1
Privileged Access

Dorit Bar-On and Crispin Wright

Contents

***

§1.1 Privileged Access: What Is the Problem?

Although some philosophers have preferred to think otherwise, it is not philosophical theory but a part of the ordinary folk notion of the mental, enshrined in literature and drama, that each of us stands in a special relationship, denied to others, to our own mental lives—that (many of) our mental states and attributes are (normally) directly available to us and only indirectly available to others, that "You cannot really know what another is thinking," for example, whereas of one's own (occurrent) thoughts one cannot but be aware. This special relationship—often referred to by the term *privileged access*—seems to embrace (some instances of) each of sensation, mood, emotion, belief, desire, fear, intention, action (what I am currently doing), memory (what I am currently remembering), perception

Expression and Self-Knowledge, First Edition. Dorit Bar-On and Crispin Wright.

(what I am seeing), thought (what I am thinking), imagination, (idiolectic) meaning … the list goes on.

The putative special relationship is frequently gestured at by the phrase "first-person authority." But the term, "authority," oversimplifies the respects in which, according to the folk notion, self-knowledge is special. As an initial approximation, there are *three* distinct features to reckon with.

First (*Immediacy*), in a wide class of cases, your knowledge of your own psychological attributes need not rely on the kind of evidence—what you say, how you act, how you look—on which others have to rely in coming to justified views about your psychological attributes, and may often seem to rely on no kind of evidence at all.

Second (*Authority*), your knowledge of a wide class of your own mental attributes does not only originate differently from the ways others can know of them—it is generally *superior*, and to be deferred to.

And third (*Salience*), we tend, in the round, to know what there is to know: our own mental attributes of the kinds listed do not, in general, elude our awareness, although those of others may often do so.[1,2]

We shall refer to these three features collectively as the *Trifecta* and, sometimes, will use "Privileged Access" and "First-Person Privilege" as catch-all terms for the three.[3]

---

[1] *Immediacy* and *Authority* approximately correspond to what Alex Byrne has termed, respectively, *peculiar* access and *privileged* access (2005: pp. 80–81; 2018: §1.3). Salience—sometimes dubbed "transparency" (Bilgrami 2006), "self-intimation" (Shoemaker 2009) or "luminosity" (Williamson 2000)—seems not to figure in Byrne's landscape of the issues. We have struggled with the choice of a term here. "Transparency," "self-intimation," and "luminosity" are all naturally understood as connoting something much more automatic than we want to imply. "Salience" itself has unwanted resonances of the visual. However, we are going to stick with "salience." The reader needs only to fix on the following intuitive idea: that those kinds of mental states, S, which allow in general of authoritative, non-inferential self-ascription are also such that a suitable conceptually endowed subject in S will at least be in position to know that they are in S.

[2] That self-knowledge of the relevant kinds of mental state can and characteristically does exhibit each of these three features is disputed by a number of writers on the topic. The philosophical problem of self-knowledge will, naturally, reconfigure itself—perhaps disappear altogether—depending on whether, or the extent to which, they are right. We will attempt to address, and to rebut, the principal arguments of four leading exponents of such skepticism in Chapter Two.

[3] Readers will also encounter uses of "distinctive security" ascribed both to avowals and to self-knowledge. We have struggled with the selection of an upfront, comprehensive, canonical terminology to hit off the explananda at issue, partly because of the switch of target involved in moving from the thought-first to language-first points of view (see §1.3) and partly because substantial philosophy is needed to say exactly what the explananda are on either viewpoint. There are some important distinctions which will emerge only in the course of subsequent discussion, so we ask for readers' indulgence at this point. Readers should also be aware that Bar-On has misgivings about whether Salience deserves its place. The matter will be discussed further in later chapters.

So, the problem—as it initially impresses—is to provide a philosophical account of these asymmetries. Knowledge generally is democratic in the sense that ways to achieve knowledge about various subject matters, while they may *in practice* be exclusive to those with special gifts of intellect, opportunity or training, are *in principle* open to anyone. Self-knowledge, by contrast, of perfectly everyday kinds, is routinely achieved by the self concerned in ways that are denied to others, even in principle. What explains this?

§1.2 The Cartesian "Solution"

There is a venerable response to this question, which springs to mind so naturally that it can seem constitutive of the very subject matter, an aspect of the "data." This is the idea, associated (perhaps erroneously) by many modern thinkers with Descartes,[4] that each one's mental life constitutes a totally transparent *inner theater*, with an audience, necessarily, of one. Others will need indirect evidence to suppose that something is happening on your inner stage, but you can just observe it. So of course you know best. And since there is total transparency, you will be able to observe what is there to be observed.

Familiarly, however, when pressed, the venerable response transpires to be deeply problematic on several counts. To begin with, the appeal to a kind of interior observation involves distortion of the phenomenology of many—too many—of the attributes that intuitively fall within the province of privileged access. For example, when you acknowledge that you believe something, or have a certain intention, there is normally no distinctive, individuative state of consciousness involved, as there is with, say, a toothache or an itch—nothing in particular that "it is like" to have that particular belief or intention. But more theoretically, the inner theater model plays a villain's part in generating the sceptikal problem of other minds. And its very coherence is put in serious question by the misgivings about "private language"—really, they concern the possibility of private *conceptualization*—original to Wittgenstein in the *Philosophical Investigations*.

Since constraints are thereby imposed on any more satisfactory account, it is worthwhile reviewing these points in some detail. For Cartesianism, our access to our mental states is afforded by a superior kind of inner perception, a faculty that brings us to awareness of the character of states, events, and processes within our inner mental lives broadly as our ordinary perceptual faculties bring us to awareness of states, events, and processes taking place in the perceptible material

[4] Again, since they have some currency, we keep the label "Cartesian" and cognates for convenience, but intend no attribution to the actual historical Descartes.

world around us. But there are crucial differences. While a perceptually normal subject paying close attention to the objects around them will normally be accorded a certain authority for what is there and how matters stand with it, and will normally be expected to notice these things, the Cartesian view, as traditionally understood, embraces far stronger—maximally strong—versions of Authority and Salience:

Absolute First-Person Authority: If a (sincere, competent) subject spontaneously avows being in mental state M, then their avowal is absolutely *indubitable, infallible*, and *incorrigible*.

Guaranteed Salience: If a subject S is in mental state M, then they are *guaranteed* to know that they are in M.

Moreover, the access given by the inner gaze is comprehensive yet local:

All-Encompassing Yet Restricted Epistemic Access: A subject's inner gaze encompasses *all* their occurrent mental attributes but exclusively concerns *their own* mental attributes. It does not take in their physical states, or the mental states of others.

And finally:

Necessarily Privileged Self-Knowledge: The matters which a subject's avowals concern are things which the subject necessarily knows.

According to the above theses, sincere avowals would constitute self-attributions that are guaranteed to be true, could never rationally be doubted, and could never reasonably be corrected by others. As such, they would present a radically unusual subclass of things we know and would contrast sharply not only with mental attributions to others but also with all bodily self-attributions. Self-attributions of bodily features and conditions such as height and weight, disease, heart rate, digestive processes, etc. are often made on the same kinds of bases as similar attributions to others. And (as we shall later see) even self-attributions of limb position and bodily orientation, as well as self-attributions of what we see, hear, or touch, or of what we are doing at a given moment—all of which do exhibit certain notable first-person/third-person contrasts—seem open to straightforward rejection and correction by others and are even open to doubt by oneself. Though one does not normally tell what the position of one's legs, or whether one is sitting down, or whether one sees something, or is doing something, in the same way in which others determine these facts about oneself, others can nevertheless be in a position straightforwardly to deny and correct a self-attribution such as "My legs are crossed" or "I am sitting down"; similarly, e.g., "I (am) hear(ing) a loud siren," or "I'm drawing a horse." By contrast, on the Cartesian conception, the corresponding avowals ("It *feels to me* as though my legs are crossed," or "I *feel* as though I am standing"; "I seem to be hearing a loud siren") are absolutely indubitable, infallible, and incorrigible: with respect to such

self-attributions, one enjoys absolute authority. (We will return to contrasts between first-person bodily and mental self-attributions later on.)

On the Cartesian conception, this absolute authority is due to the fact that we each have our own personal, absolutely secure form of access to our own present states of mind. The results delivered by this form of access are guaranteed to be true. There is a striking corollary of this. Since it seems impossible that we could have such secure access to any of our bodily states, the privileged epistemic access we each have to our mental states requires those states to have a *peculiar non-bodily nature*. Hence, as noted, the pressure to adopt the following, metaphysical thesis—one that the historical Descartes undoubtedly did hold—regarding the relation between mind and body:

Substance Dualism: The states known via the "inner gaze" are states of an immaterial substance that is (only contingently) associated with our material body.

Given this thesis, it transpires that avowals and nonmental bodily self-attributions differ in the metaphysical kinds of subject matter they concern. Avowals concern states of our immaterial minds, which are metaphysically distinct from, and only contingently coexist with, our bodies. These states are, moreover, *private* states—they cannot be objects of direct awareness to others who can know of them, if at all, only by less secure, inferential means.

As we suggested at the beginning, the Cartesian conception of these matters so elaborated is not merely a philosopher's invention, but resonates with at least one strand in a widely (if somewhat inchoately) held *folk-theory*. Its appeal, or at least its status as an implicit assumption, is witnessed by, for example, the widespread assumption that it at least makes sense to suppose that one might survive one's bodily death and by the seeming obviousness, both to freshmen and the folk, that there is indeed a "problem of other minds" as traditionally viewed. If mental states are things of which only their subject can be directly aware, then—never mind the superlative character of the knowledge they at least supposedly have of them—what grounds can *others* have for thinking that one has such states at all? And even if there are some such grounds, what reasons are there for thinking that different selves' mental theaters are populated by things of even remotely similar kinds? And if there are no strong grounds for that, what grounds are there for thinking that the meanings of expressions in our respective mental idiolects overlap to any significant degree, and that the mental vocabulary we each deploy in speaking and thinking is not our own exclusive property?—that it is not, after all, a "private language" as famously challenged by Wittgenstein in the *Philosophical Investigations*?

This Cartesian slant, we need to stress, does not exhaust the folk-conception of the mental. Indeed, there is reason to doubt that folk-philosophical thought about the mental permits of an overall, coherent systematic account. For, as we shall later have occasion to emphasize, the folk *also* appear to hold that we

normally *can* tell what others feel, want, intend, and think, very often even without consulting them. Such mental attributes are, we think, *very often manifest* in their behavior. When we take our children and other family members, or even people we casually come across and, indeed, nonhuman animals, to be agitated, scared, upset, puzzled, and so on, we do not ordinarily suppose that we are hazarding an inferential hypothesis about internal causes, designed to explain their outward behavior or predict what they will do next. Very often we take it that we can just *see* someone's fear in their eyes, *hear* their anger or nervousness in their voice, *feel* their agitation in their tense body, and so on. If there is a "commonsense" conception of our relation to other minds, it is at least not Cartesian through and through.

We will not here venture to add to the vast literature on the legacy of the Cartesian conception, nor try to add to the debates about Private Language.[5] With the exception of the following comments, we will assume—with the majority of those currently seeking an account of the asymmetries between self-knowledge and knowledge of others—that such an account had better avoid the metaphysical extravagance of Cartesian dualism. We think, indeed, there is a reason for avoiding this extravagance that goes beyond a general preference for the relative parsimony of a broadly materialist ontology (as well as more specific concerns about the dubiety of positing *immaterial* states of individuals that have the potential to interact *causally* with states of their material bodies). Rather, one further, in our view compelling reason for resisting the Cartesian view lies in the problematic character of its explanatory methodology.

In effect, the Cartesian view offers a metaphysical postulate to solve what is in essence an epistemological puzzle, one that concerns observed contrasts between our apparently privileged knowledge of our current mental states, on the one hand, and our knowledge of all other matters of fact (including others' mental states, and our own bodily states) on the other. Given the essentially epistemological character of this puzzle, it seems entirely unclear how postulating a realm of immaterial states of our minds can, by itself, help explain in any way why it is that, as subjects of such states, we are *guaranteed* to have direct and absolutely secure access to them. This will continue to remain unclear, *unless* we accept—as the Cartesian in fact appears to suppose—that immaterial states of mind, *by their very nature*, offer precisely such a guarantee. But to rely on this supposition is hardly to provide an *explanation* of the epistemic privilege we apparently enjoy. Instead, it is simply to build the epistemic privilege, *ad hoc*, into the very metaphysical characterization of the postulated states.

---

[5] For some of our contributions to this area, see Wright (1984), (1986), (2001c); Bar-On (1992), (2004a: chapters two and ten), (2009), (2016).

We indeed need to do better. And this means seeking an explanation of the Trifecta that avoids relying on otherwise unmotivated characterizations of the nature of the states we self-ascribe in our avowals, from which the Trifecta supposedly follow.[6]

§1.3 Language First or Thought First?

So, the cardinal philosophical problem of self-knowledge, as it will be interpreted in this book, is to do better than the Cartesian response: it is to achieve a satisfactory perspective on what is correct about the idea of privileged access—on what is special about a subject's self-conception—which can head off the aporetic elements of the Cartesian response, while at the same time finding a place for (perhaps refined versions of) the plausible-seeming asymmetries that, albeit dubiously, motivate it.

What will count as the achievement of such a perspective? Maybe we cannot say much more at this stage than that "we will know it when we see it," but we do well to note straight away that the Trifecta of Immediacy, Authority, and Salience allow of two radically different directions of interpretation. On one hand—as so far we have implicitly been doing—the three ingredients can be viewed as features of the *cognitive relationship* in which a subject stands to their own mental attributes. On the other, however, quite differently, they can be viewed as *features of the practice of the discourse* in which we speak of our own and others' mental attributes. The term "avowal," employed with the sense, roughly, of: authoritative, non-inferential psychological self-ascription, is used as if already familiar in Ryle's *The Concept of Mind*, and was evidently a term of art in debates in the philosophy of mind for some time before that work. But ever since the publication of Ryle's book, one important tendency of analytical philosophical discussion of psychological self-knowledge has been to focus on its linguistic

[6] There is a point of contact here with Christopher Peacocke's idea of the "Integration Challenge" that arises in a number of areas of philosophy (Peacocke 1999). Peacocke's point is that proffered accounts of the metaphysical character of a certain subject matter—for example, platonism about mathematical entities, physicalism about mental states, or Lewisian modal realism about the truth-makers for counterfactual conditionals—should harmonize with an *independently* motivated account of how our judgments about the relevant subject matter can be knowable—and knowable, ideally, by the methods we actually regard as conferring knowledge of the subject matter in question, with whatever degree of certainty we consider that those methods confer. You duck the Integration Challenge if you simply *postulate* epistemic capacities to reconcile the knowability of a certain subject matter with preconceived ideas about its metaphysical character. And you also duck it if, conversely, you simply *postulate* a metaphysical character to reconcile the knowability of a certain subject matter with preconceived ideas about how knowledge about it may be achieved. Cartesian dualism ducks the Integration Challenge.

expression and on the Trifecta as reflected in characteristics of the competent use of avowals and the competent ascription of mental states to others. How do the Trifecta present themselves when viewed in this *Language-First* way? As a first approximation:

Immediacy will have to do, e.g., with the inappropriateness of requesting grounds for a subject's avowals.

Authority will have to do, e.g., with the propriety of deference to what subjects have to avow.

Salience will be manifested, e.g., in the inappropriateness of professions of ignorance of one's own avowable states.

And these points—no doubt needing refinement to accommodate details of context and variations in the kinds of mental attribute concerned—will be regarded, when viewed from a Language-First perspective, not as reflections of differences in the character of self- and other-knowledge but as about the "language-game," that is, as grounded in our practices of ordinary psychological self- and other-ascription.

So, very early in our investigation, we confront a major fork. What comes first here: the linguistic practice, or the thoughts of the thinkers manifested in that practice? The problem of self-knowledge will look very different depending on how we choose. On the first option, we may be inclined to begin by explaining the distinctive marks of avowals without invoking the idea that they derive from first-personal cognitive advantage. Our problem will be viewed as that of accounting for the distinctive use of avowals—their "grammar" in the terminology of the later Wittgenstein—and its contrasts with the grammar of the competent ascription of mental states to others. A range of candidate explanations will then enter the fray, for instance exactly those falling under the broad rubric of the Expressivist approach under debate in this volume, which would otherwise be excluded.[7] On the latter option, by contrast, the relevant features of avowals will, from the outset, be seen as reflections of aspects of the *epistemology* of the self-directed thoughts they express. And it will seem overwhelmingly natural to suppose just what the folk notion, from which the Cartesian view springs, does suppose: that selves characteristically know of the states that give rise to avowal in a way that involves no inference from independent reasons, and which is characteristically very secure; and that the states of the relevant kind are typically salient to their subjects. The problem will then be to account for these apparent epistemic advantages in a way that guards against a slide into Cartesian privacy and its associated nemeses.

***

[7] For different versions of this approach, see, e.g., Jacobsen (1996), Falvey (2000), and Finkelstein (2003). See also Bar-On (2004a). (However, in Chapter Six, Bar-On explains why it is a misunderstanding to regard her view as "language-first.")

So much by way of introduction. The program for the discussion in this volume—not all of which is "debate," for there is much on which we are agreed—is as follows.

Chapter Two will attempt to persuade the reader that what has been sketched is, in the context of some recent skepticism about its genuineness, a well-conceived philosophical problem. We will first consider our problem in the light of the later Wittgenstein's deflationary attitude to philosophical problems in general before moving to review the arguments of four more recent proponents of skepticism about Privileged Access.

Chapter Three will review what we consider to be the shortcomings of a wide range of approaches which conceive of Privileged Access as a product of one or another putative kind of advantageous epistemic basis that selves have, qua selves, for their self-ascriptions of mental characteristics.

In contrast to the "epistemic access" proposals reviewed in Chapter Three, Chapter Four will review a range of attempts to shed light on Privileged Access by arguing for constitutive connections between a subject's mental states and what they are sincerely willing to avow or between our capacity for self-knowledge and other distinctive capacities we have as rational or moral agents. These too, we will argue, fail to meet certain desirable constraints. The chapter concludes with a compilation of desiderata that we propose a more satisfying account should meet.

Chapter Five will articulate a range of initial criticisms of traditional (or "Simple") forms of expressivism about avowals and one development of them—here termed Radical Expressivism—which differs significantly from the Neo-Expressivism that Bar-On will expound and defend in later chapters. These first five chapters conclude that part of our discussion in which we are broadly in agreement.

Thereafter, in Part Two of our exchange, we embark on our debate. In Chapter Six, Bar-On will provide an exposition of the Neo-Expressivism originally developed in her (2004a), highlighting important ways in which it departs from both Simple and Radical Expressivism, as well as from other views criticized in previous chapters. Then, partly in response to objections to Neo-Expressivism raised by critics, Chapter Seven will offer Bar-On's current take on the contribution Neo-Expressivism can make to our understanding of the epistemology of basic self-knowledge.

In Chapter Eight Wright will develop a range of criticisms which, if sustained, show that Neo-expressivism too is unsatisfactory and presents no significant gain over its Simple and Radical ancestors. Bar-On will respond to these criticisms in Chapter Nine, arguing that, on the contrary, Neo-Expressivism does represent a considerable improvement over these earlier versions of the view.

Wright will use Chapter Ten to outline a research prospectus that aspires to do better than, according to his arguments, Neo-Expressivism can. This prospectus works under the hypothesis that privileged self-knowledge admits of no

uniform account but is rather to be explained in a variety of ways, depending on the kind of mental state concerned. In the final Chapter Eleven, Bar-On will raise several difficulties for Wright's prospectus and question the cogency of the pluralist approach he advocates.