

Forthcoming in:  
Wittgenstein in America  
Edited by Timothy McCarthy  
& Sean Stidd. Oxford, 2001

## Wittgenstein's "Plan for the Treatment of Psychological Concepts"

David Finkelstein

---

Plan for the treatment of psychological concepts.

Psychological concepts characterized by the fact that the third person of the present is to be verified by observation, the first person not.

Sentences in the third person of the present: information. In the first person present: expression. ((Not quite right.))

The first person of the present akin to an expression. (Z §472)

A striking feature of Wittgenstein's later writings is his preoccupation with psychological self-ascriptions—that is, with statements such as "I am afraid," "I am expecting an explosion," and "I'm in pain." This preoccupation is evident in the remarkable passage from *Zettel* reproduced above.<sup>1</sup> In what follows, I'll be offering a reading of this passage. My primary aim will be to clarify what Wittgenstein means when he characterizes a psychological sentence in the first person of the present as "akin to an expression." Along the way, I hope to shed light on what a number of his late writings have to say about, first, the relation between an expression and that which it expresses and, second, the authority with which we speak about our own mental goings-on.

### I. DETECTIVISM

Let's begin by considering what it is that Wittgenstein rejects in his "Plan for the treatment of psychological concepts." He says that psychological sentences in the first person of the present are not verified by observation. I'm going to call the sort of view that Wittgenstein here and elsewhere opposes *detectivism*. More precisely, I'll use the term

<sup>1</sup> This passage also appears at RPP 2 §63.

“detectivist” to refer to anyone who thinks that a subject’s ability to say what she is thinking or feeling is the result of her somehow finding out—whether by observation alone, or in conjunction with inference and memory. A detectivist thinks that our ordinary consciousness of, at least, some significant range of mental states or events is explained by the fact that we are able to detect their presence.

Exactly how we should understand the process by which a person detects the presence of her own mental goings-on is something about which detectivists have held, and continue to hold, a variety of views. Many have been drawn to the idea that we inwardly observe mental items via some kind of “inner sense” or “inner eye.” Some, like Bertrand Russell in *The Problems of Philosophy*, have held that the inner sense provides us with direct, infallible knowledge of items that are, in principle, private.<sup>2</sup> Others, many contemporary detectivists among them, have claimed that we know our own states of mind thanks to a more prosaic perceptual mechanism—one understood to be on all fours with the mechanisms that enable us to see and hear, only directed toward states and events that are literally inside our heads.<sup>3</sup>

Not all detectivists posit any sort of *inward* observation. One rather unusual version of detectivism with which Wittgenstein was familiar was put forward by Russell in 1911. In *The Analysis of Mind*,<sup>4</sup> Russell provides an account both of what desires are and of how we come to know our own. According to this account, a desire is any mental occurrence that involves “discomfort”—where what it means for a mental occurrence to involve discomfort is that it causes its subject to engage in “movements tending to produce some more or less definite change involving the cessation of the occurrence.”<sup>5</sup> Russell writes:

<sup>2</sup> Bertrand Russell, *The Problems of Philosophy* (Oxford: Oxford University Press, 1912), pp. 49–51.

<sup>3</sup> Paul Churchland and Nicholas Humphrey (a neuropsychologist) provide recent statements of this sort of detectivism:

[S]elf-consciousness . . . is just a species of perception: *self-perception*. It is not perception of one’s foot with one’s eyes, for example, but is rather the perception of one’s internal states with what we may call (largely in ignorance) one’s faculty of introspection. Self-consciousness is thus no more (and no less) mysterious than perception generally. It is just directed internally rather than externally. (P. Churchland, *Matter and Consciousness* (Cambridge, Mass.: MIT Press, 1988), p. 74.)

It is as if I, like every other human being, possess a kind of ‘inner eye’, which looks in on my brain and tells me why and how I’m acting in the way I am—providing me with what amounts to a plain man’s guide to my own mind. (N. Humphrey, *The Inner Eye* (London: Faber & Faber, 1986), p. 87.)

<sup>4</sup> Bertrand Russell, *The Analysis of Mind* (New York: Humanities Press, 1911).

<sup>5</sup> *Ibid.* p. 71.

*Plan for the Treatment of Psychological Concepts* 217

The cycle [of restless movements caused by a mental occurrence involving discomfort] ends in a condition of quiescence . . . The state of affairs in which this condition of quiescence is achieved is called the "purpose" of the cycle, and the initial mental occurrence involving discomfort is called a "desire" for the state of affairs that brings quiescence.<sup>6</sup>

When I'm hungry, I'm subject to a mental occurrence that causes me to engage in restless movements likely to lead to my eating. That this occurrence is a desire to eat consists in the fact that upon eating, my restless activity comes to an end.

Given this analysis of desire, how is it that I'm sometimes able to say what it is that I desire? According to Russell, "Conscious' desire . . . consists of desire in the sense hitherto discussed, together with a true belief as to its 'purpose,' i.e. as to the state of affairs that will bring quiescence with cessation of the discomfort."<sup>7</sup> How, then, do I come to have a true belief concerning what will bring quiescence? Russell writes that "the discovery of our own motives can only be made by the same process by which we discover other people's, namely, the process of observing our actions and inferring the desire which could prompt them."<sup>8</sup> Thus, according to the sort of detectivism that Russell defends in 1921, when I say what it is that I desire, I issue a report that is based on (1) observation of my own restless behavior, and (2) an inference as to what is likely to bring the behavior to an end.

\*

In *Philosophical Remarks*, we find Wittgenstein criticizing Russell's 1921 analysis of desire:

I believe Russell's theory amounts to the following: if I give someone an order and I am happy with what he then does, then he has carried out my order.

(If I wanted to eat an apple, and someone punched me in the stomach, taking away my appetite, then it was this punch that I originally wanted.) (*PR*, III §22)

Later, in the *Investigations*, Wittgenstein is still preoccupied with the account of desire and self-knowledge that Russell defends in 1921:

Saying "I should like an apple" does not mean: I believe an apple will quell my feeling of nonsatisfaction. *This* proposition is not an expression of a wish but of nonsatisfaction. (*PI* §440)

It's not just *The Analysis of Mind's* funny sort of detectivism that Wittgenstein is concerned to criticize. His opposition to every sort of

<sup>6</sup> *Ibid.* p. 75.

<sup>8</sup> *Ibid.* p. 31.

<sup>7</sup> *Ibid.* p. 72.

detectivism emerges as a theme in his later writings, a theme that is sounded in, for example, the follow passages:

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)

When someone says "I hope he'll come"—is this a *report* about his state of mind . . .?—I can, for example, say it to myself. And surely I am not giving myself a report. (*PI* §585)

But that which is in him, how can I see it? Between his experience and me there is always the expression!

Here is the picture: He sees it immediately, I only mediately. But that's not the way it is. He doesn't see something and describe it to us. (*LaWrPP2*, p. 92)

How does Wittgenstein think we should understand psychological self-ascriptions if not as observation reports? In the next section, I shall discuss an answer to this question that I take to be unsatisfactory—both as a reading of Wittgenstein and as an account of how we should think about psychological self-ascriptions—but which, nonetheless, can easily seem to be the only alternative to detectivism.

## 2. CONSTITUTIVISM

The position I have in mind might be stated as follows: "Our mental state self-ascriptions are unlike observation reports in that they constitute the very facts to which they refer. There is no need for me to engage in anything like self-observation in order to state, for example, what I intend to do this evening because when I say that I intend to go to bed early, I *make it the case* that this is what I intend. My authority concerning my own states of mind is not epistemic; it is, rather, like that of an army colonel when he declares an area to be off-limits. The colonel needn't be better informed about the area than his soldiers are. His authority consists in the fact that what he says goes. So it is with my authority about my own states of mind. Typically, what I say goes." I shall call a philosopher who holds a view of this sort a *constitutivist*.

It might seem as if we must choose between some form of detectivism and some form of constitutivism (or else, some hybrid position) when we try to make sense of our psychological self-ascriptions. ("Either I discover my state of mind or I, as it were, make it.") If one thinks that

*Plan for the Treatment of Psychological Concepts* 219

detectivism and constitutivism together exhaust the available views of first-person authority, then it will be natural to read Wittgenstein—who often speaks against detectivism—as putting forward some version of constitutivism. This is, as it were, a negative reason to read Wittgenstein as a constitutivist: he's *not* a detectivist, so he must be a constitutivist. But there are also positive reasons to read Wittgenstein as a constitutivist. There are passages in his writings that can look like positive endorsements of some form of constitutivism.

Someone who wanted to read Wittgenstein as a constitutivist might appeal to the following passage from *Zettel* for support:

But how does the person in whom it goes on know which event the process is the expectation of? For he does not seem to be in uncertainty about it. It is not as if he observed a mental or other condition and formed a conjecture about its cause. He may well say: "I don't know whether it is only this expectation that makes me so uneasy today"; but he will not say: "I don't know whether this state of mind, in which I now am, is the expectation of an explosion or of something else."

The statement "I am expecting a bang at any moment" is an *expression* of expectation. This verbal reaction is the movement of the pointer, which shows the object of expectation. (*Z* §53)

A commentator who took Wittgenstein to be a constitutivist might put the point of this passage as follows: "An expectation is like a pointer in that it is, as it were, aimed at something—its object. Now, how is it that I am able to say at what my expectation is aimed? Many philosophers would claim that when I inwardly observe my expectation, I see the direction in which it points, but Wittgenstein rejects this view. What he is saying in the last sentence of the passage is that when I avow an expectation, I am not reporting on the direction in which my mental state points, but rather, setting or determining the position of the pointer. When I say I am expecting a bang, I make it the case that a bang is what I am expecting."

I think the most lucid proponent of a reading of Wittgenstein along these lines has been Crispin Wright. In the passage that follows, Wright states a view that he himself endorses and that he thinks Wittgenstein might have endorsed:

The authority which our self-ascriptions of meaning, intention, and decision assume is not based on any kind of cognitive advantage, expertise or achievement. Rather it is, as it were, a *concession*, unofficially granted to anyone whom one takes seriously as a rational subject. It is, so to speak, such a subject's right to declare what he intends, what he intended, and what satisfies his intentions; and his possession of this right consists in the conferral upon such

declarations, other things being equal, of a constitutive rather than a descriptive role.<sup>9</sup>

Wright's constitutivism falls out of his reading of Wittgenstein's remarks about rule-following. In the next few paragraphs, I'll sketch the connection between what Wright's Wittgenstein has to say about rule-following and what he has to say about psychological self-ascriptions.

Wright states what he takes to be a lesson of Wittgenstein's remarks about rule-following as follows:

It might be preferable, in describing our most basic rule-governed responses, to think of them not as informed by an *intuition* (of the requirements of the rule) but as a kind of *decision*.<sup>10</sup>

It is tempting to say that when someone sets out to follow a rule, she intuits, or perceives, its requirements; she sees what the rule calls for. According to Wright's Wittgenstein, to say this is to commit oneself to a problematic, platonistic conception of rules. On Wright's reading of Wittgenstein, when someone follows a rule, she doesn't perceive its requirements; she decides them. She, as it were, stipulates what the rule calls for.

There is an obvious problem with saying that when someone follows a rule, she decides or stipulates its requirements. This way of putting things suggests that when someone is confronted by a rule, she is free to decide that anything she feels like doing is what the rule calls for. Wright's answer to this problem is to say that it is only our best judgments (i.e., our best *decisions*) about a rule that determine what it requires. One may go wrong in trying to follow a rule because one may act in ways that don't conform to the best judgments about what it calls for.

By virtue of what is some judgment about a rule's requirements a *best* judgment? According to Wright, a best judgment is one that is arrived at under certain ideal conditions—under what he calls “C-conditions.” From here, Wright gets to constitutivism in two steps. The first step is to extend the sort of story he wants to tell about rules so that it applies as well to intentions, wishes, and the like. If it is platonistic to think that

<sup>9</sup> Crispin Wright, “On Making Up One's Mind: Wittgenstein on Intention,” in Weingartner and Schurz, eds., *Logic, Philosophy of Science and Epistemology* (Kirchberg: Hölder-Pichler-Tempsky, 1987), pp. 400–1.

<sup>10</sup> Crispin Wright, “Wittgenstein's Rule-following Considerations and the Central Project of Theoretical Linguistics,” in A. George, ed., *Reflections on Chomsky* (Oxford: Blackwell, 1989), p. 240; also Crispin Wright, critical notice of Colin McGinn, *Wittgenstein on Meaning, Mind* 98 (1989): 300.

*Plan for the Treatment of Psychological Concepts* 221

mere marks on a page can, of themselves, call for one behavior rather than another, then it is also platonistic to think that an intention can, of itself, determine what would fulfill it. According to Wright, the question of what it is that would satisfy an intention is, like the question of what a rule calls for, settled by a judgment made under ideal conditions. Thus, what we might call Wright's *stipulativism* is not restricted to rules or signs; it extends to mental states as well.

The second step on Wright's path from stipulativism about rules to constitutivism about mental content is to claim that under most *ordinary* conditions, when I make a judgment about my own intentional state, it is a best judgment. This is to say, for judgments or opinions that may be expressed in the form of avowals of intention, desire, expectation, and so on, what Wright calls the C-conditions are usually satisfied. Typically, according to Wright, when I say that I intend to  $\phi$ , I make it the case that  $\phi$  is what I intend to do.<sup>11</sup> Thus, he ends up with a constitutivist account of intentional state self-ascriptions:

[W]hy is it *a priori* reasonable to believe that, provided Jones has the relevant concepts and is attentive to the matter, he will believe that he intends to  $\phi$  if and only if he does? . . . [T]he matter will be nicely explained if the concept of intention works in such a way that Jones's opinions, formed under the restricted set of C-conditions, play [an] extension-determining role . . .<sup>12</sup>

Jones is generally right about his own intentions because, under ordinary conditions, his taking himself to have a particular intention fixes it that he does.

\*

I argue elsewhere that stipulativism neither captures the moral of Wittgenstein's remarks about rule-following nor turns out to be a coherent position in its own right.<sup>13</sup> Thus, I believe that Wright's constitutivism, which grows out of his stipulativism, has its roots in poor soil. I can't begin to argue this point here, however, nor would it really be to my present purpose. For even if you were to grant both that stipulativism is a hopeless strategy for explaining content *and* that Wittgenstein is no stipulativist, you might still be drawn to a constitutivist reading of Wittgenstein, one according to which—although he is

<sup>11</sup> Wright notes that this determination is defeasible. What a subject says about his own intentional states is generally allowed to stand, but subsequent events occasionally overturn his judgment.

<sup>12</sup> Wright, "Wittgenstein's Rule-following Considerations," p. 252.

<sup>13</sup> See my "Wittgenstein on Rules and Platonism," in A. Crary and R. Read, eds., *The New Wittgenstein* (London: Routledge, 2000).

not a stipulativist about content in general—he nonetheless holds that, ordinarily, the avowal of a mental state constitutes, rather than reflects, its subject matter. After all, constitutivism offers a way to explain the special authority that attaches to our mental state avowals without appealing to any sort of privileged epistemic access, and this does seem to be something that Wittgenstein is after.

But is this really what constitutivism offers? Notice that while constitutivism may have some *prima facie* plausibility when we think about intentions and expectations, it has none when we consider sensations. It would be worse than unsympathetic—it would be *crazy*—to think, for example, that toothache sufferers speak with authority about their pain because their pain is constituted by their avowals of it.<sup>14</sup>

Wright knows better than to attribute this crazy view about pain to Wittgenstein. He endorses constitutivism and reads Wittgenstein as a constitutivist, but *only* about intentional states. Wright contrasts intentional states with what he calls “genuine episodes and processes in consciousness,” characterizing the latter as “items which, like headaches, ringing in the ears, and the experience of a patch of blue, may have determinate onset and departure, and whose occurrence makes no demands upon the conceptual resources of the sufferer.”<sup>15</sup> He puts the point of contrast as follows:

[N]othing strictly introspectible has, in the case of any of these concepts [meaning, understanding, intending, expecting, wishing, fearing, hoping], the right kind of characteristics. We cannot, honestly, find anything to *be* the intention, etc., when we turn our gaze inward . . .<sup>16</sup>

According to Wright, we do sometimes “turn our gaze inward,” and when we do, we find only “episodes of consciousness.” Only these headache-like items are “strictly introspectible.” Wright avoids saddling Wittgenstein with the crazy view that headaches are constituted by avowals of them by reading him as a constitutivist about expectations and intentions, but not about pains.<sup>17</sup>

<sup>14</sup> I'm afraid that I do know of one reader of Wittgenstein—a professor in an English department—who not only attributes this view to Wittgenstein, but endorses it himself. (At least he's not a dentist.)

<sup>15</sup> Wright, “Wittgenstein's Rule-following Considerations,” p. 237.

<sup>16</sup> *Ibid.*

<sup>17</sup> One problem with this strategy is that, *prima facie*, constitutivism doesn't seem much more plausible (or sympathetic) as a story about *fear* than as a story about pain. And, Wright does suggest that Wittgenstein is a constitutivist about fear as well as expectation and intention (see “Wittgenstein's Rule-following Considerations,” p. 237). (Wright might try to analyze fear into an intentional component and a sensational component. I don't think any such strategy would work, but I won't pursue the point here.)



*Plan for the Treatment of Psychological Concepts* 223

According to the position that Wright thinks Wittgenstein might have held, a typical self-ascription of expectation has a completely different grammar from a self-ascription of pain; the former constitutes its subject matter while the latter does not. But Wittgenstein's writings tell a different story—one according to which such self-ascriptions *share* something that we must recognize if we are to rid ourselves of philosophical confusion about them. Earlier, I cited a passage from *Zettel* §53 that, I said, might seem to support a constitutivist reading of Wittgenstein. Let's look at part of it again:

The statement "I am expecting a bang at any moment" is an *expression* of expectation. This verbal reaction is the movement of the pointer, which shows the object of expectation.

Compare that remark about expectation with the following passage about pain:

[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-behaviour.

"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)<sup>18</sup>

Throughout his late writings, Wittgenstein urges us to view—not only self-ascriptions of expectation, intention, and the like, but also—self-ascriptions of pain and other sensations as, or as akin to, expressions.

If Wittgenstein held the sort of view that Wright sets out, his "Plan for the treatment of psychological concepts" ought to have recommended that we think about self-ascriptions of intention and self-ascriptions of sensation in entirely different ways. But what the "Plan" says is that *psychological* self-ascriptions—that is, sentences in the first person of the present that involve psychological concepts—are "akin to expressions" (and it is clear from what immediately follows this remark that Wittgenstein means to include sensation concepts among those which can be called psychological). Of course, we might read Wittgenstein as a more thoroughgoing constitutivist than Wright suggests we should; we might take him to be a constitutivist about sensations as well as intentional states. But this would be to attribute a crazy view of sensations to him. In what follows, I'll try to provide a reading of Wittgenstein that attributes no crazy views to him but still allows us

<sup>18</sup> See also *LW*7PP2, p. 92 and *Z* §§484–7.

to make sense of his wanting to compare both avowals of expectation and avowals of sensation to expressions.

According to the constitutivist reading of *Zettel* §53, “the movement of the pointer” – an avowal of expectation—fixes or determines what the expectation is *of*. But the final sentence of *Zettel* §53 does *not* read, “This verbal reaction is the movement of the pointer, which *determines* the object of expectation.” The sort of pointer that Wittgenstein has in mind is one that *shows* something—makes something manifest. In *Zettel* §53, Wittgenstein is trying to draw attention to the way in which a psychological self-ascription, like a facial expression, can make someone’s mental condition manifest. What *sort* of making manifest is this?

### 3. THE CEO AND THE CONTEXT PRINCIPLE

Let’s compare two ways in which something may be made manifest. Imagine that Joan rolls up her sleeves, whereupon you see that one of her wrists is badly scarred. Such a scar might be the result of an accident, but given what you know about Joan’s history, you infer that she has tried to kill herself. We might say that in rolling up her sleeves, Joan makes her scar manifest. We might also say that she makes it manifest to you that she has attempted suicide. Now, when someone’s face lights up in a joyful expression, in which of these ways does she make her joy manifest? Is joy, so expressed, analogous to Joan’s scar, or to her suicide attempt?

Neither. On the one hand, we are able to understand one another’s facial expressions without needing to make inferences. Often, the only way I am able to describe someone’s facial expression is as joyous or miserable; I don’t typically see a person’s smile or wince as a set of psychologically neutral movements and infer from these that she is in this or that mental condition:

“We see emotion.”—As opposed to what?—We do not see facial contortions and make inferences from them (like a doctor framing a diagnosis) to joy, grief, boredom. We describe a face immediately as sad, radiant, bored, even when we are unable to give any other description to the features.—Grief, one would like to say, is personified in the face. (*Z* §125; see also *PI* §537 and *Z* §218)

On the other hand, *seeing* that someone is happy or in pain is not quite like seeing that Joan’s wrist is scarred. In *Philosophical Remarks*, Wittgenstein registers his dissatisfaction with a view by describing it as follows:

*Plan for the Treatment of Psychological Concepts* 225

Pain is represented as something we can perceive in the sense in which we perceive a matchbox. (PR, p. 94)

Pain isn't the sort of thing we can perceive in the way a matchbox or a scar is perceived. Our thoughts and feelings are not hidden inside us in the way that a matchbox may be hidden in a handkerchief, or a scar inside a shirt-sleeve. If I express my pain by crying out, or my anger by scowling, this isn't like rolling up a sleeve and revealing a scar.<sup>19</sup>

Why not? What's the difference? I want to begin to characterize the difference by turning to a case in which someone speaks with a kind of first-person authority, not about her state of mind, but about the meaning of a word. Imagine a board meeting at which the chief executive officer has this to say about an employee: "Phillips is a real team-player, by which I don't mean that he's a stupid sheep, but rather that he won't help himself at the expense of this company." Now consider the question: How did the CEO find out what she meant by "team-player"? Did she listen to the first half of her sentence and interpret it? If so, why do the other board members attach such weight to *her* interpretation of the remark? Such questions are confused. The confusion may be characterized as follows: the second half of the CEO's sentence is not a *report* on the first half, but rather, an *expression* of its meaning. But now, what does *this* mean?

Part of what it means is that the members of the board hear the whole sentence and understand it as a coherent unit; the two halves of the sentence make sense in light of each other. The second half of the CEO's sentence is as much a part of her assessment of Phillips as the first. To understand what the whole sentence means—to hear what the CEO is saying about Phillips—a listener needs to take in the second half of the sentence together with the first. We can call the second half of the sentence a gloss on, or interpretation of, the first, but it isn't *merely* an interpretation; it's an elaboration, a fleshing out. It makes the CEO's meaning manifest in a way that a listener's interpretation of her remark would not. The weight that the board members accord the second half of her sentence does not reflect the CEO's superior ability to detect

<sup>19</sup> I take it that part of the point of the following passage is that in the use of the word "pain" there is no such thing as *exhibiting what one has got* if "exhibiting" is understood to have the grammar that it has when we say that someone exhibited her scar:

Do not say "one cannot", but say instead: "it doesn't exist in this game". Not: "one can't castle in draughts" but—"there is not castling in draughts"; and instead of "I can't exhibit my sensation"—"in the use of the word 'sensation', there is no such thing as exhibiting what one has got". (Z §134)

features of her own utterances or mental states. It reflects the fact that a sentence is what we might call a unit of intelligibility; the word “team-player” and the CEO’s gloss on it are two parts of a single, coherent, intelligible whole.

Consider a variation on our story: At the board meeting, the CEO says, “Phillips is a real team-player,” without elaborating on what she means by this. Five years later, her secretary is reading through the minutes of the meeting. He asks her what she meant by “team-player.” The CEO remembers neither the meeting nor Phillips. She says, “Well, that’s not a word that I often use. I suppose I might have meant it ironically—as a kind of insult to this Phillips.” The CEO interprets her own remark as if it were made by someone else; she takes up a third-person perspective on it. Her two remarks, separated by years and forgetting, don’t constitute a single unit of intelligibility. The second remark might provide information (or misinformation) about the first, but what we have in this version of the story are two separate remarks, not two parts of something that make sense together. In this version of the story, the CEO’s second remark doesn’t come with the sort of authority that we saw in the first version. Here, she is not expressing her meaning; she is merely interpreting something that she once said.

There is an analogy between the way the CEO’s expression of what she meant by “team-player” is understood to fit coherently into the context of a sentence and the way an expression of, say, anger is understood to fit coherently into a person’s behavioral and psychological life. In what remains of this essay, I’ll argue that according to Wittgenstein, the way I make my anger manifest when I express it, whether by scowling or by announcing that I’m angry, is akin to the way in which the second half of the CEO’s sentence makes the meaning of its first half manifest. Like many of Wittgenstein’s views, this one owes something to Frege—specifically, to Frege’s context principle.

\*

In the introduction to *The Foundations of Arithmetic*, Frege writes, “In the enquiry that follows, I have kept to three fundamental principles.” The second of these principles is “never to ask for the meaning of a word in isolation, but only in the context of a proposition.”<sup>20</sup> Later in the book, he writes, “It is enough if the proposition taken as a whole has a sense; it is this that confers on the parts also their content.”<sup>21</sup>

<sup>20</sup> Gottlob Frege, *The Foundations of Arithmetic* (Evanston, Ill.: Northwestern University Press, 1953), p. x.

<sup>21</sup> *Ibid.* p. 71.

*Plan for the Treatment of Psychological Concepts* 227

Frege's context principle constitutes a rejection of the view that understanding a sentence requires that one grasp the meanings of independently intelligible sentence-parts. For Frege, the sentence is the primary unit of intelligibility. A part of a sentence—a word, for example—has the meaning that it does only in the context of a sentence.

An example of Frege's helps to make the point. Compare the following sentences:

- (1) Vienna is the capital of Austria.
- (2) Trieste is no Vienna.

The logical role of the word "Vienna" is different in these sentences. In the first sentence, "Vienna" functions as proper name. In the second, it functions, Frege says, as "a concept-word, like 'metropolis.'"<sup>22</sup> It would make sense to say, "Although Trieste is no Vienna, Paris is a Vienna—the only one in France." Imagine that someone who said this also had occasion to say, "Vienna is the capital of Austria." She would not thereby commit herself to the view that the capital of Austria is in France. The word "Vienna" does not mean the same thing in sentences (1) and (2). We come to see what a particular use of a word means only when we consider it in the context of a whole sentence.

I want to suggest that what goes for the word "Vienna" also goes for the word "team-player" in the CEO example. The word "team-player" means what it does only in the context of the whole in which it appears. But in the sentence uttered by the CEO—"Phillips is a real team-player, by which I don't mean that he's a stupid sheep, but rather that he won't help himself at the expense of this company"—we see a special kind of sentential context, one that constitutes a gloss on, or interpretation of, that which it contextualizes. In the sentence "Trieste is no Vienna," the words "Trieste is no" contextualize the word "Vienna," but they do not constitute an interpretation of it. We could say that the CEO speaks with a special authority concerning what she meant by the word "team-player" because she is offering an interpretation of it that is not a *mere* interpretation. Her interpretation contextualizes that which it interprets in a way that an interpretation offered by one of the other board members would not.

\*

One can hear echoes of Frege's context principle in both the *Tractatus* and the *Investigations*. At *Tractatus* 3.3, Wittgenstein says: "[O]nly in

<sup>22</sup> Gottlob Frege, *Translations from the Philosophical Writings of Gottlob Frege* (Totowa, N.J.: Rowman & Littlefield, 1952), p. 50.

the context of a sentence has a name meaning.” And in §49 of the *Investigations*, he writes:

We may say: *nothing* has so far been done, when a thing has been named. It has not even *got* a name except in a language-game. This was what Frege meant too, when he said that a word had a meaning only as part of a sentence.

Wittgenstein doesn't merely inherit Frege's context principle; he reshapes it in a number of ways—two of which I'll call to your attention. One of them is signaled by Wittgenstein's use of the term “language-game” in the passage from the *Investigations* just cited. He speaks of language-games where Frege spoke of sentences. However it is that one ought precisely to characterize what Wittgenstein means by “language-game,” it is clear that language-games are, as it were, *wider* than sentences. Frege's point—that our words mean what they do only in their contexts—is still present in the *Investigations*, but Wittgenstein has wider contexts in mind. Consider the following passages:

“After he had said this, he left her as he did the day before.”—Do I understand this sentence? Do I understand it just as I should if I heard it in the course of a narrative? If it were set down in isolation I should say, I don't know what it's about. (*PI* §525)

The phrase “description of a state of mind” characterizes a certain *game*. And if I just hear the words “I am afraid” I might be able to *guess* which game is being played here (say on the basis of the tone), but I won't really know it until I am aware of the context. (*LaWrPP1* §50)

About the first of these passages, I want to say that there is a sense in which I *do* understand the sentence mentioned, but—and this is Wittgenstein's point—there is another sense in which I don't understand it at all. I have no idea whom the sentence is about or what condition “he” left “her” in the day before. The second passage makes a similar point. What someone is doing when he utters the words “I am afraid” depends on a context that is wider than a sentence. Wittgenstein wants to show us that the functions and meanings that our words have depend on the ways in which they are situated, not just—as Frege says—in sentences, but in conversations and stories; in stretches of discourse, thought, and behavior; in spans of human life. We could say that Wittgenstein reshapes Frege's context principle by enlarging the contexts to which it, or some descendent of it, applies.<sup>23</sup>

<sup>23</sup> Here, we might consider another variation on the CEO example: Imagine that the CEO says, “Phillips is a real team-player,” without glossing what she means by this. Later in the meeting, one of the board members says to the CEO, “A few minutes ago,

*Plan for the Treatment of Psychological Concepts* 229

So: one way in which Wittgenstein reshapes and extends Frege's context principle is by calling attention to contexts that are wider than sentences.<sup>24</sup> A second way he does this is by bringing out that it is not only linguistic items—words or sentences—that depend for their significance on their surroundings. The following passage concerns the significance of a wordless activity—the placing of a crown on a someone's head:

A coronation is the picture of pomp and dignity. Cut one minute of this proceeding out of its surroundings: the crown is being placed on the head of the king in his coronation robes.—But in different surroundings gold is the cheapest of metals, its gleam is thought vulgar. There the fabric of the robe is cheap to produce. A crown is a parody of a respectable hat. And so on. (*PI* §583)

In thinking about such examples, it helps to imagine glimpsing a few seconds out of a movie. Suppose that while channel-surfing, you catch a moment of a film in which a man is slowly closing a door. There is a sense in which you know what you are seeing—a man closing a door—but another sense in which you don't. You don't know the meaning of this activity—this door-closing—because you don't know how it figures in the story. Perhaps if you saw more of the film, you would want to say that in closing the door, the man was breaking off a love affair, or trying to stay hidden, or insulting the butler. The significance of our

you remarked that Phillips is a real team-player. I wasn't sure just how you meant that. Were you implying that he's not particularly creative or original?" The CEO replies, "No; I meant that he's good at getting along with people." In this example, the CEO's initial remark about Phillips and her gloss on it are not part of a single sentence. They are, however, part of what we can think of as a somewhat broader and looser unit of intelligibility—the CEO's remarks about Phillips at the board meeting. (The CEO speaks with authority here too, but perhaps not quite the same degree of authority as in the original example. The contextual relation is not so tight here as in the first version of the story.)

<sup>24</sup> I would argue that Wittgenstein's remarks about rule-following should be viewed in light of this point about how he appropriates Frege's context principle. The difference between the way Crispin Wright reads the remarks on rule-following and the way I read them might be summarized as follows. For Wright a central lesson of Wittgenstein's remarks about rule-following is that since no amount of interpretation can breathe meaning into the noises and ink-marks that constitute our words, they must depend for their meaning on something else—something like stipulation. On my reading, Wittgenstein's point would be better put by saying that the question "What gives meaning to the dead ink-marks and noises we produce?" only arises at all because, in doing philosophy, we're moved to consider words apart from the contexts in which they have their significance (see my "Wittgenstein on Rules and Platonism"). Thus, there's a sense in which Wittgenstein's discussions of rule-following can be viewed—indeed, I think should be viewed—as making a point similar to the one that Frege makes when he warns us not to ask after the meaning of a word apart from its context. The difference is that whereas Frege stresses a word's sentential context, Wittgenstein wants to call attention to the broader context of human activities in which words have their uses and their significance.

activities, both with and without words, depends on how they are situated in our lives:

Only surrounded by certain normal manifestations of life, is there such a thing as an expression of pain. Only surrounded by an even more far-reaching particular manifestation of life, such a thing as the expression of sorrow or affection. And so on. (Z §534)

The point here is the same as in the coronation passage—only here, the activity at issue is the expressing of a state of mind rather than the placing of a crown on a head. An affectionate glance or touch would not *be* an expression of affection, or of anything else, were it not for the “particular manifestation of life” that surrounds it.

It is not only our words and activities whose significance depends on the “manifestations of life” that surround them. According to Wittgenstein, just as an expression of love depends for its significance on the life in which it is situated, so too does the loving state of mind that it expresses. In the section of the *Investigations* that immediately precedes the coronation passage, Wittgenstein writes:

Could someone have a feeling of ardent love or hope for the space of one second—*no matter what* preceded or followed this second?—What is happening now has significance—in these surroundings. The surroundings give it its importance. (PI §583)

Frege says “never to ask for the meaning of a word in isolation, but only in the context of a proposition.” Wittgenstein might have said never to ask after someone’s mental condition in isolation, but only in the context of the events in his life:

“Grief” describes a pattern which recurs, with different variations, in the weave of our life. If a man’s bodily expression of sorrow and of joy alternated, say with the ticking of a clock, here we should not have the characteristic formation of the pattern of sorrow or of the pattern of joy.

“For a second he felt violent pain.”—Why does it sound queer to say: “For a second he felt deep grief”? Only because it so seldom happens?

But don’t you feel grief *now*? (“But are you playing chess *now*?”) (PI, p. 174)

A person *can* feel grief at a particular moment. But that it is *grief* he feels has to do with what comes before and after that moment. This is not to say that someone’s grieving at time  $t_2$  *makes it the case* that he was grieving at  $t_1$  (any more than it is to say that what a person does at  $t_1$  makes it the case that he’s grieving at  $t_2$ ). But we *can* say that at each moment, a person’s psychological condition makes sense in light of feelings, behavior, and events that precede and succeed it.



4. WITTGENSTEIN ON PSYCHOLOGICAL  
SELF-ASCRPTIONS

Earlier, I claimed that the CEO speaks with authority about what she means by the word "team-player" because her gloss on this word isn't a mere interpretation. When she says, "Phillips is a real team-player, by which I don't mean that he's a stupid sheep, but rather that he won't help himself at the expense of this company," she provides both an interpretation of "team-player" and a sentential context for it. I just now claimed that, according to Wittgenstein, a person's psychological condition at a particular moment depends on the surrounding events in her life in something like the way that, according to Frege, a word's meaning depends on its sentential context.

We are now in a position to state a significant point about the character of our mental state self-ascriptions. Just as, in the first version of the CEO story, the CEO's self-interpretation contextualizes that which it interprets and so isn't a mere interpretation, our psychological self-ascriptions contextualize that which they ascribe and so aren't mere ascriptions. Typically, when someone ascribes, for example, an expectation to himself, the ascription is an essential part of the "weave of life" in which his expectation participates and from which it draws its sense. This is part of what Wittgenstein is calling to our attention when he stresses the expressive character of psychological self-ascriptions. An avowal of expectation bears a relation to a person's psychological condition that is akin to the relation the CEO's self-interpretation bears to her use of "team-player."

Consider an example. Wittgenstein writes:

An expectation is embedded in a situation from which it takes its rise. The expectation of an explosion for example, may arise from a situation in which an explosion is *to be expected*. The man who expects it had heard two people whispering: "Tomorrow at ten o'clock the fuse will be lit". Then he thinks: perhaps someone means to blow up a house here. Towards ten o'clock he becomes uneasy, jumps at every sound, and at last answers the question why he is so tense: "I'm expecting . . ." This answer will e.g. make his behaviour intelligible. It will enable us to fill out the picture of his thoughts and feelings. (Z §67)

On the one hand, the man's expectation isn't suddenly constituted when he expresses it in words. It is, rather, "embedded in a situation from which it takes its rise," a situation in which the man has good reason to expect an explosion and in which his expectation of an explosion is expressed in a variety of ways. On the other hand, when the man described in the passage says, "I'm expecting an explosion," he isn't—

or, anyway, he needn't be—reporting a fact that he has observed. Rather, his avowal of expectation is, like his jumping at every sound, a piece of expectant behavior, an act of expecting. His avowal requires no more inner observation than does his jumpiness. Like his jumpiness and his overhearing talk about a fuse, his psychological self-ascription is an integral part of the situation in which his expectation is embedded and from which it takes its rise. It is because of *this* that it carries a kind of authority that another person's ascription of an expectation to him would not. His authority when he says what he expects derives from the fact that his avowal of expectation helps to contextualize the very thing that it is an avowal of. What's at issue isn't mere ascription; it is something more, something that could be called his expressing his expectation in words.

##### 5. "NOT QUITE RIGHT": THE ASSERTORIC DIMENSION OF PSYCHOLOGICAL SELF-ASCRPTIONS

We saw that in his "Plan for the treatment of psychological concepts," Wittgenstein writes:

Sentences in the third person of the present: information. In the first person present: expression. ((Not quite right.))

The first person of the present akin to an expression. (Z §472)

The "Not quite right" is puzzling. In other passages, Wittgenstein simply asserts that one or another psychological sentence in the first person *is* an expression. (We saw that in *Zettel* §53, he writes, "The statement 'I am expecting a bang at any moment' is an *expression* of expectation.") Why, then, in *Zettel* §472 does Wittgenstein indicate that it isn't "quite right" to say that psychological sentences in the first person of the present are expressions?

Although I have been drawing attention to what might be called the expressive dimension of psychological self-ascriptions—to the way in which an avowal of, say, desire is like a desirous facial expression—we should not overlook an important difference between an avowal of desire on the one hand, and a desirous look on the other. When I avow that I want a bite of your ice cream, I express my desire for a bite, and I *say* that I want a bite. I both show you and tell you what I want. When I stare longingly at your spoon as it moves from the dish to your lips, I express my desire, but I don't thereby assert that I want a bite.

Unlike a bodily expression, a psychological self-ascription asserts the

*Plan for the Treatment of Psychological Concepts* 233

existence of the very state of affairs that it makes manifest. It is for this reason that we are inclined to speak of *authority* in connection with psychological self-ascriptions, but not in connection with smiles or desirous looks. If you want to know my psychological condition, you ought to attend both to what I say about myself and to my bodily expressions of emotion, intention, and so on. We speak of authority, however, only in connection with the former, because when I say what's on my mind I make a *statement* that has a special claim to truth.

It is easy to get confused here—easy to think that if an utterance of “I'm so happy” is an authoritative assertion or statement, then it must carry the sort of authority that attaches to other kinds of statement. This thought will lead one to assimilate first-person authority either to the authority of an eye-witness or to the authority of the army colonel who declares an area off-limits—either, that is, in the direction of detectivism or in the direction of constitutivism. It might be less confusing if we didn't call expressive avowals “assertions” or “statements” at all. Wittgenstein notes:

To call the expression of a sensation a *statement* is misleading because ‘testing’, ‘justification’, ‘confirmation’, ‘reinforcement’ of the statement are connected with the word “statement” in the language-game. (Z §549)

Misleading as it is, however, there is an important respect in which such avowals are like run-of-the-mill statements: they are truth-evaluable. Unlike smiles and winces, they have an assertoric dimension. They are, we might say, assertions of a special sort.

It is very often assumed that when Wittgenstein characterizes psychological self-ascriptions as expressions, he means to be denying that they are assertions.<sup>25</sup> On such a reading, what Wittgenstein has to say about psychological self-ascriptions can seem fairly easy to dismiss. Consider, for example, the following from a paper by David Rosenthal:

In *Philosophical Investigations* Wittgenstein (1953) seems to have held, roughly, that although one can report that some other person is, for example, in pain, in one's case one can only express the pain, and not report it as well. If so,

<sup>25</sup> Crispin Wright speaks of “the *expressivist* tradition of commentary,” which interprets Wittgenstein as “denying that avowals are so much as *assertions*—that they make statements, true or false—proposing to view them rather as *expressions* of the relevant aspects of the subject's psychology” (Wright, “Self-knowledge: The Wittgensteinian Legacy,” in C. Wright, B. C. Smith, and C. Macdonald, eds., *Knowing Our Own Minds* (Oxford: Oxford University Press, 1998), p. 34). The present essay arises out of a conviction that we should take seriously (more seriously than, e.g., Wright himself does) Wittgenstein's oft-repeated suggestion that mental state avowals be understood as (or as akin to) expressions, *without* reading him as an expressivist.

sentences like 'I am in pain', which ostensibly report bodily sensations, actually just express them.

But however suggestive this idea may be, it is plainly possible to report explicitly that we are in such states. And it is indisputable that others sometimes assert of us that we are, or are not, in particular mental states, and we sometimes explicitly contradict what they say. It is not just that we undermine what they say, as I might by saying 'ouch' when you say I am not in pain. Rather, we literally deny what others say about us. If we were unable to report on our own states of mind, but could only express them, this direct denial of the ascriptions others make about us would be impossible. If you deny that I am in pain and I simply say 'ouch', we have not thus far contradicted each other.<sup>26</sup>

I want to suggest that when, in his "Plan for the treatment of psychological concepts," Wittgenstein says that it is "Not quite right" to assimilate psychological self-ascriptions to expressions, he is, in effect, rejecting this way of reading what he has to say about such utterances. These utterances aren't *exactly* like facial and bodily expressions: there is a crucial difference, namely, that they have an assertoric dimension. In the end, it is not particularly important whether we say that psychological self-ascriptions *are* expressions or that they are *akin* to them, as long as we keep in view both how they are like wincing of pain and how they are unlike them.

I should add that part of what is at issue in the passage from Rosenthal's paper is how to understand Wittgenstein's use of the word "report." Earlier, I quoted (part of) the following from the *Investigations*:

When someone says "I hope he'll come"—is this a *report* [*Bericht*] about his state of mind, or an *expression* [*Äußerung*] of his hope?—I can, for example, say it to myself. And surely I am not giving myself a report. (*PI* §585; I have departed slightly from Anscombe's translation.)

How should we understand the use of the word "report" that we find here? Should we, following Rosenthal, take "report" to mean, merely, *an assertion—an utterance with truth conditions*? If so, the final sentence of this passage looks mysterious. Why would Wittgenstein think it *obvious* that I cannot be saying something *true* to myself? A better gloss on what he means by "report" would be: *an attempt (or apparent attempt)<sup>27</sup> to inform someone of some fact or facts that the speaker has*

<sup>26</sup> David Rosenthal, "Thinking That One Thinks," in M. Davies and G. W. Humphreys, eds., *Consciousness* (Oxford: Blackwell, 1993), p. 203. For another reading along these lines, see R. J. Fogelin, *Wittgenstein* (London: Routledge, 1987), pp. 197–8.

<sup>27</sup> A report may be a lie.

*Plan for the Treatment of Psychological Concepts* 235

*learned*. On this understanding of "report," that final sentence makes sense. But notice that on this understanding of "report," it does not follow from a self-ascription's not being a report that it is not an assertion.

## 6. CONCLUSION

There's a moment in *LaWrPP2* at which Wittgenstein's interlocutor asks, "Can one know what goes on in someone else in the same way he himself knows it?" Many philosophers—for example, Russell in 1912, in *The Problems of Philosophy*—would answer that one cannot. According to Russell in 1912, what goes on in a person is, in principle, accessible only to him. A few philosophers—for example, Russell in 1921, in *The Analysis of Mind*—would answer that one can know what goes on in someone else in the same way he himself knows it, by observing his behavior and inferring his state of mind. Still others, including a number of contemporary philosophers, would answer that although *at present* one cannot know what goes on in someone else in the way he himself knows it, this is a contingent limitation on our powers.<sup>28</sup>

How does Wittgenstein answer his interlocutor's question? He doesn't. Here is the whole passage from *LaWrPP2*:

"Can one know what goes on in someone else in the same way he himself knows it?"—Well, how does he know it? He can express his experience. No doubt within him whether he is really having this experience—analagous to the doubt whether he really has this or that disease—comes into play; and therefore it is wrong to say that he knows what he is experiencing. But someone else can very well doubt whether that person has this experience. Thus doubt does come into play, but, precisely for that reason, it is also possible that there is complete certainty. (p. 92)

According to Wittgenstein, his interlocutor's question presupposes too much, or better—it comes too late. Regardless of whether one answers

<sup>28</sup> Thus, David Armstrong writes:

We can conceive being directly hooked up, say by a transmission of waves in some medium, to the body of another. In such a case we might become aware, for example, of the movement of another's limbs, in much the same sort of way that we become aware of motion of our own limbs. In the same way, it seems an intelligible hypothesis (a logical possibility) that we should enjoy the same sort of awareness we have of what is going on in our own mind. A might be 'introspectively' aware of B's pain, although A does not observe B's behavior." D. M. Armstrong and N. Malcolm, *Consciousness and Causality* (Oxford: Blackwell, 1984), p. 113.

Yes or No—whether one affirms the privacy of the mental or denies it—one misses the essential point: that the relation a person bears to his own experience is not *epistemic*; it is not a matter of his knowing or failing to know something in this way or that. To get this relation into view, we must attend to the expressive dimension of mental state self-ascriptions.<sup>29</sup>

At *PI*, p. 222, Wittgenstein describes the ambition of this region of his philosophizing as one of condensing a whole cloud of philosophy into a drop of grammar. My aim in this paper has been twofold: first, to characterize a cloud of philosophy, one that we might call the dialectic between detectivism and constitutivism; and, second, to show how Wittgenstein's "Plan for the treatment of psychological concepts" offers the sketch of a plan for condensing this cloud into a few drops of grammar—drops of grammar such as, "The first person of the present akin to an expression."

<sup>29</sup> On p. 221 of *WNL*, Wittgenstein has just been criticizing the idea that certain inner goings-on are private, whereupon an interlocutor replies: "But do you really wish to say they are not private? That one person can see the picture before the other person's eye?" Wittgenstein would not have us *deny* that inner goings-on are private. *Given* a picture of the mental according to which inner items appear before the mind's eye, such items *ought* to be called private. But it is precisely this picture (and related ones) that Wittgenstein would have us call into question. He writes:

The great difficulty here is not to represent the matter as if there were something one *couldn't* do. As if there really were an object, from which I derive its description, but I were unable to shew it to anyone. (*PI* §374)