language would be more suitable for understanding psychology than the concepts of our language. (ibid.)

When someone says “I hope he’ll come”—is this a *report* about his state of mind, or a *manifestation* of his hope?—I can, for example, say it to myself. And surely I am not giving myself a report. It may be a sigh; but it need not. If I tell someone “I can’t keep my mind on my work today; I keep on thinking of his coming”—*this* will be called a description of my state of mind. (PI 585)

In PI 586 Wittgenstein makes similar observations and then goes on to draw a difference between the “exclamation” ‘I’m longing to see him!’’ and a different use of the same words: “But I can utter the same words as the result of self-observation, and then they might mean: ‘So, after all that has happened, I am still longing to see him’”. In PI 587, talking about love, he writes: “It makes sense to ask: ‘Do I really love her, or am I only pretending to myself?’ and the process of introspection is the calling up of memories; of imagined possible situations, and of the feelings that one would have if …”. In PI 588, discussing the role of self-ascriptions of intentions, Wittgenstein writes: “‘I am revolving the decision to go away tomorrow’ (This may be called a description of a state of mind.) (…). I say at the end of a quarrel ‘All right! Then I leave tomorrow!’; I make a decision”. References could be multiplied *ad libitum*. They all go in the direction of distinguishing between different uses or functions the same words can have and, at least in the case of self-ascriptions of intentions, between the self-ascription of a disposition and of an on-going mental state, with the attendant diagnosis of their underlying epistemology. None in the latter case, and third-personal in the former.

1. **Conclusion**

We have now examined at length Wittgenstein’s expressivism. We can thus sum up how it fares with respect to an adequate theory of self-knowledge. In the case of *first*-personal self-knowledge,[[20]](#footnote-20) it respects groundlessness—that is, the idea that one’s psychological avowals are not based on the observation of one’s own mental states, or on inference to the best explanation starting with the observation of one’s own behaviour. Nor are they justified, for Wittgenstein, by one’s own going mental states. Indeed, avowals are not expressions of beliefs about one’s own mental states, which may be more or less justified, or not at all. Rather, they are immediate expressions of the mental states one is in. Hence, they are simply not up for being epistemically supported one way or the other.

It also respects transparency—that is, the fact that if one has (or is in) the mental state M, one is immediately in a position to avow it, provided, of course, one has the relevant concepts. Indeed, the expressive role of avowals has built into it the very idea that they are direct linguistic expressions of one’s on-going mental states.

Finally, authority—that is, the fact that in the relevant circumstances a subject’s avowals are not challenged or doubted, unless there is any reason to suspect of a subject’s sincerity—is respected as well. In particular, it is seen as a product of our linguistic practice. That is to say, once the relevant conditions obtain, it is part of our practice of making psychological avowals that there is no room for doubt regarding the fact that a subject, who does avow her own mental state, actually has it.

At the same time, Wittgenstein was well aware of the fact that many of our psychological self-ascriptions—that is, those having a descriptive function—are epistemically supported. He noticed that they were arrived at not just by inference to the best explanation starting with the observation of one’s overt behaviour, but also by means of inference to the best explanation starting with the observation of one’s inner promptings, as it were. Similarly, he was responsive to the fact that these epistemic procedure might sometimes go wrong and that just as much as the resulting self-ascriptions would often express genuine instances of knowledge, they might fail to do so, at least occasionally.

Still, Wittgenstein’s expressivism is not without problems. For he ended up endorsing a strong form of semantic contextualism, which would pair a change of function with a change in meaning. In connection with avowals of belief and other propositional attitudes, moreover, it is not clear what kind of pre-linguistic behavior avowals would substitute. Finally, their equation with the assertion of their embedded contents is highly problematic, as his treatment of Moore’s paradox reveals.

These problems notwithstanding, Wittgenstein’s treatment of self-knowledge, while critical of what Strawson would have called “hard” naturalism, is certainly hospitable to “soft” or “catholic” forms of it.[[21]](#footnote-21) The stress on the instinctive behavioural manifestations and the idea that avowals are linguistic accompaniments or even replacements of those, and are thus constitutive our “second nature”, clearly testifies to that. Equally, his stress on the possibility for some of our psychological self-ascriptions to express genuine self-knowledge, arrived at through inference to the best explanation starting from various kinds of promptings, attests to his endorsement of the application of *third*-personal, naturalistically amenable, epistemic methods to acquire genuine knowledge of our own mental states.

**References**

Bar-On, D. (2004). *Speaking My Mind. Expression and Self-Knowledge*. Oxford: Oxford University Press.

Bilgrami, A. (2006). *Self-Knowledge and Resentment*. Cambridge (MA): Harvard University Press.

Chalmers, D. (2003). “The content and epistemology of phenomenal belief”. In Q. Smith and A. Jokic, (eds.), [*Consciousness: New Philosophical Perspectives*](http://philpapers.org/rec/SMICNP), Oxford: Oxford University Press, 220-72.

Coliva, A. (2010). *Moore and Wittgenstein. Scepticism, Certainty and Common Sense*. London: Palgrave.

Coliva, A. (2015). How to commit Moore’s paradox. *The Journal of Philosophy*, CXII(4), 169-192.

Coliva, A. (2016). *The Varieties of Self-Knowledge*. London: Palgrave.

Fodor, J. (1998). *Concepts. Where Cognitive Science Went Wrong*. Oxford: Oxford University Press.

Jacobsen, R. (1996). Wittgenstein on self-knowledge and self-expression. *Philosophical Quarterly,* 46, 12-30.

Russell, B. (1912). *The Problems of Philosophy*. Oxford: Oxford University Press.

Shoemaker, S. (1996). *The First Person Perspective and Other Essays*. Cambridge: Cambridge University Press.

Strawson, P. (1985). *Skepticism and Naturalism. Some Varieties*. London: Methuen.

Travis, C. (1975). *Saying and Understanding*. Oxford: Blackwell.

Williams, M. (2004). Wittgenstein, truth and certainty. In M. Kölbel and B. Weiss, eds., *Wittgenstein’s Lasting Significance*. London-New York: Routledge, 249-284.

Wittgenstein, L. (1953). *Philosophical Investigations*. Oxford: Blackwell.

Wittgenstein, L. (1958). *The Blue and the Brown Books*. Oxford: Blackwell.

Wittgenstein, L. (1969). *On Certainty*. Oxford: Blackwell.

Wright, C. (1989). Wittgenstein’s rule-following considerations and the central project of theoretical linguistics. In A. George, ed., *Reflections on Chomsky*. Oxford: Blackwell. Reprinted in *Rails to Infinity*. Cambridge (MA): Harvard University Press, 2001, 170-213.

Wright, C. (2004). Warrant for nothing (and foundations for free?). *The Aristotelian Society Supplementary Volume*, 78, 167-212.

1. See in particular Bar-On (2004). I have discussed Bar-On’s position at length in Coliva (2016, chapter 6). [↑](#footnote-ref-1)
2. See in particular Bilgrami (2006), Shoemaker (1996), Wright (1989). I have discussed their positions at length in Coliva (2016, chapter 7). [↑](#footnote-ref-2)
3. I have presented, explained and defended these traits in Coliva (2016, chapter 3). Transparency amounts to the idea that, when one has the relevant mental states and concepts, one is *ipso facto* in a position to self-ascribe that mental state. Groundlessness amounts to the view that such a self-ascription is not underwritten either by observation of one’s own mental states, or by inference to the best explanation starting with the observation of one’s own behavior. More strongly, it can also entail the idea that the mental state does not serve as a justification for its self-ascription. Finally, authority amounts to the idea that, when the relevant conditions obtain, a subject’s psychological self-ascription cannot be challenged. [↑](#footnote-ref-3)
4. See Coliva (2016, chapters 7, 8). [↑](#footnote-ref-4)
5. This is the position I myself have developed in Coliva (2016). [↑](#footnote-ref-5)
6. Arguably that is what Russell (1912, chapter 5) proposed. See also Chalmers (2003). [↑](#footnote-ref-6)
7. Two brief remarks. First, if, like Chalmers (2003), one contented oneself with short-lived demonstrative concepts, the ostensive procedure could go through, but it would actually require much more than what proponents of it are usually prepared to recognize. For it would require *attending* to one’s sensation *as such*, as opposed to merely having it. Furthermore, it would require the ability to demonstratively refer to it in thought and to be disposed to use that demonstrative concept in a potentially infinite number of occasions. Given the short-lived nature of that concept, further occurrences in thought would not be genuinely demonstrative ones. Rather they would be like definite descriptions in disguise. Alternatively, they would have to rely on memory. In this case, Wittgenstein’s criticism that anything it would seem right to a subject would be right would apply. These aspects make the demonstrative strategy much more complex and at the same time less intuitive and appealing than it first seems. Furthermore, it then becomes quite clear that it would depend on having quite complex abilities, even conceptual ones, which could hardly be developed in a purely private setting. For these reasons, far from what Chalmers claims in his paper, it is dubious that his strategy can really avoid the kind of criticisms Wittgenstein developed against the idea of a private language and of private ostensive definitions.

Second, if one held that concepts are words of *mentalese*, along the lines of Fodor (1998), brought about by the causal interaction with the relevant properties, we would have no clear explanation of the conceivability of “inverted-spectrum” scenarios. For, if the concept elicited by the causal interaction with green objects looking red is red, after all, inverted-spectrum subjects would have a different concept than normal ones, who would conceptualize the very same items as green. Yet, the very point of inverted-spectrum scenarios is that despite different perceptual appearances, subjects would still be able to categorize objects in exactly the same way. If, in contrast, the concept elicited by the causal interaction with green objects looking red is green, there is no causally respectable story, which could explain the asymmetry. To account for it, one should be prepared to say that inverted-spectrum subjects have an abnormal conceptual faculty, rather than simply an abnormal perceptual mechanism, such that different phenomenal properties would give rise to the same concepts. [↑](#footnote-ref-7)
8. Notice that “ouch!” is a thoroughly linguistic item. In Italian it translates with “ahi!”. The usual sequence in Italian is “ahi!”, “Bua”, “Male”, “Ho/fa male”. “Bua” is considered child-speech, “ahi!” the universal exclamation of pain, “male” the ungrammatical but still acceptable expression of pain, if it is used by a child. Finally, “Ho/fa male” is the fully grammatical sentence to be used to express one’s pain. That very sentence, however, can be used to describe one’s own standing situation. For instance, when one goes to see a doctor and is asked to describe one’s symptoms. [↑](#footnote-ref-8)
9. *Pace* Jacobsen (1996). [↑](#footnote-ref-9)
10. *Contra* Jacobsen (1996). [↑](#footnote-ref-10)
11. At least in a large amount of cases, not always, though. [↑](#footnote-ref-11)
12. By strong contextualism, I have in mind what is nowadays called “semantic eliminativism”, a view maintained by Wittgenstein and later by Travis (1975). I do not mean anything having to do with forms of assessment sensitivity. Nor is the label “contextualism”, as used here, indebted to Kaplan’s notion of context and to Grice’s distinction between semantics and pragmatics. Crudely, for Wittgenstein the notion of context roughly coincides with that of language-game, and is certainly not exhausted by the triad subject-place-time or any slight extension of it. Moreover, given his equation of meaning and use, there is no room for a sharp division between semantics and pragmatics. [↑](#footnote-ref-12)
13. For an extended examination, see Coliva (2010, chapter 2). There I also draw the connection with Wittgenstein’s earlier treatment of the use of “I know” in connection with one’s own mental states, to point out the continuity of his thought. [↑](#footnote-ref-13)
14. The correct translation is “incorrectly” rather than “without right”. [↑](#footnote-ref-14)
15. See, in particular, Wright (2004) and Williams (2004). [↑](#footnote-ref-15)
16. The argument would simply be a *non sequitur*. Consider the analogy with colors. Clearly, there are cases of vague predication, where one subject would say “This is red” and the other “This is orange”, or where the same subject might oscillate between the two. That, however, would not mean that she could not know, for focal cases of redness, that a given object is instantiating that property. [↑](#footnote-ref-16)
17. The notion of transparency involved here is not equivalent to the one described in fn. 3. Rather, it is the one subsequently built upon by Evans (1982), Moran (2001) and several other theorists. I discuss this issue in Coliva (2016, chapter 5). [↑](#footnote-ref-17)
18. This is a remarkable, even though often under-appreciated point, of Wittgenstein’s treatment of Moore’s paradox. I have built on it in Coliva (2015), reprinted in Coliva (2016, Appendix). [↑](#footnote-ref-18)
19. Cf. ibid. [↑](#footnote-ref-19)
20. In Coliva (2016) I distinguish between first- and third-personal self-knowledge to make it as clear as possible that psychological self-ascriptions, though prima facie identical, may serve different purposes and may actually be arrived at in totally different ways. [↑](#footnote-ref-20)
21. See Strawson (1985, chapter 1). [↑](#footnote-ref-21)