The Authority of Expressive Self-Ascriptions*

A. MINH NGUYEN *Eastern Kentucky University*

ABSTRACT: What explains first-person authority? What explains the presumption that an utterance is true when it is a sincere intelligible determinate first-person singular simple present-tense ascription of intentional state? According to Rockney Jacobsen, self-ascriptions each enjoy a presumption of truth because they are systematically reliable. They are systematically reliable because they are typically both truth-assessable and expressive. Such self-ascriptions, if sincere, are certain to be true. This article presents a defence and a critique of Jacobsen's theory. It is argued that the purported prevalence of expressive self-ascriptions is at best contingent. This contingency cannot explain why self-ascriptions are necessarily authoritative.

RÉSUMÉ : Comment expliquer l'autorité de la première personne? Comment expliquer la présomption qu'un énoncé est vrai lorsque c'est une attribution d'un état intentionnel, sincère et intelligible, à la première personne du singulier de l'indicatif présent. Selon Rockney Jacobsen, les auto-attributions jouissent d'une présomption de vérité parce qu'elles sont systématiquement fiables. Elles sont systématiquement fiables parce qu'elles ont pour caractéristique d'être à la fois expressives et évaluables en termes de vérité. Si de telles auto-attributions sont sincères, elles sont certainement vraies. Cet article présente une défense et une critique de la théorie de Jacobsen. L'auteur soutient que la prétendue prévalence des auto-attributions expressives est au mieux contingente. Cette contingence ne peut pas expliquer pourquoi il est nécessaire que les auto-attributions fassent autorité.

***Dialogue* XLVII (2008), 103-36**
© 2008 Canadian Philosophical Association /Association canadienne de philosophie

https://doi.org/10.1017/S0012217300002407 Published online by Cambridge University Press

1. First-Person Authority Characterized

It is widely assumed that every person enjoys special authority with respect to his own present intentional states. According to this assumption, it is necessary that, for any person, if he sincerely ascribes the presence, or absence, of a particular intentional state to his present self, then there is a legitimate presumption that what he says is true. By contrast, it is not necessary that, for any person, if he sincerely ascribes the presence, or absence, of a particular intentional state to another person, or to his own non-present self, or if he sincerely makes a claim about a certain physical aspect of the external world, then there is a legitimate presumption that what he says is true. To say that, for any sincere intelligible determinate first-person singular simple present-tense ascription of intentional state that any competent speaker makes, there is a presumption that it is true amounts to saying this. Given that a speaker with an adequate command of appropriate concepts sincerely says that he is, or that he is not, in a certain intentional state, his utterance is presumed to be true. One shall treat his utterance as if it were true unless or until one has sufficient evidence or other epistemic grounds to the contrary.[1]

First-person authority is not to be confused with any one of the following discredited versions of so-called privileged access:

(1) *Infallibility*: It is necessary that if a person sincerely ascribes a particular intentional state to his present self, then he is in that state;

(2) *Self-Intimacy*: It is necessary that if a person is in a particular intentional state, then he sincerely ascribes that state to his present self;

(3) *Indubitability*: It is necessary that if a person sincerely ascribes a particular intentional state to his present self, then no one has sufficient grounds for doubting that he is in that state;

(4) *Incorrigibility*: It is necessary that if a person sincerely ascribes a particular intentional state to his present self, then no one else has sufficient grounds for showing that he is not in that state.[2]

First-person authority does not amount to any one of the above. First, it does not amount to infallibility because first-person authority allows for the possibility that some self-ascriptions, though necessarily presumed to be true, are in fact false, but infallibility rules this out on the grounds that such error is impossible. Second, it does not amount to self-intimacy because first-person authority allows for the possibility that one fails to sincerely ascribe to one's present self some of the intentional states

https://doi.org/10.1017/S0012217300002407 Published online by Cambridge University Press

that one is instantiating, but self-intimacy rules this out on the grounds that such innocence is impossible. Third, it does not amount to indubitability because first-person authority allows for the possibility that some self-ascriptions are subject to serious sceptical challenge, but indubitability rules this out on the grounds that such doubt is impossible. Fourth, it does not amount to incorrigibility because first-person authority allows for the possibility that the totality of publicly available evidence in one's possession at the time entitles one not only to question but also to overthrow some self-ascriptions, but incorrigibility rules this out on the grounds that such correction is impossible. First-person authority as currently conceived is thus a very weak form of privileged access, for it is consistent with a modest measure of error, innocence, doubt, and correction concerning one's own present intentional states. First-person authority can obtain even if infallibility, self-intimacy, indubitability, and incorrigibility all fail.

Many philosophers have taken it to be an integral part of first-person authority that we usually apply intentional predicates such as "believe that *p*" and "desire that *q*" to others on the basis of behavioural evidence but to our present selves without benefit of such aid. Some have gone so far as to insist that it is constitutive of first-person authority that, normally, sincere present-tense intentional-state self-ascriptions are not mediated by perception, memory, testimony, reasoning, rational insight, or any other cognitive process. Following Donald Davidson, I do not accept either one of the views just described. Davidson rightly points out that, in general, claims made without epistemic support are neither more probable nor more trustworthy than claims made with such support, and so "it is a strange idea that" claims made without epistemic support should "carry more authority than" and "be favored over" claims made with such support.[3] The ideas that self-ascriptions carry special authority and that self-ascriptions involve no cognitive achievement seem not only independent of but also contrary to each other. That at least one of the ideas is false seems not only possible but also necessary.

What notion of necessity do the typical proponents of first-person authority have in mind when they claim that it is necessary that if a person sincerely ascribes a particular intentional state to his present self then there is a legitimate presumption that what he says is true? There are compelling reasons to believe that the intended notion is one of metaphysical or conceptual necessity. "The existence of first-person authority," writes Davidson, "is not an empirical discovery, but rather a criterion, among others, of what a mental state is."[4] Similarly, Sydney Shoemaker states that "it is of the essence of mind,"[5] "it belongs to [the] very nature [of mentality],"[6] "it is constitutive or definitive of mental states, or of minds, or of the concepts of these"[7]—in other words, it is a matter of metaphysical or conceptual necessity—that each of a person's beliefs, whether verbally

https://doi.org/10.1017/S0012217300002407 Published online by Cambridge University Press

expressed or not, about his own present mental states has a presumption in favour of its truth. Thus,

> at a minimum . . . it is in some sense necessary that our beliefs about our own mental states . . . be for the most part correct, and that a person's belief that she has such a state creates a presumption that she has it, in a sense in which it is not [necessarily] true that someone's having a belief that some *other* person has such a state creates a presumption that that other person does indeed have the state.[8]

Let us call the characterization just offered, with the relevant modality specified, "the standard characterization of first-person authority." As shall be contended in §8, both Crispin Wright and Rockney Jacobsen accept the standard characterization. Jacobsen, whose explanation of first-person authority we shall examine in depth and detail, never explicitly says that he accepts the standard characterization. But he makes his acceptance of it fairly non-opaque because he accepts Wright's formulation of the problem of first-person authority, and Wright formulates the problem by means of the standard characterization.

2. A Philosophical Problem about First-Person Authority

First-person authority raises a philosophical issue, or so it has been argued. According to Davidson, unless first-person authority is adequately explained, it will invite the conceptual problem of other minds,[9] for, if there is no accounting for the fact that sincere present-tense intentional-state self-ascriptions carry authority, whereas other-ascriptions do not, then it is a serious possibility that the intentional states that one ascribes to one's present self and those that one ascribes to others are not the same sort of psychological state. Likewise, if there is no accounting for the fact that intentional predicates are necessarily presumed to be true of a subject when self-applied in the present indicative active, whereas not so when applied to him by another, then it is a serious possibility that these psychological terms have neither the same sense nor the same reference in first-person and in third-person contexts.[10] Considering the insalubrity of the conceptual version of scepticism about other minds, or the disunity that results from bifurcating several of content, attitude, sense, and reference, first-person authority cannot be deemed a brute fact or left unaccounted for. If first-person authority is known or reasonably believed to be unexplainable or unexplained, then the sceptic is surely justified in asking how we know that intentional predicates are univocal or unireferential across first-person and third-person contexts, and how we know that intentional ascriptions uttered in the two contexts pertain to the same subject matter.[11]

https://doi.org/10.1017/S0012217300002407 Published online by Cambridge University Press

Jacobsen conceives the predicament somewhat differently, framing the problem not so much in terms of the threat of scepticism about other minds but in terms of the seemingly more immediate danger of courting absurdity due to an apparent incompatibility between the dispositionality of intentional states and the authority of intentional-state self-ascriptions. Jacobsen writes approvingly thus:

> On Crispin Wright's formulation of the difficulty, intentional states are two-faced in this way: they are disposition-like states, and so the truth of an ascription of, say, belief, whether in the first or third person, is answerable to behavioural manifestations that are as easily knowable to others as to the person of whom the ascription is made. But this aspect of mental states appears to make unsalvageable the widespread thought that all of us enjoy a privileged and reliable authority concerning our own mental contents. If we suppose that we have no introspective access to our own states of belief, then these two aspects of belief must strike us as incompatible.[12]

The Davidson and Jacobsen ways of conceiving the problem are in fact related. Treatments of intentional concepts are usually multipartite and are such that the third-person clauses affirm the dispositionality of intentional states, whereas the first-person present-tense clauses affirm the authority of intentional-state self-ascriptions.[13] If these two aspects of intentional states strike us as incompatible, we shall be inclined to regard the corresponding clauses as incapable of picking out the same sort of psychological state. Jacobsen seems aware of this when he remarks that in contrast to the Cartesian, Rylean, and bifurcationist, the sceptic about mental representations reacts to the purported two-faced character of intentional states by declaring our ordinary concepts of intentional psychology unintelligible on the grounds that it is impossible for any content-bearing state to have both attributes.[14] Conversely, if the absence of an adequate explanation of first-person authority strikes us as providing a hotbed for other-minds scepticism, we shall be inclined to regard the difficulty as stemming from an incompatibility between the authority of intentional-state self-ascriptions and the possession on the part of intentional states of some kind of extrinsic or relational property, whether it be dispositional, causal, or functional. Since discovering such properties, whether in one's own case or in the case of others, requires observation of the subject's behaviour over a span of time, and since behavioural facts properly construed are, in principle, equally accessible to all mature cognitive agents, it is doubtful that dispositional ascriptions in general and intentional-state self-ascriptions in particular are specially authoritative. Davidson seems aware of this when he remarks that intentional states are multitracked dispositions whose manifestations occur over time and whose ascriptions are therefore subject to error, doubt, and correction.[15]

https://doi.org/10.1017/S0012217300002407 Published online by Cambridge University Press

It is unclear how the failure to explain first-person authority is supposed to raise the spectre of scepticism about other minds or how the dispositionality of intentional states is supposed to conflict with the authority of intentional-state self-ascriptions. Davidson's and Jacobsen's arguments can be reformulated as follows:

(1) *Davidson's First Other-Minds Argument*: The intentional states that a subject ascribes to his present self are such that the corresponding ascriptions carry authority. The intentional states that he ascribes to others are such that the corresponding ascriptions do not carry authority. Therefore, the intentional states that a subject ascribes to his present self are not the same sort of psychological state as the intentional states that he ascribes to others.

(2) *Davidson's Second Other-Minds Argument*: The intentional predicates that a subject applies to his present self are necessarily presumed to be true of him. The intentional predicates that another applies to the same subject are not necessarily presumed to be true of him. Therefore, intentional predicates have neither the same sense nor the same reference in first-person and in third-person contexts.

(3) *Jacobsen's Original Incompatibility Argument*: Intentional states are disposition-like states. Ascriptions of disposition-like states, whether in the first person or third person, are neither privileged nor reliable. Therefore, ascriptions of intentional states, whether in the first person or third person, are neither privileged nor reliable.

(4) *Jacobsen's Incompatibility Argument Modified*: Intentional states are such that ascriptions of them in the first-person present indicative active are privileged and reliable. Disposition-like states are such that ascriptions of them, whether in the first person or third person, are neither privileged nor reliable. Therefore, intentional states are not disposition-like states.

It is commonly acknowledged that there is a problem with using epistemic or otherwise intentional properties—such as being known, believed, presumed, justified, warranted, reasonable, authoritative, privileged, and trustworthy—to differentiate one entity from another. To use intentional properties for this sort of purpose amounts to committing what logicians call an "intensional fallacy." The following argument, which parallels those stated above, illustrates the fallacy:

https://doi.org/10.1017/S0012217300002407 Published online by Cambridge University Press

(5) *An Intensionally Fallacious Argument*: Ronaldinho is universally considered the best soccer player of his generation. Ronaldo Assis de Moreira is not universally considered the best soccer player of his generation. Therefore, Ronaldinho is not identical with Ronaldo Assis de Moreira.

Since Davidson's and Jacobsen's arguments fare no better than this patently invalid argument, we have decisive reasons for rejecting them.

Criticism of Davidson and Jacobsen notwithstanding, first-person authority gives rise to a philosophical issue, for the mere fact that a property is capable of being instantiated by a subject does not in general confer a presumption of truth on his sincere and intelligible utterance that he instantiates it. Such fact alone does not in general entitle us to assume in advance that his utterance is true. It is possible that the subject habitually fails to complete the relevant kind of investigation or that the kind of investigation which he manages to complete is inconducive to truth. In either case, we have no valid grounds for taking what he says to be authoritative because what he says is neither reliable nor trustworthy. Furthermore, even if we are actually entitled to presume what the subject says is true, that does not mean, *pace* Tyler Burge, that the "entitlement . . . comes with being a rational agent" or that its existence is a matter of metaphysical or conceptual necessity.[16] It does not belong to the nature or concept of testimony, of a testifier, of a testifiee, or of just any subject matter testified to, that the granted entitlement is in place. It is not metaphysically or conceptually necessary that if a person sincerely says that his eye colour, nose shape, fingerprint pattern, and other polygenic traits are thus and so, then there is an automatic presumption in favour of what he says. As Wright explains,

> the presumptive acceptability of *original* testimony—testimony for which the source is not itself testimony—extends no further than to subject-matters which an informant is deemed competent to know about. . . . I may have such poor colour vision that you rightly come to distrust my testimony on matters of colour. I may, unwittingly, have a very bad memory, and, learning of this, you may rightly come to a state of wholesale suspicion about my testimony on matters of personal recall. But no corresponding wholesale suspicion concerning my attitudinal avowals is possible. . . . Wholesale suspicion about my attitudinal avowals—where it is not a doubt about sincerity or understanding—jars with conceiving of me as an intentional subject at all.[17]

This raises the question: why is the situation any different when the properties to be avowed or self-ascribed are mental, in particular, intentional?[18]

https://doi.org/10.1017/S0012217300002407 Published online by Cambridge University Press

3. Expressivism and Minimalism

What then explains first-person authority? What explains the presumption that an utterance is true when it is a sincere intelligible determinate first-person singular simple present-tense ascription of intentional state? Jacobsen has an answer. He claims to have found in the writings of Ludwig Wittgenstein two apparently opposing strands of thought that not only can be reconciled, but the reconciliation also puts one in a position to account for the systematic reliability and presumptive correctness of self-ascriptions of intentional states.[19] The two strands of thought are expressivism about self-ascription and minimalism about truth.

Expressivism is the view that self-ascriptions are typically expressions rather than descriptions of one's own present intentional states. Self-ascriptions typically express rather than describe the intentional states denoted by the main psychological terms. Despite their expressiveness, self-ascriptions convey information about such intentional states via Gricean mechanisms. But self-ascriptions tend not to assert the information being conveyed. In the typical case, when I say, "I believe that *p*," I do not assert the information that I believe that *p*. I do, however, convey that information and I do assert some information, namely, that *p*. In the language of speech-act theory, self-ascriptions are typically not assertives but expressives. There are exceptions. When we ascribe an intentional state to ourselves on the basis of self-observation or self-reflection (e.g., "I know that I have two hands and I am as certain of it as one can be of anything"), our self-ascription counts as a description. The same holds when we self-ascribe by invoking the testimony of someone who interprets our behaviour or when we self-ascribe without specifying the content of the ascribed intentional state (e.g., "I believe the same as Derrida").[20] Because of the caveats—normally self-ascriptions do not serve as self-descriptions, but sometimes they do; usually their job is not to self-report, but sometimes it is—expressivism cannot be refuted by the following sorts of consideration: it is improbable that after we have inventoried the various expressive functions of self-ascriptions, it will turn out that no such utterances actually have a descriptive role, that none of them really count as fact-stating assertions about our own present intentional states; any progress in understanding self-other asymmetries that we have struggled to forge can only be hampered by an analysis that implies that we cannot think or talk about the very same matters that other persons do when they think or talk about what we believe or desire.[21]

Minimalism is the view that the Disquotational Schema (henceforth DS), ⟨"*p*" is true if and only if *p*⟩, tells us all there is about truth. DS embodies the concept of truth. Our grasp of the concept is therefore exhausted by our understanding of DS or its representative instances. An apparent corollary of minimalism is that a meaningful sentence is

https://doi.org/10.1017/S0012217300002407 Published online by Cambridge University Press

truth-assessable or truth-apt—i.e., apt to be assessed in terms of truth and falsity—as long as it is a suitable substitute for "*p*" in DS.[22] As long as it can serve as the antecedent of a conditional and admits of a negation, it is in the business of being true or false.[23] All that is needed to have a truth-value is to be meaningful and exhibit certain surface syntactic features. Is Noam Chomsky's oft-quoted sentence "Colourless green ideas sleep furiously" truth-assessable? Assuming that this perfectly grammatical sentence makes no sense, the minimalist will answer in the negative, because on the minimalist account the syntactic well-formedness of a declarative sentence by itself does not entail its truth-assessability. Minimalism thus makes two related claims. First, there is nothing more to truth than what is stated or implied by DS. Second, a meaningful sentence is truth-assessable if it can embed in more complex logical constructions such as negation and the conditional. The first constitutes the minimalist account of the nature of truth; the second constitutes the minimalist criteria of truth-assessability.[24]

Of the two claims, Jacobsen wisely subscribes only to the latter. The former he does not need because, as will be apparent shortly, the minimalist criteria of truth-assessability and the expressivist theory of self-ascriptions are all that his explanation of first-person authority makes use of. Another reason Jacobsen keeps his minimalist commitments minimal is this. Following Wright, he asserts that the minimalist criteria of truth-assessability can be derived from the minimalist account of the nature of truth, though the converse does not hold.[25] He maintains, however, that the minimalist criteria of truth-assessability can be supported by other considerations such as those having to do with the surface features of syntax, considerations that Peter Geach famously develops.[26] By arriving at the minimalist criteria of truth-assessability without making a stop at the minimalist account of the nature of truth, Jacobsen hopes to make his explanation of first-person authority appealing to those who may not find the latter attractive.[27]

From a minimalist standpoint, self-ascriptions are more like theoretical claims and observational reports than smiles and winces. To the extent that self-ascriptions are authoritative, they must be presumably true and hence truth-assessable. Theoretical claims and observational reports are truth-assessable whereas smiles and winces are not. From an expressivist standpoint, however, self-ascriptions are more like smiles and winces. Insofar as we speak with authority about our own mental states, it seems that we must be able to say that we occupy such and such mental states without appeal to evidence. Non-appeal to evidence, as well as expressive character, is something that self-ascriptions have in common with smiles and winces. We usually smile and wince without needing or using evidence. By contrast, our theoretical claims and observational reports are neither probable nor trustworthy in the absence of evidential support. Per-

https://doi.org/10.1017/S0012217300002407 Published online by Cambridge University Press

haps self-ascriptions are typically both truth-apt and expressive? For this to be true, minimalism and expressivism must be consistent with each other.

4. Jacobsen's Account of First-Person Authority

But they seem not, for the following reason. Self-ascriptions are suitable substitutes for "*p*" in DS. Assuming minimalism, if anything is a suitable substitute for "*p*" in DS, then it is truth-assessable. So self-ascriptions are truth-assessable. If anything is truth-assessable, then it has assertoric status. So self-ascriptions have assertoric status. If self-ascriptions have assertoric status, then they are assertions by means of which one describes the intentional states denoted by the main psychological terms. So self-ascriptions are assertions by means of which one describes the intentional states denoted by the main psychological terms. Assertions about the existence and character of one's own intentional states are neither identical with nor sufficient for expressions of these very same states. So, contrary to expressivism, self-ascriptions are neither identical with nor sufficient for expressions of the intentional states denoted by the main psychological terms.[28]

Jacobsen correctly observes that the "minimalist" argument against expressivism rests upon a mistake. There is no legitimate basis for attributing to minimalists the claim that if anything is truth-assessable then it has assertoric status. Minimalism holds that every suitable substitute for "*p*" in DS is truth-assessable. It does not require that any truth-assessable expression have assertoric status. Indeed, an utterance of a truth-assessable sentence which means that *p* may not be an act of asserting that *p*. It may be stripped of any assertoric status it might otherwise have had because the context clearly indicates that, by uttering the sentence, the speaker does not seriously represent himself as uttering a sentence he believes true. Such etiolating contexts include the following situations: where a truth-assessable sentence is used as part of a negation, disjunction, conditional, modal claim, or propositional attitude report; where it is employed in play-acting, cracking a joke, teaching a language, giving an example, or practising oratory; and where it is exploited for the performance of a performative, a speech act that is not an assertive, for instance, a directive ("You have to play fair"), a commissive ("I will try my best"), a declarative ("I shall call him 'Minime'"), or an expressive ("I am so excited").[29] In each of these contexts, truth-assessability and assertoric force come apart. In none of them is there any affirmation of the relevant truth-evaluable content.[30]

The upshot of the preceding discussion is that the "minimalist" argument against expressivism fails because of its questionable attribution of a questionable claim to minimalists. As things stand, no good argument exists for the incompatibility between minimalism and expressivism. The

https://doi.org/10.1017/S0012217300002407 Published online by Cambridge University Press

minimalist criteria of truth-assessability do not contradict the expressivist analysis of self-ascriptions. Self-ascriptions, though truth-assessable, are typically non-assertoric because they are already fully employed in the business of expressing the intentional states they only appear to describe. So employed, they are dispossessed of any assertoric property they might otherwise have possessed. Armed with minimalism and expressivism, Jacobsen offers the following explanation of first-person authority. Typically a sincere utterance of "I believe that *p*" expresses the utterer's belief that *p*. Something expresses the utterer's belief that *p* only if the utterer believes that *p*. So typically a sincere utterer of "I believe that *p*" believes that *p*. An utterance of "I believe that *p*" is true if and only if the utterer believes that *p*. So typically a sincere utterance of "I believe that *p*" is true. If typically a sincere utterance of "I believe that *p*" is true, then there is a legitimate presumption that it is true. So there is a legitimate presumption that a sincere utterance of "I believe that *p*" is true. This line of reasoning generalizes to self-ascriptions of other types of intentional state. So first-person authority is fully explained.[31]

5. Jacobsen's Account Defended

The suggestion, then, is that self-ascriptions are systematically reliable and thus each enjoy a presumption of truth because they are typically both expressive and truth-apt. To deny that self-ascriptions are typically both expressive and truth-apt would be to misconstrue their character and raise an insuperable obstacle to a satisfactory explanation of first-person authority. What are we to make of this? It seems that the expressivist account faces a serious problem, for it fails to accommodate the deeply entrenched intuition that, all else being equal, the more authority one possesses regarding one's own mental states, the more knowledge one enjoys regarding their existence and character. All else being equal, a person with more first-person authority possesses more first-person knowledge because authority of this sort has to do with reliability and reliability has to do with justification and knowledge.[32] But instead of embracing the intuition, the expressivist account resolutely opposes it.

It is worth noting that, though related, first-person authority and first-person knowledge are nonetheless distinct. First-person authority implies that, necessarily, any self-ascription that any speaker makes is presumed to be true. This necessity is consistent with the possibility that many self-ascriptions are in fact false and hence insufficient for knowledge. To carry a presumption of truth, especially one that is rebuttable, such as the presumption defining first-person authority, is not to carry truth itself. A statement carrying such a presumption may be false, and demonstrably so. Conversely, first-person knowledge or knowledge of one's own present intentional states does not suffice for special authority with respect to these states. Just because one gets a bit of self-knowledge by making a

https://doi.org/10.1017/S0012217300002407 Published online by Cambridge University Press

(Gettierized justified) true self-ascription, that is not a valid reason to think that the self-ascription in question, let alone every self-ascription, is necessarily presumed to be true. The mere fact that a knowledge claim—or, better, a belief—happens to be true or happens to qualify as knowledge does not entitle us to take its truth for granted. We may know nothing about its alethic or justificative credentials. We may know nothing about its being true, likely to be true, justified, or justifiable. Moreover, if such entitlement were the norm, then knowledge of other minds and knowledge of the external world would entail special authority with respect to these matters, which is absurd.

To show that the expressivist account severs the link between first-person authority and first-person knowledge, consider my counterpart in a slightly different possible world. He makes all the self-ascriptions that I make plus some expressive self-ascriptions not in my repertoire. Assuming that our shared descriptive self-ascriptions are fallible and possessed of the same veridicality rate, my counterpart's self-ascriptions are on average more reliable and so more authoritative than mine because all the additional self-ascriptions that he makes are expressive and so guaranteed to be true if sincere, and this drives up the overall success rate of his self-ascriptions. But it does not follow that he has more self-knowledge. After all, expressive self-ascriptions are not assertions. They are not judgements or any other subjects of epistemic appraisal. As Jacobsen puts it, they "are no more appropriate for epistemic appraisal than moans."[33] Expressive self-ascriptions express no beliefs about mental states. Making them does not enlarge our self-conception, let alone self-knowledge.

We can reach the same conclusion in a somewhat different way. Suppose that my counterpart and I make exactly the same self-ascriptions as far as truth-assessable contents are concerned. The only difference is that, with respect to a number of self-ascriptions, his are expressive whereas mine are descriptive. As a self-ascriber, he makes more self-expressions but fewer self-descriptions than I. The total number of self-ascriptions that we each make, however, is the same. Suppose further that, whether shared or not, our descriptive self-ascriptions are equally fallible. It is clear that my counterpart's self-ascriptions are on average more reliable and so more authoritative than mine, for, with respect to the differentiating self-ascriptions, his are expressive and so guaranteed to be true if sincere, whereas mine are descriptive with no such guarantee. But, again, this does not entail that he has more self-knowledge. On the contrary, it is I who enjoys the advantage. After all, I make more descriptive self-ascriptions. Such ascriptions are assertions expressing judgements or beliefs about my own mental states. Making them enlarges my self-conception and under suitable conditions enhances my self-knowledge.

The expressivist account fails to accommodate the intuition that, all else being equal, the more authority one has regarding one's own mental

https://doi.org/10.1017/S0012217300002407 Published online by Cambridge University Press

states, the more knowledge one enjoys regarding their existence and character. This failure is shown by the fact that, given the expressivist account, a person with more first-person authority may have less, not more, first-person knowledge. Conversely, the expressivist would have us believe that a critically self-reflective person who exemplifies all kinds of epistemic virtue may be more knowledgeable but less authoritative about his own mental states than his epistemically nonchalant albeit self-expressive counterpart. However, this conflicts with the intuition that, all else being equal, the critically self-reflective and epistemically virtuous person ought to be counted as having not only more first-person knowledge but also more first-person authority.

Despite its intuitive plausibility, the preceding criticism is flawed because it betrays a misunderstanding of the expressivist account. In response, the expressivist happily concedes the point. After all, he does not treat first-person authority as an epistemic privilege or achievement. He denies that any cognitive capital is needed or spent to secure a fit between sincerity and truth in our normal self-ascriptions. All self-ascribing subjects, including those who exercise little or no epistemic care, have the ability to self-ascribe mental states correctly without needing or using evidence even when evidence is available. In light of this, it is only natural for the expressivist to urge that we withhold the accolade of knowledge from such effortlessly earned truths.[34] A disconnect between first-person authority and first-person knowledge is to be expected if the authority of self-ascriptions is due not to their epistemic prowess but to their expressive character. First-person authority is not an epistemic or cognitive notion. It neither derives from epistemic acumen nor co-varies with self-knowledge. The criticism levelled at the expressivist account is baseless, for it rests on an apparently intuitive assumption which the expressivist already means to cast aside. If this is correct, then insisting that first-person authority has nothing to do with epistemic privilege or achievement also enables the expressivist to counter a similar objection raised by Wright.[35]

6. The Problem with Jacobsen's Account

Let us now turn to the real objection. The real objection is that the typicality of expressive self-ascriptions that Jacobsen alleges is at best contingent. Such contingency may explain why self-ascriptions are authoritative as a matter of fact, but it cannot explain why authoritativeness is a necessary feature of self-ascriptions. In default of this further explanation, first-person authority remains unexplained. There is no denying that some human speakers occasionally use some psychological verbs, including those of the intentional variety, in the first-person singular simple present-tense indicative active just to express the mental states denoted by the verbs. As James Urmson rightly notes in an article aptly entitled "Paren-

https://doi.org/10.1017/S0012217300002407 Published online by Cambridge University Press

thetical Verbs," some people sometimes employ epistemic constructions such as "I know (believe, guess, suppose, suspect, estimate, etc.) that . . . ," not to claim that they know (believe, guess, suppose, suspect, estimate, etc.) that things are thus and so, but to state in essence that the nested statement is true, as well as "to indicate the evidential situation in which the statement is made . . . and hence to signal what degree of reliability is claimed for, and should be accorded to, the statement to which they are conjoined."[36]

Consider a typical case of efficient communication. You ask, "Where is the cat?" I answer, "I believe the cat is on the mat." Assuming that my response is sincere and cooperative, I do not utter this intending that you should believe that I believe that I believe that the cat is on the mat. By making the utterance, I do not mean that I believe that the cat is on the mat. I do not intend my utterance to be a self-report or self-description. The accepted purpose or direction of the talk exchange in which you and I are engaged requires me at this stage to name the external position of the cat. Talking instead about the existence and character of an internal state of mine, even if it is a doxastic state about the cat's location, only renders my conversational contribution less relevant and informative than is required. Given the key elements that sustain the conversational context, my response is intended as a hesitant and suggestive way of making the contained assertion and expressing the affiliated belief that the cat is on the mat. I assert this while suggesting that even though it is not true that I just plump for one alternative rather than another without any grounds, which I would do if I answered, "I guess the cat is on the mat," neither is it true that I have all the evidence that I need to guarantee the truth or likelihood of the embedded statement so that the statement should be accepted, if at all, with the right amount of discretion.

We have analyzed the case as one in which I mean both that the cat is on the mat, therein expressing my belief that the cat is on the mat, and that my belief that the cat is on the mat falls considerably short of certainty. To be sure, I do not say such things because the sentence that I utter does not conventionally mean such things. I only mean in the sense of suggesting, implying or, to use a neologism coined by Paul Grice, "implicating" that both of these states of affairs obtain. That the cat is on the mat and that my belief that the cat is on the mat falls considerably short of certainty are my "implicatures."[37] A relevant alternative to our analysis treats the case under consideration as one in which I assert that I believe that the cat is on the mat, therein expressing my belief that I believe that the cat is on the mat, and implicate both that the cat is on the mat and that the belief which I report myself as having falls considerably short of certainty. The first analysis is more accurate, however, because the second analysis would have me make in addition a self-assertion when the current purpose or direction of the talk exchange between us requires me to make only an

https://doi.org/10.1017/S0012217300002407 Published online by Cambridge University Press

assertion about the cat and qualify it if necessary. In effect, it would have me, a competent party to the conversation, be both more informative—hence violating Grice's Maxim of Quantity—and less relevant—hence violating Grice's Maxim of Relation—than is required.

Some might argue that the second analysis is nonetheless more accurate because it explains error better than the first analysis. Suppose the cat is not on the mat. Can you reasonably complain that I asserted that the cat was on the mat even though it was not there? Or can I sensibly respond by saying, "I only said that that was what I thought; I did not know the cat's location and so did not assert what you are claiming I did"? If "only knowledge warrants assertion," as Timothy Williamson proposes, and the assertion that I made was warranted, then what I asserted could not be that the cat was on the mat because I did not know this to be true. The second analysis is more precise, providing my assertion conformed to the fundamental constitutive rule of assertion that one should assert that *p* only if one knows that *p*.[38] There are several reasons we must resist this line of thinking. To begin with, the preceding reply in support of the second analysis denies that the relevant error has been committed rather than explaining its existence. Second, if the first analysis entails that the relevant error has been committed, then so does the second analysis because both state that I implicated that the cat was on the mat. Third, while it is legalistic and defensive to respond to the complaint by saying, "I only said that that was what I thought," it is more factual and honest to admit in light of the existing features of the conversational context that what I implicated was false. Finally, even if Gricean implication is a form of assertion, knowledge is the norm of assertion, and my implicature was false and unwarranted, the assertion that I made to the same effect could still be reasonable.

Concerning this last point, Williamson writes:

> The account can be roughly summarized in the slogan "Only knowledge warrants assertion." "Warrant" is used here as a term of art, . . . This use need not correspond exactly to that of "warrant" in everyday English. It is not denied that false assertions are sometimes warranted in the everyday sense that they are sometimes reasonable; the claim is rather that the reasonableness of such assertions is explicable as the joint outcome of the knowledge rule and cognitive considerations not specific to assertion.[39]

Williamson goes on to articulate such considerations. For instance, the urgency of the situation may require one to assert that *p* even though one knows that one does not know that *p*. Thus, upon seeing a train come into the station, I shout, "Hurry, hurry, that is your train," knowing that I do not know that it is, because I believe on the basis of evidence that it probably is, that you have only moments to catch it, and that it is urgent for

https://doi.org/10.1017/S0012217300002407 Published online by Cambridge University Press

you to reach your destination.[40] It is also reasonable for one to assert that *p*, even though it is not true that *p* and therefore one does not know that *p*, provided one reasonably believes that *p* and one reasonably believes that one knows that *p*. Thus, it is reasonable for me to assert that there is snow outside when "it is winter, and it looks exactly as it would if there were snow outside, but in fact that white stuff is not snow but foam put there by a film crew of whose existence I have no idea."[41] If it is reasonable for me to believe not only that *p* but also that I know that *p*, then, since it is reasonable for me to assume that knowledge warrants assertion, it is reasonable for me to believe that I have warrant to assert that *p*. If it is reasonable for me to believe that I have warrant to assert that *p*, then, *ceteris paribus*, it is reasonable for me to assert that *p*. Hence, it is reasonable for me to assert that there is snow outside even though the assertion is both false and unwarranted.[42]

But the fact that some or all mature human speakers occasionally use some or all intentional predicates expressively does not mean that they all normally do. Moreover, even if they all normally do, that does not mean that it should rank as a metaphysically or conceptually necessary truth. Consider some possible world in which full-fledged self-ascribing subjects are exactly like mature human speakers except that they are simply incapable of uttering "I believe (desire, hope, fear, etc.) that *p*" with the intention of expressing, as opposed to describing, their own present belief (desire, hope, fear, etc.) that *p*. I use "simply incapable" for clear contrast; "mostly incapable," with its advantage of making our story less fanciful, will do as well because the expressivist proposal is doomed if expressive self-ascriptions are anything but typical. For ease of expression, I shall call the possible world under consideration and its inhabitants "D-world" and "D-worlders," respectively.

D-worlders never express their propositional attitudes by saying that they have them. They do not get to possess, let alone exercise, this mode of competence. Their utterance of "I believe (desire, hope, fear, etc.) that *p*" is invariably assertoric. It is invariably achieved in a matter-of-fact tone, accompanied by assertive gestures, and so forth. Although D-worlders inherit an incapacity for expressing their relevant intentional states by means of self-ascriptions, they are nonetheless capable of expressing these states by standard alternative means such as intonation, body posture, hand gestures, gaze patterns, facial expressions, and less explicit linguistic signals. For instance, a D-worlder can use "*p*" to express his belief that *p*, and use "I believe that *p*" to express his belief that he believes that *p*, though he cannot use the latter to express his belief that *p*. The fact that he is a self-ascriber with a capacity for some form of linguistic self-expression does not imply that his self-ascriptions are ever expressive, or largely expressive, in Jacobsen's sense. We can make D-world even more credible by stipulating that its inhabitants "have what

https://doi.org/10.1017/S0012217300002407 Published online by Cambridge University Press

they regard as important ideological reasons for behaving as they do, and they have, through years of training, learned to live up to their own exacting standards . . . and after millions of years they begin to have children who are born fully acculturated."[43]

To guard against misunderstanding, I should point out that I am not committed to the claim that D-worlders are purely Rylean subjects who come to know their own mental states only in the same way as others come to know them. First of all, the thought experiment about D-world is not meant to support the behaviourist approach to the epistemology of self-knowledge, but only to highlight and make coherent the possibility that self-ascriptions typically have a different illocutionary force from what the expressivist claims for them. That D-worlders enjoy a non-inferential and perhaps even privileged access to their own mental states is consistent with the fact that they use, and choose to use, self-ascriptions to express their higher-order beliefs, some of which they acquire through the access, rather than the mental states which those beliefs are about. Second, even if self-ascriptions in D-world are all introspective self-reports, and even if introspection is a cognitive process or faculty, it does not have to be construed as retrospection *à la* Gilbert Ryle.[44] Among contemporary philosophers and psychologists, there is a determinate and settled consensus that while the title of being the true account of introspection is still up for grabs, the retrospectionist account is not a serious contender. It is broadly accepted that introspection is not a matter of examining our present and past behaviour, discerning the patterns in that behaviour, and ascribing apposite mental states to ourselves—hence the *quasi* truism that introspection is not the making of inductive inferences about our own mental states on the basis of sensory observation of what we overtly say or do. Besides, there are arguably better epistemic accounts of introspection, such as the classical inner-sense model and the contemporary brain-scanner theory.[45]

There is no contradiction in the scenario just described. D-worlders' lifelong incapacity for making expressive self-ascriptions has a compelling genetic and ideological explanation. Nor is there any change of meaning in any intentional predicate. D-worlders believe and desire just as we do. Their attitudes are multitracked dispositions that manifest themselves over a span of time just as ours are. The only difference is that their capacity for linguistic self-expression is on average a little bit less versatile than ours. Despite the difference, a normal D-worlder is still better at linguistically expressing himself than an otherwise normal human with a damaged language organ. But we do not in general regard such a human as incapable of first-order or higher-order thought. So the difference makes no difference as far as meaning is concerned. The above scenario also does not entail any untestable hypothesis. Academic scepticism about other minds aside, if there were D-worlders, we would in principle be able to detect them and distinguish them from ourselves and from crea-

https://doi.org/10.1017/S0012217300002407 Published online by Cambridge University Press

tures that do not self-ascribe psychological states, for we can in principle determine whether a physically instantiated string of signs or sequence of sounds is a self-ascription and whether a self-ascription plays a descriptive or expressive role. Expressive self-ascriptions are expressions of the intentional states denoted by the main psychological predicates. Descriptive self-ascriptions are assertions, and thus expressions of higher-order beliefs, to the effect that the self-ascriber occupies the intentional states in question. Because intentional states are dispositional states, the truth of an ascription of any of them—whether first-order or higher-order, doxastic or otherwise—is answerable to behavioural manifestations that are publicly observable. But if the above scenario does not involve any logical inconsistency, conceptual incoherence, semantical change, or untestable hypothesis, then D-world is possible. Since self-ascriptions in D-world are not expressive, expressive character cannot explain why they carry authority.

7. Objections and Replies

There are at least five general lines of response to this objection. None of them proves effective. The first line of response simply denies that D-world is possible. The denial seems unwarranted because D-world is described in adequate detail and the description involves no logical, conceptual, semantical, or verificational infelicities. The fact that a situation is coherently conceived and intelligibly described constitutes at least *prima facie* evidence that it is possible. We are entitled to accept the possibility of D-world unless presented with strong evidence to the contrary. The best option at this point for the expressivist is to challenge the claim that the intentional predicates employed by D-worlders in their predominantly descriptive self-ascriptions mean the same as our intentional predicates. It is a moot question, the expressivist might urge, whether such predicates are even psychological predicates at all. Quoting Wittgenstein with approval, Jacobsen explains as follows:

> [I]f our self-ascriptions did not combine the appearance of descriptions with an expressive role, then psychological concepts would lack their distinctive character. . . . [T]he distinctive character of psychological terms is traced to their expressive uses. . . . Terms lacking an expressive use would not be *psychological* terms at all, since "Not until it finds its particular use in the first person does it acquire the meaning of mental activity."[46]

D-worlders are supposed to use intentional predicates with the same meaning as they have in the actual world. But if the expressivist is right, the first-person present-tense indicative active of an intentional verb does not have any psychological reference and so the actual meaning of, say, "believe" when used in a sentence of the form "I believe that *p*" does not

https://doi.org/10.1017/S0012217300002407 Published online by Cambridge University Press

include any such reference as a part. Instead, the meaning of an expression is exhausted by its use in a normal context of communication and in such a context the first-person present-tense indicative active of "believe" serves to present the embedded statement, convey the lightness of the evidence for the statement, and suggest that it be accorded only a modest measure of reliability. The expressivist thus stands resolute against the notion that D-worlders use "believe" with its actual meaning, for D-worlders always use this verb to designate a certain cognitive attitude and use a sentence of the form "I believe that *p*" to report that they harbour the attitude towards the proposition in question. Intentional predicates do not retain their meanings when they lose their expressive roles. Nor do they retain their status as psychological predicates because having an expressive role is necessary for a predicate to be a psychological predicate. Deprived of any expressive role in D-world, "believe," "desire," and other intentional predicates must have their semantics radically altered.

Aside from a question-begging appeal to the expressivist hypothesis ("if the expressivist is right"), this response faces a couple of problems. As noted in §6, since the expressivist scheme is unworkable if expressive self-ascriptions are anything but normal, there is no need, except for the purpose of providing a dramatic contrast, to stipulate that D-worlders are utterly incapable of using intentional predicates expressively. D-worlders may use intentional predicates expressively as often as they please provided that such expressive employments are not typical. But the main problem with the expressivist response is that if it were to succeed intentional predicates would be neither univocal nor unireferential across first-person present-tense contexts, on the one hand, and other-person or other-tense contexts, on the other. The expressivist would have us believe that self-ascribers use intentional predicates differently in the contrasting contexts and that intentional predicates possess different semantic values in the contrasting contexts given the Wittgensteinian conception of meaning as use. It is not difficult to see, however, that this kind of multivocality and multireferentiality of intentional predicates leads to a pernicious form of bifurcationism about psychological meaning and reference and threatens the conceptual version of scepticism about other minds as a result. Moran explains as follows:

> A good account of the conceptual structure of commonsense psychology should . . . preserve the assumption that terms like "belief" and "desire" have the same meaning when used in first-person and in third-person contexts. . . . Without this assumption of the univocality of psychological terms, we could not say that two people could be talking about the same thing, or even the same *sort* of thing, when, as we say, they talk together about the beliefs of one of them. . . . The denial of univocality would . . . be a form of solipsism.[47]

https://doi.org/10.1017/S0012217300002407 Published online by Cambridge University Press

The expressivist is thus not justified in contesting the claim that the intentional predicates employed by D-worlders mean the same as those employed by actual self-ascribers such as ourselves. What the expressivist describes as a case of change of meaning will be described more correctly, or at least less contentiously, as a case of change of force. Predominantly expressive, or so it seems, in the actual world, self-ascriptions are predominantly assertive in D-world. Their typical force goes from being a matter of expressing and manifesting to a matter of asserting and describing. All this happens without meaning change in any psychological term or truth-value change in any psychological sentence. We use a sentence of the form "I believe (desire, hope, fear, etc.) that *p*" to express, then to describe, then back to express, then back to describe our own present belief (desire, hope, fear, etc.) all the time. No change of meaning or truth-value occurs but only a change of force. Admittedly, the kind of piecemeal change observed in the actual world is much more pedestrian and less profound than the kind of wholesale change involved in the case of D-world. But the difference is one of degree rather than one of kind. Jacobsen ought to be agreeable to our description and analysis of D-world, for he explicitly invokes the extensively embraced distinction between truth-conditional meaning and illocutionary force to dispel a persistent myth about the incompatibility between minimalism and expressivism and account for the semantics and pragmatics of first-person present-tense psychological utterances.[48]

This concludes the discussion of the first line of response, which denies that D-world is possible. The second line of response concedes that D-world is possible but denies that the expressivist account fails to accommodate the possibility. The denial again seems unwarranted, for expressive self-ascriptions are atypical in D-world and the expressivist lacks the conceptual wherewithal to explain the authority of those self-ascriptions that are not expressive because for him such self-ascriptions count as self-descriptions and hence carry no more authority than descriptions of the physical world or other minds.[49] The expressivist claims to have traced the source of first-person authority to the generally expressive employment of self-ascriptions. According to this claim, self-ascriptions are authoritative because they are typically employed to serve expressive functions. But if expressive employments are non-existent or rare, as is the case in D-world, then it is a mystery how all self-ascriptions get to be authoritative.

The third line of response concedes that the expressivist account fails to explain the authority of D-worlders' self-ascriptions but denies that the failure has any bearing on whether it succeeds in explaining the authority of our own self-ascriptions. Now if we are concerned with only the authority that every mature human subject enjoys with respect to his own present intentional states, then of course D-world is importantly different from the actual world, thus effectively irrelevant, and so properly ignored. But

https://doi.org/10.1017/S0012217300002407 Published online by Cambridge University Press

the range of authoritative self-ascriptions is much wider. According to the standard characterization of first-person authority, which Jacobsen seems to accept, all self-ascriptions carry authority, regardless of whether they come from human speakers or not. Consequently, insofar as the authority of self-ascriptions realized in D-world or any other world is not explained, neither is the authority of self-ascriptions full stop. The third response is thus inadequate, for it underrates the range of authoritative self-ascriptions and so understates the challenge of providing a correct and complete account of first-person authority.

The fourth line of response concedes that the expressivist account fails to explain the authority of D-worlders' self-ascriptions but denies that the account cannot be supplemented by one capable of dealing with it. The optimism seems undercut, however, by the expressivist's conviction that descriptive self-ascriptions function as self-reports and enjoy no more epistemic privilege than reports concerning other subject matters. Furthermore, assuming that such a supplement is available, do we really need the expressivist account, with its tendentious claims of minimalism and expressivism, to explain first-person authority? Would it not be simpler and more economical to explain first-person authority by rejecting wholesale expressivism, acknowledging the contingency of expressive self-ascriptions, regarding these as self-ascriptions *manqué*, and explaining the authority of full-blooded (i.e., descriptive) self-ascriptions in the manner of the so-called supplement? So-called because the "supplement" is now the main thing. Simplicity and economy are no theoretical virtues if one discounts the truth. But the truth is that expressivism does not hold when D-world is possible. The expressivist account is either redundant or worse.

The fifth line of response concedes that the expressivist account cannot deliver without shedding a key doctrine about the nature and function of descriptive self-ascriptions or else becoming superfluous, but denies that it thereby cannot play any crucial role in explaining first-person authority. All we have to do is accept the possibility of D-world without the authority of D-worlders. The reasoning goes something like this. The expressivist account is probably the only one that can explain the actual cases of first-person authority. If it cannot explain certain alleged cases of first-person authority, then one just has to read the writing on the wall and conclude that one's allegation and attendant impression about the range of authoritative self-ascriptions are mistaken. D-worlders *qua* authoritative self-ascribers are not only fictitious but also impossible. The range of authoritative self-ascriptions ought to be constricted. Unfortunately, this kind of argument is not likely to impress proponents of first-person authority. To begin with, the inference is problematic. Just because one cannot explain something does not mean that it does not exist. The fallacy committed is akin to the fallacy of ignorance, with failure or inability to explain replac-

https://doi.org/10.1017/S0012217300002407 Published online by Cambridge University Press

ing failure or inability to know. Second, the argument relies on the assumption that either the expressivist account has no rivals or they all are inferior. The first disjunct is patently false. As for the second, it may well be true. Without detailed critical engagement, however, proponents of first-person authority are duly bound to remain dubious.[50]

8. The Standard Characterization of First-Person Authority

I want to dislodge a difficulty that arises in connection with my critique of the third and fifth lines of response. In seeming contradiction to what I claim above, Jacobsen never explicitly says that he accepts the standard characterization of first-person authority. This characterization states that insincerity and lack of understanding aside, all self-ascriptions carry authority in the sense of enjoying a rebuttable presumption of truth, regardless of whether or not they come from human speakers. But Jacobsen accepts Wright's formulation of the problem of first-person authority, and Wright's characterization of first-person authority is the standard characterization, the one developed in depth and detail at the beginning of this article.

Wright formulates the problem of first-person authority as follows:

> [H]ow is it possible to be effortlessly, non-inferentially, and generally reliably authoritative about psychological states which have no distinctive occurrent phenomenology and which have to answer, after the fashion of dispositions, to what one says and does in situations so far unconsidered?[51]

Wright's crisp formulation is echoed in the following passage from Jacobsen:

> The problem takes . . . the form of a dilemma. Any adequate, non-revisionary account of our concept of belief must be crafted to accommodate both (1) the status of ascriptions of intentional states as answerable for their truth to their overt behavioural manifestations and to their inferential relations to other such states, what Wright calls their "theoreticity"; and (2) the apparently effortless and reliable authority of such ascriptions in the first person present indicative. . . . But efforts to illuminate the one face of belief invariably leave the other face in shadow.[52]

Wright and Jacobsen hold that it belongs to the nature of intentional states, or to the nature of the concepts of these, that they are both disposition-like and avowable. That they have both characteristics is constitutive of intentional states, or definitive of intentional concepts, in the sense that it is an essential part of what makes these intentional entities intentional. But first-person avowability seems to conflict with theoreticity.[53] This prompts Wright and Jacobsen to ask how it is possible for intentional

https://doi.org/10.1017/S0012217300002407 Published online by Cambridge University Press

states to have both characteristics. Are our ordinary concepts of intentional psychology unintelligible?

First-person avowability is inextricably linked with first-person authority. The link consists in the fact that intentional states are avowable in the first-person present-tense indicative active if and only if ascriptions of intentional states in the first-person present-tense indicative active are authoritative. Now the conception of first-person authority that informs Wright's work on the epistemology of self-knowledge is as follows:

> [T]he authority standardly granted to a subject's own beliefs, or expressed avowals, about his intentional states is a *constitutive principle*: something that . . . enters primitively into the conditions of identification of what a subject believes, hopes, and intends. The language game of intentional states, so conceived, would function roughly like this. A subject's sincere dispositions of avowal—or, better, his beliefs—concerning his intentional states would stand by default, unless there was positive reason to reject them."[54]

A similar characterization occurs in the following passage:

> [The] no-self-deception condition is *positive-presumptive*. By that I mean that, such is the "grammar" of ascriptions of intention, one is entitled to assume that a subject is *not* materially self-deceived, or unmotivatedly similarly afflicted, unless one possesses determinate evidence to the contrary. Positive-presumptiveness ensures that, in all circumstances in which one has no countervailing evidence, one is *a priori* justified in holding that the no-self-deception condition is satisfied.[55]

Still another variation on the same theme, albeit one which appears after the publication of Jacobsen's Wittgenstein article, is this:

> What distinguishes the presumptive acceptability of attitudinal avowals . . . is that the authority which attaches to them is, in a certain sense, *inalienable*. There is no such thing as showing oneself chronically unreliable in relation to the distinctive subject-matter of attitudinal avowals. . . . [No] wholesale suspicion concerning my attitudinal avowals is possible. You may not suppose me sincere and comprehending, yet chronically unreliable, about what I hope, believe, fear, and intend. Wholesale suspicion about my attitudinal avowals—where it is not a doubt about sincerity or understanding—jars with conceiving of me as an intentional subject at all.[56]

Wright advances three contentions in the preceding quotation montage. First, the authority granted to a subject's self-ascriptions is a constitutive principle, something that enters primitively into the conditions of identification of what a subject believes, desires, hopes, and fears. Second, the

https://doi.org/10.1017/S0012217300002407 Published online by Cambridge University Press

language game of intentional states and the logical grammar of self-ascriptions function in such a way that, in all circumstances in which one has no countervailing evidence, one is *a priori* justified in holding that the subject is free of any condition which might somehow impede his ability to reliably certify his own present intentional states. Third, the authority which attaches to attitudinal avowals is inalienable in the sense that wholesale suspicion about a subject's attitudinal avowals, where it is not a doubt about sincerity or understanding, jars with conceiving of him as an intentional subject at all. Given this trio of contentions, the existence of first-person authority is a deeply non-contingent fact.[57] If the authority accorded self-ascriptions but denied other-ascriptions enters into the conditions of identification of intentional states, if the logical grammar of self-ascriptions not only defeasibly immunizes self-ascriptions against error but also plays an indispensable role in governing the language game of intentional states, and if extensive scepticism about a subject's self-ascriptions clashes with conceiving of him as an intentional subject, then it is constitutive of intentional states and subjects, or definitive of intentional predicates and concepts, that every sincere present-tense intentional-state self-ascription carries authority. As a matter of metaphysical or conceptual necessity, every such self-ascription enjoys a rebuttable presumption of truth. Whether it be known as defeasible authority, presumptive acceptability, or positive presumptiveness, truth as a default position is an essential or analytic feature of every such self-ascription.

9. Authoritative Self-Ascriptions in D-World

As just contended, Wright's characterization of first-person authority is the standard characterization. Since Jacobsen accepts Wright's formulation of the problem of first-person authority and Wright's formulation is guided by the standard characterization, it is reasonable to conclude that Jacobsen accepts the standard characterization of first-person authority as well. But why should Jacobsen accept that every possible self-ascription carries authority in the sense of enjoying a rebuttable presumption of truth? It seems generally recognized that not all self-ascriptions are authoritative in this sense. Convinced by a psychoanalyst that he harbours a deeply repressed desire to drink a cup of mud, an analysand may be ready to self-ascribe it. But there would be no presumption that his self-ascription is true. It would not be the case that what he says goes and others have to defer to him. So which self-ascriptions do carry authority? Here the expressivist argues that only expressive self-ascriptions come with this presumption of truth. Descriptive self-ascriptions issue from the same cognitive processes by which we form beliefs and judgements and make assertions and claims about the rest of the world. There is no presumption that our descriptions of our own intentional states are true, just as there is no presumption that our descriptions of the rest of the world

https://doi.org/10.1017/S0012217300002407 Published online by Cambridge University Press

are true. Granted, we often have better cognitive access to our own intentional states than the psychoanalytic patient and doctor do in relation to the former's unconscious desires and conflicts. But we also often have better cognitive access to our own immediate environment than others do. Contrary to my critique of the fifth line of response, the expressivist contends that people in D-world have no more authority over the contents of their own minds than over the contents of their own closets.

The existence of unconscious intentional states and descriptive attitudinal self-ascriptions does not threaten first-person authority as standardly conceived. It is not compulsory to deny that all propositional attitudes fall within the scope of first-person authority or that all self-ascriptions come with a presumption of truth. Following Davidson, I submit that insofar as psychological concepts are extended to include the unconscious, a *bona fide* psychoanalytic patient does lose authority over some of his intentional states. Indeed, it is loss of authority that gives unconscious mental states their distinctiveness. As expected, all instances of propositional attitudes that are not unconscious, whether they be conscious or preconscious, remain as subject as ever to first-person authority. But more noteworthy is the fact that

> in psychoanalytic practice, recovery of authority over an attitude is often considered the only solid evidence that the attitude was there before being noninferentially appreciated by its holder. Thus those cases of unconscious mental states that were unsystematically recognized to exist before Freud are indirectly included in the scope of first person authority by psychoanalysis.[58]

In view of this, a person loses authority over unconscious mental states only in the sense that he is utterly unable or deeply disinclined to self-ascribe them, or if he manages to do so by way of the talking cure, the presumption of truth that accompanies his self-ascriptions is poised to be reversed. Neither case suffices to confute the standard characterization of first-person authority. As argued in §1, first-person authority is compatible with a modest measure of innocence, error, doubt, and correction concerning one's own present intentional states.[59] Failures of self-cognition come from many sources, including the exercise of such psychological defence mechanisms as repression, projection, rationalization, and intellectualization.[60]

In addition to conflating the case in which a presumption of truth simply does not exist and the case in which a presumption of truth does exist but is soundly rebutted, the expressivist argument against granting first-person authority to inhabitants of D-world depends on the tendentious stipulation that no descriptive self-ascriptions be authoritative. Since the stipulation is implausible, the argument is implausible. To see the implausibility of the stipulation, recall that first-person authority is characterized in terms of a presumption rule according to which, given that a person sin-

https://doi.org/10.1017/S0012217300002407 Published online by Cambridge University Press

cerely ascribes a particular intentional state to his present self, we shall proceed as if his utterance were true unless or until we have sufficient reason to believe otherwise. In other words, given that a person makes a sincere intelligible determinate self-ascription, there is a presumption that the self-ascription is true. The rule sanctions the passage from a self-ascription's being made to the self-ascription's being true, with the passage blocked only if or only when we have sufficient reason to believe that the self-ascription is false. Note that when the passage is blocked in a given case, what is rebutted is not the general presumption rule that defines first-person authority but rather the specific presumption of truth that attaches to such and such self-ascription. This is analogous to a certain aspect of the current legal practice in many Western countries. The practice is such that if a jury has sufficient reason to believe that so-and-so is not innocent, the presumption that so-and-so is innocent is rebutted. What is not rebutted is something universally known simply as "the presumption of innocence," which, strictly speaking, is a presumption rule to the effect that, given that a person is charged with a crime, jury members shall proceed as if the person so charged were innocent unless or until they have sufficient reason to believe that he is guilty. In order to invalidate the presumption rule just described, one has to do much more than just rebut the presumption of innocence enjoyed by this or that defendant at trial. Likewise, in order to invalidate the presumption rule defining first-person authority, one has to do much more than just rebut the presumption of truth that attaches to this or that self-ascription, no matter whether the self-ascription is descriptive.

In response, the expressivist might agree that this last observation is correct. But he might also point out that all self-ascriptions in D-world are *ex hypothesi* descriptive. Since the presumption of truth that supposedly accompanies every descriptive self-ascription cannot help being overridden, this in effect means that no self-ascriptions in D-world are really authoritative. No persons in D-world really enjoy first-person authority. Appeal to D-world is a pointless distraction because it does nothing to undermine the expressivist strategy. Now this response ought not to be decisive, for it is predicated on the dubious assumption that all it takes to nullify the presumption of truth that accompanies descriptive self-ascriptions are facts about their descriptive character. The assumption is dubious because it is not facts *simpliciter* but rather established facts or, more precisely, genuine evidence about facts within one's ken that rebuts a presumption. It takes the production, appreciation, and submission of countervailing evidence on one's part, or at the very least the possession of such evidence, to bring about the rebuttal. Otherwise, plain old objective fact that things are not thus and so would by itself rebut the presumption that things are thus and so.[61] Of course, the mere fact that a defendant is guilty or suspicious constitutes only the existence of reasons for believing that he is not innocent. But it is the jury members' having those rea-

https://doi.org/10.1017/S0012217300002407 Published online by Cambridge University Press

sons that counts so far as defeating the presumption of his innocence goes. Similarly, the mere fact that a self-ascription is false or descriptive constitutes at best the existence of grounds for disbelieving the self-ascription. But it is the deliberators' possessing those grounds that matters as far as overcoming a presumption in favour of its truth is concerned. Because it takes the possession of countervailing evidence to bring about the rebuttal of a presumption, and because the acquisition of such evidence requires a non-trivial amount of cognitive energy and effort, the expressivist is not entitled to assume that the presumption of truth that attaches to every descriptive self-ascription cannot but be rebutted.

Just as it takes the possession of countervailing evidence to bring about the rebuttal of a presumption, so too it requires the possession of evidence for the presumption-raising fact to trigger the injunction to proceed as if the presumed fact were to obtain. Thus, with respect to the legal presumption that a child born during wedlock is legitimate, if jury members find evidence for a child's birth-during-wedlock, then by the legal presumption in question they are to take its legitimacy for true. The objective but hitherto unestablished fact of the child's birth-during-wedlock, however, is not enough to entitle any deliberator to make any assumption about its legitimate status. In light of this, the twin stipulations—that no descriptive self-ascriptions be authoritative and that only expressive self-ascriptions be presumably true—can be bought only at an exorbitant price. The reason is as follows. Given the stipulations, one is entitled to assume ahead of time that self-ascription *S* is true if and only if one knows, or has good reason to believe, that *S* is not descriptive but rather expressive. That is so because a descriptive self-ascription does not fall within the range of authoritative self-ascriptions and hence does not carry a presumption of truth. Moreover, the objective but hitherto unestablished fact of *S*'s expressive character is not enough to entitle any deliberator to make any assumption about *S*'s being true. But the requirement that one know, or have good reason to believe, that *S* is expressive—the requirement which must be met if one is to be entitled to assume in advance that *S* is true—is so excessive as to make the presumption rule defining first-person authority hardly of use as a guide for action.

To see this, note that in order to ascertain *S*'s illocutionary force, one has to determine whether the speech act is a matter of expressing or describing the relevant psychological state. In order to determine whether the speech act is a matter of expressing or describing the relevant psychological state, one has to ascertain the speaker's communicative intention. In order to ascertain the speaker's communicative intention, one has to determine his beliefs and desires. After all, the question whether someone intends to express or describe a particular psychological state by means of a particular self-ascription depends in part on whether he wants to express or describe the psychological state and believes that by making the self-

https://doi.org/10.1017/S0012217300002407 Published online by Cambridge University Press

ascription he will express or describe the psychological state. It also depends on whether this belief and desire have caused, in the right way, a desire to make the self-ascription. Thus, in order to be in a position to presume the truth of *S*, one has to determine the speaker's communicative intention and the underlying matrix of beliefs and desires. Sometimes it is fairly easy to ascertain a speaker's communicative intention. If you ask me whether the cat is on the mat and I say, "I think so," it would not make sense to interpret me as asserting something about my doxastic state rather than about the cat's location. It would be absurd to take me to be describing, as opposed to expressing, my belief.[62] But, if I say unpromptedly, "I believe the cat is on the mat," or "I want to feed the cat," it is not as easy to determine whether I am expressing or describing my mental state. Since ascertaining a speaker's mental states, in particular those that constitute his communicative intentions, is normally more challenging than just getting oneself in a position to presume that his self-ascriptions are true, the aforementioned requirement and the twin stipulations which underpin it must be rejected. It is neither analytic to the concept of first-person authority nor constitutive of a reasonable conception of it that one needs to know the communicative intentions of the self-ascribing subject to presume his self-ascriptions true. Any stipulations that require one to have such knowledge are therefore unacceptable. If first-person authority is of any real theoretical interest, every sincere intelligible determinate first-person singular simple present-tense ascription of intentional state, regardless of its illocutionary force, must be authoritative.

Jacobsen has never given, or attempted to give, an argument for expressivism, or for minimalism for that matter. Expressivism for him is more of a pretheoretical intuition or platitude than a considered judgement or theory.[63] As such, it is presumed to be true. One is entitled to take it for granted unless one has sufficient evidence to the contrary. Countervailing reason is thus required to override this entitlement. So far the only serious argument against expressivism has been the one to the effect that expressive character clashes with truth-aptness. If this is a bad argument—and Jacobsen stands on good ground here—then we can safely return to our pretheoretical platitude. The suggestion does not succeed, however, because pretheoretical platitudes may stem from cultural indoctrination and lack evidential support. In the absence of a cogent argument for expressivism and the presence of considerable alternatives, a sceptic is surely justified in asking how we know that expressivism obtains. How do we know that concepts and predicates that serve to describe a person's state of mind in their second-person, third-person, past-tense, or future-tense use normally shed that function and assume a wholly different one in their first-person present-tense use?[64] In any event, a return to our pretheoretical platitude seems impossible or at least premature, for, if my objection holds up, the place the expressivist calls home is just not there.

https://doi.org/10.1017/S0012217300002407 Published online by Cambridge University Press

The expressivist proposal, as Wright colourfully puts it, "flies rather further than is usually thought. But it is a dead duck all the same."[65]

Notes

* I would like to thank Brandon Absher, David Alexander, Crystal Brothers, Kevin Bryant, Dorothy Carter, Bobi Conn, Helen Davis, Ezra Engling, Roberta Gerwing, Laura Gibson, Amy Jessie, Stuart Kendall, Rebecca King, Charles Larmore, Marlowe Moore, Amy Newman, Russ Payne, Baron Reed, Emily Shearer, Justin Smith, Rion Smith, Michelle Turk, members of the EKU Philosophy Reading Group, and two anonymous referees from this journal for their helpful comments and suggestions on an earlier draft of this article. Special thanks to Rockney Jacobsen, who read through multiple drafts. Above all, I would like to thank Nhi Huynh for her constant support and encouragement.

1 For a characterization of first-person authority along these lines, see Sydney Shoemaker, "First-Person Access," in his *First-Person Perspective and Other Essays* (Cambridge, MA: Cambridge University Press, 1996), pp. 50-73, esp. p. 51; Crispin Wright, "Wittgenstein's Later Philosophy of Mind: Sensation, Privacy, and Intention," *Journal of Philosophy*, 86 (1989): 622-34, esp. pp. 632-33; and Donald Davidson, "First Person Authority," *Dialectica*, 38 (1984): 101-11, esp. p. 101.

2 For a systematic critical exploration of these and related first-person notions, see William Alston, "Varieties of Privileged Access," *American Philosophical Quarterly*, 8 (1971): 223-41.

3 See Donald Davidson, "Knowing One's Own Mind," *Proceedings and Addresses of the American Philosophical Association*, 60 (1987): 441-58, esp. p. 442; and Davidson, "First Person Authority," p. 103.

4 Donald Davidson, "Davidson, Donald," in *A Companion to the Philosophy of Mind*, edited by Samuel Guttenplan (Oxford: Blackwell, 1995), pp. 231-36, esp. p. 234. If silence implies consent, then Davidson has to commit himself to the idea that the true-presumptiveness of self-ascriptions is a matter of logical necessity. The passage, which he quotes from Alston ("Varieties of Privileged Access," p. 235), characterizes the special status of self-ascriptions in terms of a certain logical impossibility, but his sweeping criticism of the characterization says nothing about Alston's use of this modal notion. See Davidson, "First Person Authority," p. 107.

5 Shoemaker, "First-Person Access," p. 50.

6 Ibid., p. 51.

7 Sydney Shoemaker, "On Knowing One's Own Mind," in his *First-Person Perspective and Other Essays* (Cambridge, MA: Cambridge University Press, 1996), pp. 25-49, p. 25.

8 Shoemaker, "First-Person Access," p. 51; italics in original. In what sense is it necessary? Shoemaker's ensuing discussion on p. 52 provides a clue, for he uses the expression "conceptually necessary" therein.

https://doi.org/10.1017/S0012217300002407 Published online by Cambridge University Press

9 For the distinction between the conceptual problem and the epistemological problem of other minds, see Alec Hyslop, "Other Minds," in *The Stanford Encyclopedia of Philosophy*, edited by Edward Zalta (Winter 2005 Edition, URL = <http://plato.stanford.edu/archives/win2005/entries/other-minds/>). Hyslop provides a more comprehensive examination of the epistemological problem in his *Other Minds* (Dordrecht: Kluwer, 1995). Anita Avramides does the same for the conceptual problem in her *Other Minds* (London: Routledge, 2001).

10 See Davidson, "Davidson, Donald," p. 235; "Knowing One's Own Mind," p. 442; and "First Person Authority," p. 101. Davidson seems to think that the reason there is a problem about first-person authority can be elucidated in the modality of either the ontology of psychological states or the semantics of psychological expressions. This is a suggestive hypothesis, but I shall not explore any connection between ontology and semantics here.

11 Compare Richard Moran, "Interpretation Theory and the First Person," *Philosophical Quarterly*, 44 (1994): 154-73, esp. pp. 156-57.

12 Rockney Jacobsen, "Wittgenstein on Self-Knowledge and Self-Expression," *Philosophical Quarterly*, 46 (1996): 12-30, p. 13. Compare Crispin Wright, "Critical Notice of Colin McGinn, *Wittgenstein on Meaning*," *Mind*, 98 (1989): 289-305, p. 294.

13 See, for instance, Christopher Peacocke, *A Study of Concepts* (Cambridge, MA: MIT Press, 1992), pp. 147-76.

14 Jacobsen, "Wittgenstein on Self-Knowledge and Self-Expression," p. 13.

15 Davidson, "First Person Authority," p. 103.

16 Tyler Burge, "Content Preservation," *Philosophical Review*, 102 (1993): 457-88, p. 467. Burge offers a defence of what he calls "the Acceptance Principle," which states that "[a] person is entitled to accept as true something that is presented as true and that is intelligible to him unless there are stronger reasons not to do so" (ibid.). The principle, he contends, is neither "a statistical point" (ibid., p. 468) nor "an empirical principle. . . . The justificational force of the entitlement described . . . is not constituted or enhanced by sense experiences or perceptual beliefs" (ibid., p. 469). This principle is meant to apply to anything that is presented as true and that is intelligible to a rational interlocutor. It makes no difference whether the subject matter is one's own present intentional states or something else. But if the Acceptance Principle holds, the parity of entitlement across first-person and third-person contexts will obliterate the asymmetry defining first-person authority. Perhaps aware of this consequence, Burge characterizes the special status of self-ascriptions not in terms of their presumptive acceptability but in terms of something much stronger, namely, their self-verifying character. He argues that self-judgements such as *I doubt that water is more common than mercury* "are self-verifying in an obvious way: making these judgements itself makes them true" (Tyler Burge, "Individualism and Self-Knowledge," *Journal of Philosophy*, 85 [1988]: 649-63, p. 649).

https://doi.org/10.1017/S0012217300002407 Published online by Cambridge University Press

17 Crispin Wright, "Self-Knowledge: The Wittgensteinian Legacy," in *Knowing Our Own Minds*, edited by Crispin Wright, Barry Smith, and Cynthia Macdonald (Oxford: Clarendon Press, 1998), pp. 13-45, esp. pp. 17-18; italics in original.

18 Compare Wright, Smith, and Macdonald, "Introduction," in ibid., pp. 1-11, esp. p. 1.

19 See Jacobsen, "Wittgenstein on Self-Knowledge and Self-Expression." See also his "Self-Quotation and Self-Knowledge," *Synthese*, 110 (1997): 419-45, and his "Semantic Character and Expressive Content," *Philosophical Papers*, 26 (1997): 129-46. I shall not try to determine whether the views that Jacobsen ascribes to Wittgenstein are really Wittgenstein's. Jacobsen articulates and defends these views, and they are important enough to merit serious discussion. Variants of Jacobsen's Wittgensteinian account include: Dorit Bar-On, *Speaking My Mind: Expression and Self-Knowledge* (Oxford: Clarendon Press, 2004); Kevin Falvey, "The Basis of First-Person Authority," *Philosophical Topics*, 28 (2000): 69-99; David Finkelstein, *Expression and the Inner* (Cambridge, MA: Harvard University Press, 2003); Andy Hamilton, "The Authority of Avowals and the Concept of Belief," *European Journal of Philosophy*, 8 (2000): 20-39; and Victoria McGeer, "Autistic Self-Awareness," *Philosophy, Psychiatry, and Psychology*, 11 (2004): 235-51. For simplicity, I shall consider only Jacobsen's account, but the critique to be advanced arguably applies to its variants as well.

20 Jacobsen, "Wittgenstein on Self-Knowledge and Self-Expression," pp. 13-17, 20, 30.

21 Richard Moran advances such considerations in his *Authority and Estrangement: An Essay on Self-Knowledge* (Princeton, NJ: Princeton University Press, 2001), pp. 71-72, 103.

22 Compare Jacobsen, "Wittgenstein on Self-Knowledge and Self-Expression," pp. 21-22: "On the minimalist account of truth, the Disquotational Schema (hereafter DS), '"*p*" is true if and only if *p*,' tells us all there is to know about truth. . . . By DS, a truth-assessable sentence is any suitable substitute for 'p' in the schema."

23 It is unclear what the intended modality of the "can" in the above statement is. But here I am just paraphrasing Jacobsen's formulation of the minimalist requirements for truth-assessability: "A meaningful sentence of a language counts as *truth-assessable* if it can occur without (surface) syntactic incongruity as an antecedent-clause and has a significant negation" (Jacobsen, "Wittgenstein on Self-Knowledge and Self-Expression," p. 22; italics in original).

24 Ibid., pp. 21-22.

25 Crispin Wright, *Truth and Objectivity* (Cambridge, MA: Harvard University Press, 1992).

26 Peter Geach, "Assertion," *Philosophical Review*, 74 (1965): 449-65.

27 Jacobsen, personal communication, August 15, 2005.

28 Jacobsen, "Wittgenstein on Self-Knowledge and Self-Expression," p. 23.

https://doi.org/10.1017/S0012217300002407 Published online by Cambridge University Press

29 Compare ibid., pp. 26-27 (italics added): "The mere appearance of the performative verb '*V*' in the prefix ('I *V* that . . . ') introducing the content clause *divests those utterances [of explicit performatives] of assertoric status*, assigns them another performative status and hence another expressive character. . . . We are obliged, by minimalism, to regard explicit performatives as truth-assessable utterances . . . but not as expressions of belief. . . . [N]othing confers on our utterances of explicit performatives *the particular non-assertoric status they will have (by default)* apart from the meaning of the performative verbs that prefix their content clauses."

30 Ibid., pp. 23-28. See also Bar-On, *Speaking My Mind*, pp. 235f.; Falvey, "The Basis of First-Person Authority," p. 72; and Wright, "Self-Knowledge: The Wittgensteinian Legacy," p. 36. Unlike Jacobsen, Bar-On, Falvey, and Wright, Finkelstein seems innocent of, or oblivious to, the independence between truth-assessability and assertoric force, for he apparently equates the two: "[I]t does not follow from an avowal's not being a report that it lacks a truth-value. An avowal of hope that is not a report may yet be a true assertion. . . . [P]sychological self-ascriptions . . . have an assertoric dimension; they may be true or false" (Finkelstein, *Expression and the Inner*, p. 97). He reasonably claims that being a non-report does not entail lacking a truth-value. But his example about an avowal of hope suggests that for him the possession of a truth-value confers not only truth-evaluability but also assertoric status. The suggestion is confirmed by his subsequent elucidation of the notion of an assertoric dimension in terms of the capacity for being true or false. It may seem uncharitable to accuse Finkelstein of missing the obvious fact to which Jacobsen, Bar-On, Falvey, and Wright draw our attention—the fact that not every utterance of a truth-assessable sentence which means that *p* is an act of asserting that *p*. If the accusation is false, however, then the only way to make sense of the above passage is to view Finkelstein as trying to advance the highly implausible thesis that self-ascriptions are assertions as well as expressions of the same psychological states even when wholly and unequivocally expropriated of assertoric status and credentials by etiolating features of context. This interpretation is hardly more generous than the last one.

31 Jacobsen, "Wittgenstein on Self-Knowledge and Self-Expression," pp. 29-30.

32 Jacobsen takes himself to be offering a full account of "the reliable authority all persons enjoy concerning their own mental states" and "a speaker's effortless and reliable ability to provide true self-ascriptions." See ibid., pp. 12, 13, 29, 30.

33 Ibid., p. 16.

34 Ibid., pp. 15-16.

35 See Wright, "Self-Knowledge: The Wittgensteinian Legacy," p. 37.

36 James Urmson, "Parenthetical Verbs," *Mind*, 61 (1952): 480-96, p. 485. Compare David Rosenthal's observation: "'I think that *p*' can, of course, be used to qualify the firmness of one's conviction that *p* is the case, in ways that simply asserting '*p*' doesn't" ("Moore's Paradox and Consciousness," in *Philo-*

https://doi.org/10.1017/S0012217300002407 Published online by Cambridge University Press

sophical Perspectives, vol. 9, *AI, Connectionism, and Philosophical Psychology*, edited by James Tomberlin [Atascadero, CA: Ridgeview Publishing, 1995], pp. 313-33, p. 331).

37 Paul Grice, "Logic and Conversation," in his *Studies in the Way of Words* (Cambridge, MA: Harvard University Press, 1989), pp. 22-40.

38 Timothy Williamson, *Knowledge and Its Limits* (Oxford: Oxford University Press, 2000), p. 243.

39 Ibid., p. 243.

40 Ibid., p. 256.

41 Ibid., p. 257.

42 Ibid.

43 The quotation comes from Hilary Putnam's science fiction story about the super-super-spartans or super-super-stoics, imaginary individuals who are capable of feeling pain but, unlike us, have the ability to successfully suppress all involuntary pain behaviour and even all talk of pain. See his "Brains and Behavior," reprinted in *The Nature of Mind*, edited by David Rosenthal (Oxford: Oxford University Press, 1991), pp. 151-59, esp. pp. 154-55.

44 See Gilbert Ryle, *The Concept of Mind* (London: Hutchinson, 1949), p. 166.

45 See John Locke, *An Essay Concerning Human Understanding*, edited by Peter Nidditch (Oxford: Clarendon Press, 1975), p. 105, and David Armstrong, *A Materialist Theory of the Mind* (London: Routledge, 1968), pp. 323-38.

46 Jacobsen, "Wittgenstein on Self-Knowledge and Self-Expression," pp. 16-17; italics in original. Moran rejects something very much like this view, which he calls "the Presentational View," and the attribution of it to Wittgenstein in his *Authority and Estrangement*, pp. xvii, xix, 70-73, 100-107; see also Moran's "Interpretation Theory and the First Person," p. 157, n.3.

47 Ibid., pp. 156-57; italics in original.

48 Jacobsen, "Wittgenstein on Self-Knowledge and Self-Expression," pp. 23-28, esp. p. 26. See also n.30 above.

49 Jacobsen, "Wittgenstein on Self-Knowledge and Self-Expression," pp. 15, 30.

50 I am not hostile to the idea of banishing first-person authority as traditionally conceived; on the contrary, I am quite receptive to that idea. See my "Davidson on First-Person Authority," *Journal of Value Inquiry*, 38 (2004): 457-72; and my "Why There Is No Such Thing as First-Person Authority," *Southwest Philosophy Review*, 16 (2000): 165-89. I am hostile only to the idea of banishing first-person authority on the grounds that one cannot explain it.

51 Wright, "Critical Notice of Colin McGinn, *Wittgenstein on Meaning*," p. 294.

52 Jacobsen, "Wittgenstein on Self-Knowledge and Self-Expression," p. 13.

53 The phrase "first-person avowability" occurs in Wright, "Critical Notice of Colin McGinn, *Wittgenstein on Meaning*," p. 293.

54 Wright, "Wittgenstein's Later Philosophy of Mind: Sensation, Privacy, and Intention," pp. 632-33; italics in original.

55 Crispin Wright, "Wittgenstein's Rule-Following Considerations and the Central Project of Theoretical Linguistics," in *Reflections on Chomsky*, edited by

https://doi.org/10.1017/S0012217300002407 Published online by Cambridge University Press

Alexander George (Oxford: Blackwell, 1989), pp. 233-64, esp. pp. 251-52; italics in original.

56 Wright, "Self-Knowledge: The Wittgensteinian Legacy," pp. 17-18; italics in original. See also ibid., p. 41.

57 Compare André Gallois, *The World Without, the Mind Within: An Essay on First-Person Authority* (Cambridge, MA: Cambridge University Press, 1997), p. 31; and Bernhard Thöle, "The Explanation of First Person Authority," in *Reflecting Davidson*, edited by Ralf Stoecker (Berlin: Walter de Gruyter, 1993), pp. 213-47, esp. p. 217.

58 Davidson, "First Person Authority," p. 105.

59 Compare ibid., p. 103: "Even in the exceptional cases, however, first person authority persists; even when a self-attribution is in doubt, or a challenge is proper, the person with the attitude speaks about it with special weight."

60 For a canonical discussion of psychological defence mechanisms, see Anna Freud, *The Ego and the Mechanisms of Defense* (New York: International Universities Press, 1967).

61 See Edna Ullmann-Margalit, "On Presumption," *Journal of Philosophy*, 80 (1983): 143-63, esp. pp. 150-51.

62 See David Rosenthal, "Thinking that One Thinks," in *Consciousness*, edited by Martin Davies and Glyn Humphreys (Oxford: Blackwell, 1993), pp. 197-223, esp. p. 214.

63 Jacobsen, personal communication, August 10, 2004.

64 Moran, *Authority and Estrangement*, pp. 71-72.

65 Wright, "Self-Knowledge: The Wittgensteinian Legacy," p. 38. Although I concur with Wright's conclusion, my critique of the expressivist proposal is not quite the same as his. See the end of §5 of the present article.

https://doi.org/10.1017/S0012217300002407 Published online by Cambridge University Press