Search in this book

CHAPTER

Introduction 🔓

Crispin Wright (ed.), Barry C. Smith (ed.), Cynthia Macdonald (ed.)

Subject: Epistemology, Philosophy of Language, Philosophy of Mind

Collection: Oxford Scholarship Online

I The Basic Problem

The temple of the Delphic oracle carried the legend, γνωθι σαυτον, 'know thyself'. If our powers of self-knowledge are as ordinarily conceived, then nothing is easier. Each of us effortlessly knows an enormous amount about those of our attributes which go with our rationality, sentience, and affective susceptibilities: our beliefs, hopes, desires, and fears, whether we have a headache or an itchy toe, whether we are elated or depressed, whom we love or hate, what attracts or repels us. We are unhesitatingly sure about the normal run of our intentional states, both content-bearing (states that *p*) and object-directed (thinking of *x*, minding about *y*), our sensations, and our emotions. Having great sweeps of such knowledge is the normal condition of a mature human being. True, it is also normal to know at least some other folk very well. However, knowledge of the sensations, emotions, and intentional states of others demands reliance upon independently articulable grounds— considerations concerning their sayings and deeds and context. By contrast, self-knowledge is characteristically immediate, at least in basic cases (though self-interpretation has its place).

This immediacy would already be enough to raise a philosophical issue. For, of course, the mere fact that an attribute belongs *to* a particular subject will not in general enable her to know of it differently from the way anyone else can know of it. If someone wants to know their height, or location, or precise date of birth, they must carry out exactly the same kind of investigation as anyone else. So why, when the case is that of (some) psychological attributes, is the situation any different? But there is more. Not merely is it not necessary in general, in order for them to count as knowledge, that one's opinions concerning one's own psychological attributes be grounded in evidence. Typically, they are received as authoritative just in virtue of their being held. Not only do you know *differently* from others what you think, hope, and feel; you ↳ are also (defeasibly) regarded as knowing *best*. If you take it that you exemplify a certain mental attribute, then, *ceteris paribus*, that is an authoritative indication that you do. Indeed, the converse also obtains: mental attributes are characteristically *salient* to the subject—if you hope that *p*, or are frightened of *x*, or have toothache, then it is to be expected that you will know it.

These three features of the epistemology of self-knowledge—*immediacy*, *authority*, and *salience*—combine to set a certain explanatory problem to which the first six chapters in this collection react, in their various ways. Crispin Wright is concerned with the issue in the context of the philosophy of mind of the later Wittgenstein. He provides detailed stage-setting for the explanatory problem, and argues that the Cartesian conception of the mind—the acceptance of mental states as composing an inner realm, directly knowable only by their subject—is best viewed as a product of an attempt to solve the problem by means of a broadly observational conception of self-knowledge. He contends that it is a major achievement of Wittgenstein's later philosophy to expose the essential inadequacy of the Cartesian response, by considerations prior to and independent of its culmination in aporia, in the sceptical problem of other minds. However, the primary focus of Wright's paper is the elusiveness of any constructive alternative in Wittgenstein's work. The 'expressive' concept of first-person psychological claims which some early commentary found in the *Philosophical Investigations* is explored, and, after some refurbishment, is shown to be useless for the purpose. Rather, Wright suggests, we should regard Wittgenstein's attitude here as of a piece with his general admonitions against projects of philosophical explanation. He concludes on a note of qualified dissatisfaction with this response.

Wright's chapter (1) is, in part, a continuation of an earlier exchange with John McDowell,[1] and McDowell here responds again (Ch. 2). He is dissatisfied with Wright's contention that the captivating power of the Cartesian conception may be traced to the apparent explanation it offers of the various points of asymmetry between self-knowledge and knowledge of others, arguing that it is a point of contrast between self-knowledge and genuinely perceptual knowledge—a point which the Cartesian conception simply ignores—that the former is *groundless* in a way in which the latter is not. The proper explanation of the attractiveness of Cartesianism is, in

p. 3   McDowell's view, different, having to do ↳ with the distinctively seductive shape that the Myth of the Given takes in reflection about the inner.[2]

The other principal point of difference between Wright and McDowell is this. In Wright's treatment, Wittgenstein, although he offers no constructive philosophical explanation of the distinctive differences between self-knowledge and knowledge of others, nevertheless makes a positive contribution: the thesis that demand for explanation is inappropriate—something which Wright allows may be defensible, but, if so, on grounds that Wittgenstein himself never fully provided. McDowell responds that since the attraction of Cartesianism was not that of an explanation, its demise need not be seen as generating an explanatory vacuum; nor does he see Wittgenstein as recommending the positive thesis that the attempt to provide an explanation here would be inappropriate. For McDowell is not satisfied that it emerges sufficiently clearly from Wright's account what exactly the demand for explanation is supposed to come to.

Christopher Peacocke's careful, fine-grained contribution (Ch. 3) focuses on the case of conscious occurrent attitudes, and brings to bear the kind of apparatus of concept individuation via possession conditions which he pioneered in *A Study of Concepts*.[3] One effect of the view which Wright (*pace* McDowell) finds in Wittgenstein would be that the superiority of a thinker's views concerning her own mental attributes would be a *constitutive principle*—a primitive constraint on what anyone should take it to be correct to think about those attributes—rather than a reflection of any genuinely cognitive advantage. It is very natural to think that if both perceptual and inferential accounts of the cognitive mechanics of self-knowledge are rejected, this broad kind of 'default' view remains the only option. An important contention of Peacocke's discussion is that this is a false trichotomy. The core of his proposal is an account of the possession conditions of the concept of belief in particular which explains how, contrary to the kind of view which Wright elicits from Wittgenstein (and the 'no reasons' account of Sydney Shoemaker[4]), a conscious belief can *give reason for* the corresponding second-order judgement without it being the case either that the latter should be regarded as justified by inference or that it is justified by introspection. In Peacocke's view, we need, and can provide for, a rational basis for the self-

p. 4   ascription of psychological states which is neither inferential nor observational, yet which allows such ↳ ascriptions to be answerable to—to 'track'—independently characterizable first-order states. Peacocke proceeds to elaborate this account of conscious attitudes in such a way that, as he observes, it clashes with a

number of recently advocated theses about consciousness: for instance, that the consciousness of a mental state can be explained in terms of a tendency of states of that kind to generate self-ascriptions, that a purely reliabilist account of second-order beliefs can explain their status as knowledge, and that the consciousness of occurrent propositional attitudes consists in their extensive inferential integration.

Michael Martin's response (Ch. 4) focuses on the core of Peacocke's account. He brings forward examples of self-ascriptions which are based on conscious episodes but are nevertheless not authoritative, thereby suggesting that Peacocke's treatment stands in need of some limiting principle to characterize just which kinds of conscious thought can provide a rational basis for authoritative self-ascription. Martin suggests that one such class of cases is where knowledge of one's belief issues from one's apparent knowledge of how matters stand in the world. The question therefore arises as to how apparent knowledge of how the world is can provide a reason for the self-ascription of a belief; and Martin argues that the possession-conditions account, unsupplemented, is not sufficient to answer this question.

A contention *ob iter* of Wright's discussion is that the drawbacks of Cartesianism afflict any broadly observational model of self-knowledge. Cynthia Macdonald (Ch. 5) disagrees, defending a qualified observational model which centres on two features of ordinarily observable characteristics that help to explain a subject's direct awareness of them: that they are basic—one does not have to know of any underlying fact in virtue of which they apply when they do—and that it is generally necessary and sufficient for the application of such a characteristic that it seem to a normal observer in normal circumstances that it does apply. Macdonald argues that analogues of these features apply in certain cases of authoritative self-knowledge, specifically— taking perhaps the most apparently unpromising case—in the '*cogito*-type' cases (like 'I am currently thinking that philosophy is difficult') highlighted in Tyler Burge's writings.

Macdonald defends her view by responding to two well-known misgivings about modelling self-knowledge on observation: first, that there is a telling structural disanalogy, since observation normally involves three components, namely, the item perceived, an intermediary, non-conceptual sensation state, and a judgement grounded in that sensation state, whereas self-knowledge of an intentional state seemingly involves analogues of the first and third components only, namely, a first-order (content-bearing) ↳ state and the second-order state which it validates; and second that, whereas the relations between a perceptual state and the item perceived are causal and contingent, those between first- and corresponding second-order intentional states are in general non-contingent, and so not (merely) causal. Each of these concerns, she contends, can be met.

p. 5

The trichotomy—observation, inference, groundlessness—which has tended to structure so much work on these issues was remarked upon above. But of course no one could coherently suppose that self-knowledge was *in general* inferential; that idea would be viciously regressive, since before she can take on board any inference-based claim about her own psychological states, a thinker must first recognize that she believes the premises from which that conclusion may be inferred.[5] So in the basic case the trichotomy must reduce to a dichotomy between two kinds of account, one positing some form of direct cognitive access to their own mental states which subjects enjoy uniquely and which ordinary epistemic practice reflects and is vindicated by, the other viewing the authority accorded to self-ascriptions not as reflecting a commonly known, prior fact of special access, but as being, rather, partly determinative of the identity of the states in question, so that the self-ascription of a particular mental state would be a criterion, in something like the popular Wittgensteinian sense, of the subject's being in that state. Elizabeth Fricker's contribution (Ch. 6) examines this dichotomy— special access accounts versus constitutive accounts—and concludes that, unqualified, it is a false one. First, not just Cartesianism, but a number of quite different kinds of account of the nature of mental states and of our self-knowledge of them, come under the special access heading. One such group comprises functionalist accounts—special access theories which involve non-Cartesian conceptions of the individuation conditions of mental states, and need make no play with any form of 'inner perception'. Fricker is thus in substantial (though not terminological) agreement with both Macdonald and Peacocke in holding that a number of possibilities come under the heading of special access accounts which are not discredited by the demise of Cartesianism.

Second, there is in any case, she argues, theoretical space between the poles—room for intermediate theories according to which 'grammar' and empirical regularities collaborate inextricably to hold our actual mental state concepts together, and play a joint role in explaining the reliability of our basic (non-inferential) self-ascriptions of mental states. Finally, Fricker contends that there is no coherent account of mental concepts which allows the ↳ authority of self-ascriptions to be *wholly* 'grammatical'—wholly a constitutive matter. Rather, all viable concepts of mental states which may competently be non-inferentially self-ascribed must be multi-criterial concepts.

II The Place of Self-Knowledge: Rationality and Agency

Chapters 1–6 are all broadly concerned with issues to do with the explanation of certain distinctive features of self-knowledge. A closely related but distinct set of concerns has to do with the conceptual liaisons of self-knowledge, with the role of our knowledge of our own psychological states in our functioning as rational agents. It is these latter concerns which, though their arguments also bear on the explanatory issues, are central to the contributions of Akeel Bilgrami and Tyler Burge.

Bilgrami (Ch. 7) champions a distinctive and original form of the constitutive view. Specifically, he argues, in a fashion acknowledgedly inspired by Strawson's famous line of reasoning in 'Freedom and Resentment',[6] that self-knowledge is implicated in free, responsible agency in a way which requires the following strong biconditional thesis: wherever responsible agency is operative in thought or action, then the subject has an intentional state that P if and only if she believes that she has it. We are thus committed to treating a subject's self-conception as accurate, and her mental states as essentially salient to her, just by and to the extent that we regard her as a fully responsible agent. It follows, Bilgrami argues, that the connection between first- and second-order attitudes cannot be, in however extended a sense, a causal one. If it were causal, then it might, in a particular case, break down. There would then be a mismatch between such a subject's self-conception and her actual first-order intentional states. But we would never be prepared to countenance such a mismatch: rather, to the extent that we take ourselves to understand a subject's self-ascriptions, we take it that they are expressions of responsible agency, and are thus committed, if Bilgrami's argument is correct, to regarding her first-order states as being in accord with them.

The primary focus of Tyler Burge's chapter (8) is on certain infrequently explored aspects of the role played in thought and action by possession of the concept expressed by the first-person pronoun—the concept of *oneself*. Burge argues that only one who possesses that concept is in a position fully to articulate certain fundamental, a priori aspects ↳ of the concept of reason. A theoretical reasoner recognizes commitments flowing from, and warrants for, particular thoughts. A practical reasoner recognizes how particular sets of attitudes rationalize, require, or clash with particular courses of action. These are forms of rational evaluation. But it is only in cases where the thoughts and attitudes considered come marked as *one's own*—indexed by the first-personal concept, as it were —that such rational evaluation can connect immediately with an implementation—with, say, revising one's thoughts, or selecting one choice of action among alternatives. If a subject scrutinizes a set of attitudes and finds them rationally wanting, that is so far nothing to her—unless they come flagged to her as her own. It is the first-person concept which, by simultaneously designating the agent of appraisal and flagging the attitudes and thoughts appraised, makes the appraisal immediately relevant to changing or maintaining those attitudes.

This line of argument is developed in the first half of Burge's chapter. Though Burge does not himself take it in this direction, it might encourage a version of the constitutive view, since it implicates both authority and salience in a subject's capacity for effective self-criticism as a thinker and agent, and it is arguably a commitment merely of regarding someone as a person that she be deemed an effective self-critic. In the second, more speculative part of his chapter Burge sketches a view of the asymmetries between first- and third-

personal knowledge of mental states in which a basic role is played by a non-empirically warranted sensitivity to the difference between commitments that are one's own and commitments that are not, and canvasses a line of argument therefrom that non-empirically warranted knowledge of other minds is possible via non-empirically warranted understanding of speech. These intriguing suggestions are directions for further work.

III Self-Knowledge and Externalism

The sense that the ordinary conception of self-knowledge makes a philosophical problem out of it has intensified recently with the realization of the apparent tension between immediacy, authority, and salience, on the one hand and, on the other, certain widely accepted arguments that psychological content is *externally*—socially and environmentally—determined. If the contents of our attitudes characteristically depend on relations in which we stand to others in our speech community and to objects and kinds in the material world, then what states of mind we are in at any given point is going to be a function of prevailing empirical ↳ circumstance of a kind to which, in our ordinarily effortless and groundless self-ascriptions, we pay absolutely no attention. Is there here a compelling case for revision of the ordinary conception—should our customary treatment of self-knowledge claims be modified to accommodate externalism? Or should their authority somehow be regarded as extending over the various external conditions of content? Or is there perhaps no genuine tension: can a compatibilist position be made out according to which an ignorance of the external factors that condition mental content can be reconciled with our possession of non-empirical knowledge of the contents of our thoughts?

Paul Boghossian (Ch. 9) advances the case that the tension is genuine: that if the ordinary conception of self-knowledge is retained alongside an externalist conception of mental content, it must follow that certain propositions can be known immediately, without empirical investigation, which everyone agrees can be known only on the basis of empirical investigation. His argument is of the same general type as that offered by Michael McKinsey in his 'Anti-Individualism and Privileged Access', but is distinctive in its focus on the 'empty case'—the case in which a thinker gives expression to what he takes to be a genuine thought concerning a natural kind where there is in fact no relevant natural kind, so that the putative natural kind term in the expression of this thought is actually devoid of reference. In such a case, on an externalist conception of the meaning of natural kind terms, a thinker *delusively* takes himself to have a thought of a certain kind. And such a delusion is not consistent with first-person authority over the nature of one's thought as traditionally conceived.

Brian McLaughlin and Michael Tye reject that conclusion. Their chapter (10) is concerned with versions of content externalism which may be supported by Twin Earth thought experiments alone. They review in detail a leading line of argument for incompatibilism contained in the literature: that, like Boghossian's, which seeks to develop the thought that combination of the traditional conception of self-knowledge with content externalism leads to paradoxical over-extension in the domain of what can be known non-empirically. They argue in detail that this line of argument does not succeed.

Martin Davies's chapter (11) notes an analogy between the latter form of the problem posed by externalism and the structurally parallel problem posed by certain substantive philosophical theses about the requirements for being a thinking being at all, specifically certain requirements on internal cognitive architecture. In both cases an argument is presented of the following form (MC): ↳

1. I have mental property M.

2. If I have mental property M, then I meet condition C.

3. ∴ I meet condition C.

Self-knowledge as traditionally conceived allows a subject to have non-empirical knowledge of appropriate instances of premiss (1); and content externalism, or the considerations which Davies adverts to concerning thinking and its internal cognitive requirements, may be thought respectively to provide a priori support for appropriate instances of premiss (2). Granted only that what is entailed by non-empirically known premisses may itself be non-empirically known, it would follow that the appropriate instances of (3) may also be known non-empirically. And this seems quite wrong for the instances in question. For in the case of externalism, the conclusion will say something about the world outside the subject's head; and in the case of many philosophical theses about the requirements of thought, the conclusion will say something about the unobserved cognitive architecture within his head.

McLaughlin and Tye contend in effect that there are no genuinely problematical instances of the argument form (MC). Davies is unconvinced. He therefore proposes, as an insurance, to invoke principles limiting the transfer of epistemic warrant across known entailments. He notes the precedent of such a limitative principle canvassed in Crispin Wright's 'Facts and Certainty',[7] though some modification is called for to motivate the specific kind of principle needed in the present context. Just how the principle might have to be improved in the face of certain apparent counter-examples is left for further work.

Diana Raffman (Ch. 12) contends that Davies misdiagnoses the trouble with the architecturalist and externalist arguments that are his targets. Whether or not there are independent grounds for principles limiting the transfer of epistemic warrant across known entailments, the primary problem with the two types of argument, in her view, is, rather, that they equivocate. In each case, if premiss (1), that I have mental property M, expresses something about which the subject is non-empirically authoritative, then, Raffman contends, it should be viewed as empty of substantive empirical content; whereas if the second premiss expresses something knowable a priori, then since its consequent involves substantive empirical content, so must the antecedent. In brief: if both premisses, (1) and (2), express something knowable non-empirically, it must follow that there is an ambiguity in what is expressed by 'I have ↳ mental property M', and hence that the *modus ponens* step is flawed by equivocation.

In the last chapter in this group, Jim Edwards (Ch. 13) argues that externalism may issue in a hitherto unnoticed, rather subtler form of compromise of the first-personal transparency of mental contents. This argument depends on two collateral assumptions: John McDowell's contention, developed in his *Mind and World*,[8] that experience is passively structured by concepts, and the so-called Simple Theory of Colour, supported by writers such as John Campbell,[9] according to which those properties which actually are the semantic values of colour concepts are just as they appear to be to normal observers under standard conditions. Edwards develops an example to suggest that it must nevertheless be an epistemic possibility that the semantic values of such concepts, in so far as they are externally determined, are not such properties—not the properties which the Simple Theory takes them to be. If such a possibility were to obtain, there would be a mismatch between the properties which colour concepts present as their semantic values and their actual, externally determined semantic values; and such a mismatch, he crucially contends, would induce incoherence into colour concepts of such a sort that they would be unfitted passively to structure colour experience. Since the difference between this situation and that which the Simple Theory supposes actually to obtain, wherein the semantic values of colour concepts are just as it requires them to be, cannot be salient to a subject just on the basis of the quality of her experience, it follows that it is not transparent to a subject whether her colour experiences are indeed conceptually structured, *à la* McDowell, by concepts answering to the Simple Theory of colour, or accordingly what content is carried by such experience.

IV Knowing What One Means

The chapters mentioned so far are all concerned with our knowledge of our own psychological states. But conscious knowledge of some of the properties of one's own language—for instance, of what particular expressions mean, where their use would be deviant, under what circumstances it would be appropriate to utter them, and the like—is both commonplace and typically arrived at without reflection on evidence, ↳ and thus presents a special case of the issues concerning immediacy, authority, and salience.

This facet of self-knowledge is explored in the contributions by Barry Smith and James Higginbotham. Smith (Ch. 14) argues that any attempt to ground our knowledge of our own minds on knowledge of what we mean when we speak our minds—as in Davidson's famous discussions[10]—must either lead to circularity or leave a lacuna in the explanation. Closing the lacuna will demand an independent account of our first-person knowledge of meaning. Smith proceeds to consider the form which a full explanation of linguistic understanding would take, drawing on first-personal, third-personal, and subpersonal elements of linguistic knowledge, and argues that all these elements must be harnessed to explicate what is made available and expressed by instances of disquotational knowledge, typified by the knowledge that 'Snow is white' means that snow is white. Such expressions, he contends, do express substantial knowledge of what we mean just when they are the upshot of facts about our internal linguistic systems and about the effects these systems have upon our conscious experience of speech and our communicative dealing with others. Thus there is a genuine species of immediate, substantial knowledge of one's own meanings which may be expressed by disquotation, but the authority attaching to whose expression is not (merely) an artefact of disquotation.

Higginbotham (Ch. 15) disagrees, arguing that nothing more than thinly conceived disquotational knowledge is needed to account for first-personal knowledge of one's meanings. However, he distinguishes the kind of authority that attaches to disquotational claims from that which attaches to our intuitive judgements about what we mean. The latter may fall short of genuine knowledge while still involving *entitlement* and a presumption of correctness. Higginbotham assesses each of these species of authority in turn, with respect to both the discussion in Smith's paper and some wider issues, and defends the view that neither is threatened by

externalism. ↳

Notes

1   See their respective contributions, 'Wittgenstein's Later Philosophy of Mind: Sensation, Privacy and Intention' and 'Intentionality and Interiority in Wittgenstein', in K. Puhl (ed.), *Meaning Scepticism* (Berlin and New York: de Gruyter, 1991), 126–47, 148–69.

2   This idea is further elaborated in McDowell's 'One Strand in the Private Language Argument', *Grazer Philosophische Studien*, 33/4 (1989), 287–303.

3   Christopher Peacocke, *A Study of Concepts* (Cambridge, Mass.: MIT Press, 1992).

4   See Sydney Shoemaker, 'Special Access Lies Down with Theory-Theory', *Behavioral and Brain Sciences*, 16 (1993), 78–9 10.1017/S0140525X00029149 .

5   Cf. Paul Boghossian, 'Content and Self-Knowledge', *Philosophical Topics*, 17 (1989), 5–26, at 8–10.

6   P. F. Strawson, 'Freedom and Resentment', *Proceedings of the British Academy*, 48 (1962), 1–25.

7   Crispin Wright, 'Facts and Certainty', *Proceedings of the British Academy*, 71 (1985), 429–72.

8   John McDowell, *Mind and World* (Cambridge, Mass.: Harvard University Press, 1994).

9   See John Campbell, 'A Simple View of Colour', in J. Haldane and C. Wright (eds.), *Reality: Representation and Projection* (New York: Oxford University Press, 1993), 257–68.

10   See Donald Davidson, 'First Person Authority', *Dialectica*, 38 (1984), 101–12 10.1111/j.1746-8361.1984.tb01238.x , and 'Knowing One's Own Mind', *Proceedings and Addresses of the American Philosophical Association*, 60 (1987), 441–58 10.2307/3131782 .