

Précis of *Expression and the Inner*

David H. Finkelstein

Expression and the Inner (hereafter, “E&I”) has two parts. In Part I, which comprises the book’s first three chapters, I outline and criticize three approaches toward understanding what philosophers call “first-person authority,” i.e., the sort of authority with which we speak about our own attitudes, feelings, emotions, or sensations. In Part II, which comprises the next three chapters, I spell out what I take to be a better approach toward making sense of first-person authority, one that has its roots (and many of its branches too) in Wittgenstein’s late writings. Also in Part II, I discuss a variety of related topics, among them: Wittgenstein on rule-following and on the expressive dimension of psychological self-ascriptions, the distinction between conscious psychological states and unconscious ones, and what the mental life of a non-linguistic animal has in common with our sort of mental life. In what follows, I’ll say a little bit about each of the book’s six chapters.

Chapter 1. Detectivism

At least since Locke’s *Essay*, one influential approach toward understanding self-knowledge and first-person authority is what I call “detectivism.” A detectivist is someone who holds that we speak with authority about our own mental states and events because they are conscious and that they are conscious by virtue of a mechanism or process that enables us to detect them. Almost all detectivists think there is a significant parallel between the sort of relation we bear to the objects and events that we perceive (with, e.g., our eyes) and the sort of relation we bear to our own mental states and events. According to what I call “old detectivism,” there are, nonetheless, significant disanalogies between “outer sense” and “inner sense.” An old detectivist thinks of self-knowledge in something like the way that Russell (1912) thought of acquaintance with sense data: we know our own states of mind thanks to an infallible sort of detection that provides us with access to them that is more direct and certain than seeing or hearing can be. In 20th Century

analytic philosophy, we find a displacement of old detectivism by what I call “new detectivism.” New detectivists typically seek to domesticate old detectivism — to render it less Cartesian and more naturalistic — by understanding self-awareness as a species of ordinary, fallible, garden-variety perception. The burden of Chapter 1 is to show that neither old nor new detectivism provides a satisfactory explanation of consciousness or first-person authority.

Chapter 2. Constitutivism

In his late work, Wittgenstein often expresses opposition to detectivism. Beginning in this chapter of E&I and continuing off and on through the rest of the book, I discuss and criticize two opposed views of what we can learn from Wittgenstein about first-person authority — one suggested by Crispin Wright and one by John McDowell. Wright holds (correctly, I think) that Wittgenstein’s position concerning psychological self-ascription is intimately bound up with his views about meaning and rule-following. In Wright (2001b) and (2001a), he offers an account of what Wittgenstein can teach us about the self-ascription of attitudes that (more or less) falls out of his reading of Wittgenstein on rule-following. In Chapter 2 of E&I, I discuss Wright’s views about both rule-following and first-person authority.

Wright spells out what I call a “constitutivist” account of the self-ascription of attitudes. A constitutivist claims that a person speaks with authority about her own, e.g., intentions because when she avows that she intends to ϕ , she makes it the case that she does so intend. According to constitutivism, the authority with which one speaks about one’s own attitudes is more like that of an army colonel when he *declares* an area off-limits than it is like that of an eye-witness when he says what he saw. I argue that we should not be satisfied with either Wright’s account of rule-following or his constitutivist account of first-person authority. I go on to argue that any constitutivist account of first-person authority is liable to misrepresent the responsibility that we bear for our own states of mind — either by portraying us as active in their constitution when we are not or by mischaracterizing the ways in which our activity figures in their constitution.

Chapter 3. Between Detectivism and Constitutivism

In this, the last chapter of Part I, I turn to an account of how we know our own states of mind that combines elements from both detectivism and constitutivism. I call this the “middle path account” or “MPA”. As I read McDowell, he defends MPA both on his own behalf and on Wittgenstein’s. I discuss McDowell’s reading of Wittgenstein in Chapter 6 of E&I; in Chapter

3, I set Wittgenstein aside and focus on the way in which MPA makes its appearance in McDowell's *Mind and World*.¹

In *Mind and World*, McDowell recommends that we “think about ‘inner sense’ in parallel with ‘outer sense’ to the fullest extent that is possible” [McDowell (1996), p. 22]. He argues that, in order to avoid succumbing to the Myth of the Given in our theorizing about perceptual knowledge, we should understand perceptual judgments to be justified by sensory impressions that have conceptual — indeed, propositional — contents. And to avoid the Myth in our theorizing about self-knowledge, McDowell thinks we should understand a subject's judgments about his own “perceptions, thoughts, sensations, and the like” [ibid., p. 18] to be justified by impressions that are akin to those of “outer sense,” only directed inward. He says that these “impressions of ‘inner sense’ must be, like the impressions of ‘outer sense,’ passive occurrences in which conceptual capacities are drawn into operation” [ibid., p. 22].

So far, McDowell's view sounds like a kind of detectivism according to which a person would judge that he was in pain on much the same sort of basis as he might judge that there was a mug on his desk. But there is, McDowell says, a crucial difference between judgments of “inner sense” and those of “outer sense,” viz., “[O]bjects of ‘inner sense’ are internal accusatives to the awareness that ‘inner experiences’ constitute; they have no existence independently of that awareness” [ibid., p. 21]. The mug that I see on my desk exists independently of my visual awareness of it. But the pain that I feel in my wrist has no existence independently of my awareness of it; it is nothing over and above the actualization of certain conceptual capacities. Thus there is a constitutive element in McDowell's story: an item in a stream of consciousness is constituted by a subject's awareness of it. (McDowell speaks here of a “limiting case of the structure of awareness and object” [ibid., p. 38].) In §3.4 of E&I, I argue that McDowell's story ends up bringing awareness and object *too* close together and so doesn't leave enough room for a person to have a sensation whose features he only gradually comes to be aware of.

Chapter 4. Meaning, Expression, and Expressivism

In PI §431, Wittgenstein has an interlocutor say:

“There is a gulf between an order and its execution. It has to be filled by an act of understanding.

“Only in the act of understanding is it meant that we are to do THIS. The order — why, that is nothing but sounds, ink-marks. — ”

The interlocutor's thought might be stated as follows: “Any spoken or written order (or instruction or statement of a rule) is, in itself, just a string of empty

sounds or marks. For such a thing to *call for* one activity rather than another, it must be supplemented by ‘an act of understanding’ — something like an interpretation.” This conception of the relation between a rule and its application gives rise to a regress of interpretations and the famous paradox of PI §201: “no course of action could be determined by a rule, because every course of action can be made out to accord with the rule.” Wittgenstein avoids the paradox — not by telling a story about something that bridges a gulf between every order and its execution, but by questioning, and ultimately rejecting, the impulse to say that whenever someone is given an order or a rule, he is confronted by a string of sounds or ink-marks in need of supplementation.

Just as it can seem, when we are doing philosophy, as if we must understand orders to be really nothing more than sounds and ink-marks — items without semantic significance — it can seem as if we must understand human (and animal) behavior as mere bodily movement without any intrinsic psychological import. The notion of expression figures in Wittgenstein’s late philosophy in a way that’s closely related to the thought — given voice in PI §201 — that “there is a way of grasping a rule which is *not an interpretation*, but which is exhibited in what we call ‘obeying the rule’ and ‘going against it’ in actual cases.” We are often (“in actual cases”) able to hear the meaning in an order or an instruction, and we are often able to see the psychology in a facial or bodily expression, e.g., in a wince or a laugh:

“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 [Z §225].

One of my aims in Chapter 4 of E&I is to introduce the notion of expression as it figures in Wittgenstein’s thought in such a way that its connection to the rule-following problematic comes clearly into view.

Another aim is to criticize a widespread interpretation of an important strand in what he says about expression. Wittgenstein says that the expression of a mental state can take the form of a self-ascription of it. In the *Investigations*, this suggestion is first made in connection with pain, in §244. But the point is not limited to avowals of sensation; throughout Wittgenstein’s late writings, we find remarks such as: “The statement ‘I am expecting a bang at any moment’ is an *expression* of expectation” [Z §53]. Wittgenstein proposes that we understand — not only self-ascriptions of pain and expectation, but also — self-ascriptions of hope [PI §585] and irritation [LWPPII, p. 70], indeed all manner of psychological self-ascriptions [Z §472], as expressions (or as akin to expressions). Such remarks are often read as showing that he held (or perhaps merely flirted with) a patently unsatisfactory view according to which psychological self-ascriptions express their subject matter *rather than* saying anything truth-evaluable. I claim (in §4.3) that there is no compelling reason

to saddle Wittgenstein with this kind of expressivism, and that we ought to try harder to do justice to the suggestion that we think about avowals as, or as akin to, expressions.

Chapter 5. Authority and Consciousness

Throughout E&I, I understand the phenomenon of first-person authority in terms of the following pair of facts: (1) If you wish to know my state of mind, I am usually the best person to ask; and (2) psychological self-ascriptions do not, as a rule, seem to be based on behavioral evidence. Having introduced, in Chapter 4, Wittgenstein's proposal that one of the ways in which we express our psychological states is by self-ascribing them in words, I am able, in §5.1, to offer a very brief account of these two facts and so of first-person authority: First, if you wish to learn my state of mind, I'm usually the best person to ask for the same reason that my face is usually the best one to look at. My psychological self-ascriptions often express, and so make manifest, that which they are self-ascriptions of. And second, I no more need to rely on behavioral evidence in order to avow my happiness than I need behavioral evidence in order to express my happiness by smiling.

If psychological self-ascriptions are so like facial expressions, why do self-ascriptions, but not smiles or winces, present us with a philosophical puzzle about their authority? This, I claim, is at least in part because of a difference between (e.g.) smiles and self-ascriptions of happiness. The latter are truth-apt, while the former are not. This difference leads us, on the one hand, to speak about "authority" in connection with avowals (no one speaks about the "authority" of smiles) and, on the other hand, to find it puzzling that psychological self-ascriptions aren't typically based on inference or observation. (Detectivism just is the postulation of some such basis for them.) But, as long as we keep in view the expressive side of psychological self-ascriptions, there is no reason to find this baselessness puzzling. When Wittgenstein suggests, in PI §244, that "the verbal expression of pain replaces crying," part of what he has in mind could be put this way: When a child learns to speak, and the expression of pain takes a new, linguistic form, a great deal changes. But there's no reason to think that, suddenly, in order for the child to express his pain, he requires epistemic grounds.

In §5.3, I extend the brief account of first-person authority offered in §5.1 by further elucidating the notion of expression and, with it, the phenomenon of first-person authority. I discuss the following example: At a corporate board meeting, the chief executive officer says, about a midlevel employee, "Phillips is a real team-player, by which I mean — not that he's a stupid sheep, but — that he won't help himself at the expense of this company." In the first half of her sentence, the CEO uses an ambiguous term. In

the second half, she says what it means. I suggest that she is able to speak with authority about what she means by “team-player” because the second half of her sentence is not *merely* an interpretation of the first half; it’s an elaboration of it. The weight that the other board members are liable to give the CEO’s interpretation of her own term reflects the fact that the two halves of her sentence make sense together, in light of each other, as (what we might call) a single unit of intelligibility. This example leads to a discussion of Frege’s context principle and how it is taken up and extended by Wittgenstein. By the end of §5.3, I am suggesting that a mental state and its expression should be thought of as making sense together in a way that’s akin to the way in which the parts of sentence make sense together.

E&I arrives at many of its main conclusions in Chapter 5. One of these concerns the distinction between conscious and unconscious mentality. In Chapter 1, much was made of the fact that we don’t speak with first-person authority about *all* of our own attitudes and emotions — only about the conscious ones. A person might come to realize, on the basis of compelling behavioral evidence, that he harbors unconscious anger toward his mother. While such a person might self-ascribe his anger, he could not speak with first-person authority about it. The claims that we make about our own unconscious states of mind are only as good as the evidence we have to back them up. There is, then, a tight connection between first-person authority and consciousness. In §5.4, I take the lessons learned thus far about first-person authority and incorporate them into an account of the distinction between conscious states of mind and unconscious ones. I argue that a mental state is conscious if and only if the subject is able to express it merely by self-ascribing it. Where the subject lacks this ability, his mental state is unconscious.

I go on, in §5.5, to argue that both consciousness and first-person authority should be understood to come in degrees. Sometimes a person self-ascribes a mental state that is neither entirely conscious nor entirely unconscious. In such cases, the self-ascription may be said to lean on, but not rest squarely upon, evidential considerations. (Part of the reason that detectivism can be appealing is that we do often need to consider evidence — both the evidence of our own behavior and that of our feelings and passing thoughts — when we ascribe attitudes to ourselves.)

Chapter 6. Sensations, Animals, and Knowledge

At the start of his *Philosophy and the Mirror of Nature*, Richard Rorty wonders whether we mean anything coherent by the word “mental.” He writes, “The attempt to hitch pains and beliefs together seems ad hoc — they don’t seem to have anything in common except our refusal to call them ‘physical’” [Rorty (1979), p. 22]. Part of what’s at issue in E&I’s final chapter

is the mental as such. I bring together three questions: (1) What does our sentience have to do with our sapience? (2) What does the mental life of a non-linguistic animal have in common with our sort of mental life? (3) How are unconscious mental states related to conscious ones?

Consider the following states of mind: my conscious desire to complete this précis, my longstanding unconscious anger at my brother, the pain I'm feeling in my wrist from too much typing, a leashed dog's desire to chase a nearby squirrel, and a pain felt by an arthritic cat. I argue that we should think of these psychological states as all sharing a kind of intelligibility that the states of mountain ranges and coffee mugs don't share. Part of what this claim comes to is that attitudes, emotions, and sensations may be expressed in behavior, regardless of whether they figure in the life of a human being or a brute, and regardless of whether they are conscious or unconscious. And what this, in turn, comes to can be understood in terms of a thought already alluded to above (in connection with §5.3) and developed in Chapter 6 — that whether we are talking about sentience or sapience, human beings or animals, conscious mental states or unconscious ones, the inner and the outer make sense together, in light of each other, in what I call “the logical space of animate life.”

I devote a fair bit of Chapter 6 to a critical discussion of the way in which McDowell understands a few difficult and important passages in Wittgenstein's *Investigations* — passages having to do with sensations, privacy, and behaviorism, along with what Wittgenstein calls “the grammar that tries to force itself on us” [§304] when we think philosophically about the inner. I'll say a little bit about this in my reply to McDowell's contribution to this symposium.

Postscript. Deliberation and Transparency

After I thought I had finished writing E&I, I read Richard Moran's *Authority and Estrangement* and decided to add a postscript (largely) about the account of first-person authority presented there. According to Moran, the question of whether I believe that p is, for me, “transparent to” the outward-directed question of whether I ought rationally to believe that p . This is to say: I can answer the former question by addressing the latter, and so I can say what I believe without relying on any sort of self-observation. Moran extends this thought about the self-ascription of beliefs to other attitudes: I can, according to him, say what I fear or want or hope by consideration of what I rationally ought to fear or want or hope. Or, rather, I can do this on the condition that the attitude I, in fact, have is in accord with my assessment of what it rationally ought to be. Moran calls this the “Transparency Condition.”

Moran claims that a person has “genuine first-person awareness” [Moran (2001), p. 107] only when he can avow his attitude by addressing the outward-directed question of what the appropriate reasons require of him — so, only

when the Transparency Condition obtains. But there are a wide range of cases in which a person speaks with first-person authority about his state of mind even though he cannot do this. Here is one kind of case: Harry suffers from astraphobia — a fear of thunder and lightning. Harry doesn't take this fear to be rational; he knows that thunder can't hurt him and that lightning is unlikely to, especially when he is indoors. Nonetheless, when he hears a burst of thunder, peeks out from the blanket under which he's huddled, and says, "I'm so afraid," his self-ascription is not based on evidence; he speaks with first-person authority about his state of mind. In that example, Harry's fear is, by his own lights, irrational. Here is another kind of case: I'm fond of my friend Jim's cat, Marmalade. I don't view it as rationally incumbent upon me to be fond of cats in general, or Marmalade in particular. I cannot answer the question of whether I'm fond of Marmalade by addressing a question about whether I rationally ought to be. Nonetheless, I *am* fond of Marmalade, and I can speak about this fondness with first-person authority.

Moran writes, "A proper philosophical account of self-knowledge should tell us how it is that a person can speak about his own mind, without appealing to evidence about himself," where his not appealing to evidence "contributes rather than detracts from the authority of what he says" [ibid., p. 135]. In E&I's postscript, I argue that — once it becomes apparent that one may speak with first-person authority regardless of whether one is talking about a belief that is dictated by the conclusion of theoretical reflection, an irrational fear, or an attitude that is neither dictated by reasons nor irrational — it is hard to be convinced by Moran's account of self-knowledge and first-person authority.

Department of Philosophy
The University of Chicago
 1115 E. 58th St. Chicago, IL 60637, USA
 E-mail: davidf@uchicago.edu

NOTES

¹ McDowell takes issue with the reading of him that I set out in E&I and that I am summarizing here. (See his contribution to this symposium along with my reply.)

REFERENCES

Abbreviated References to Wittgenstein's Writings:

- LWPPII: WITTGENSTEIN, L. *Last Writings on the Philosophy of Psychology*, vol. 2 C.G. Luckhardt and M.A.E. Aue (trans), G.H. von Wright and H. Nyman (eds.), Oxford: Blackwell, 1992.
- PI: *Philosophical Investigations*, G.E.M. Anscombe (trans.), New York: Macmillan Publishing Co., Inc., 1958.
- Z: *Zettel*, G.E.M. Anscombe (trans.), G.E.M. Anscombe and G.H. von Wright (eds.), Berkeley: University of California Press, 1981.

Other References:

- FINKELSTEIN, D.H. (2003), *Expression and the Inner*, Cambridge, MA: Harvard University Press.
- MCDOWELL, J. (1996), *Mind and World*, Cambridge, MA: Harvard University Press.
- MORAN, R. (2001), *Authority and Estrangement: An Essay on Self-Knowledge*, Princeton, NJ: Princeton University Press.
- RORTY, R. (1979), *Philosophy and the Mirror of Nature*, Princeton: Princeton University Press.
- RUSSELL, B. (1912), *The Problems of Philosophy*, Oxford: Oxford University Press.
- WRIGHT, C. (2001a), "Excerpts from a Critical Study of Colin McGinn's *Wittgenstein on Meaning*," in *Rails to Infinity: Essays on Themes from Wittgenstein's Philosophical Investigations*, Cambridge, MA: Harvard University Press.
- (2001b), "Wittgenstein's Rule-Following Considerations and the Central Project of Theoretical Linguistics," in *Rails to Infinity: Essays on Themes from Wittgenstein's Philosophical Investigations*, Cambridge, MA: Harvard University Press.