

FIRST-PERSON AUTHORITY ABOUT ONE'S OWN MIND¹

Bruno Mölder

University of Tartu

Abstract. According to the thesis of first-person authority, my knowledge of my own mental states is substantially different from my knowledge of the mental states of others. In this paper, the thesis of first-person authority is examined and refuted. The thesis consists of four different theses: self-intimacy, infallibility, indubitability and incorrigibility. It is possible to develop counter-arguments to all of them. Also, I attempt to show that ineffable awareness could not guarantee first-person authority. An alternative account to first-person authority is suggested that is free of the Cartesian Model of Mind. The account rests upon the folk theory of mind and on the nature of interpretative practice. We ascribe first-person authority to ourselves and interpret other people as if they would enjoy privileged access to their own mental states. Although first-person authority forms the central part of folk psychology, it lacks metaphysical power – i.e., one cannot build coherent metaphysical systems based on folk-psychologically interpreted first-person authority. The reason for this is first, that folk psychology as a theory need not be true and second, that the first-person authority that belongs to the folk psychology is only contingent.

Introduction

The dissatisfaction with the Cartesian conception among philosophers of mind is revealed by their rejection even of notions that seem *prima facie* intelligible. One of those notions is the conviction that one enjoys a kind of special epistemic authority about one's own mental states. First-person authority is tried to explain away, because it tends to cause serious obstacles to different philosophical programs (e.g. behaviourism, identity theory, central-state materialism, semantic externalism). In this paper, I discuss the arguments presented against various forms of the thesis of first-person authority. I conclude that it is possible to support the epistemic authority *vis-à-vis* one's own mind without committing oneself to the Cartesian picture. Viewing first-person authority as an integral part

¹ I would like to thank Jüri Allik and Madis Kõiv for their valuable comments and discussion of the earlier versions of this paper. Special thanks to Wolfgang Drechsler for his editorial suggestions. I also profited from discussions of an earlier version of this paper at the Seminar of the Estonian Society of Analytic Philosophy.

of folk psychology without metaphysical flavour enables to avoid the problems of the Cartesian Model of Mind² without giving up the common-sense intuitions.

The Cartesian Model and the thesis of first-person authority

According to the Cartesian Model, the human mind is like a private "theatre", where the conscious self watches the private objects on the stage – representations, appearances, sense data, qualia. The common property of the mental objects, which discriminates the mental from the physical, is the indubitability of their existence for the self. The *Cogito*³ was indubitable for Descartes, for while it is possible to doubt in every matter of fact (an evil demon could deceive us after all), the *cogito* is the logical precondition of the deception, since being deceived is a matter of having false beliefs and believing is a case of *cogito* (Shoemaker 1990). The mental states are transparent to the self according to the Cartesian Model – if a mental state enters the theatre, then I could not fail to know that and cannot be mistaken about its nature. My mental states are entities that I know infallibly and incorrigibly. The third person has no reason to doubt in my reports and beliefs of my own mental states. This all gives me the special epistemic authority concerning my own mental states which I do not have concerning other people's mental and physical states.

The notion of first-person authority – or, as some have named it, privileged access – denotes several theses of different strength. It is important to notice that these are theses, not definitions.

The first is the thesis of **self-intimation** or the thesis of **omniscience**: all mental states are such that if I am in state *x*, then I believe⁴ that I am in state *x*. It is impossible for me not to believe that I am in state *x*, when I am in that particular state.

The second is the **infallibility** thesis: all mental states are such that if I believe that I am in state *x*, then I am in state *x* and not in state *y*, since I cannot have false beliefs about my own mental states.

The third is the **indubitability** thesis: all mental states are such that if I believe that I am in state *x*, then I have no reason to doubt whether I am in that state or not.

² Among the most celebrated attacks on the Cartesian Model are Ryle (1949), Rorty (1979), and Dennett (1991), but they stress other problems and therefore my description of the Cartesian Model might slightly differ from their account. This description does not strive to be an authentic summary of René Descartes' views, for it is still an open question whether Descartes was a proper Cartesian. The Cartesian Model denotes mainly the tacit presumptions that are shared by many philosophers and psychologists in the present century.

³ Descartes meant by the *Cogito* not only *thinking*, but *every mental state we are aware of*, including understanding, wanting, believing, imagining and sense-perception (see Kenny 1973: 114–119). It is important to notice that such a definition excludes the possibility of unconscious mental states.

⁴ In this paper I use the word "belief" to mean the possibility of report. The sufficient condition for P's believing that *x* is that P would report that *x* if he were asked, if he understood the question and if he is sincere (Cf. Alston 1971: 229).

The fourth is the **incorrigibility** thesis: all mental states are such that if I believe or report that I am in state *x*, then nobody has reasons to correct my belief or report in saying that actually I am not in that particular state.

It may be said that the thesis of first-person authority is the disjunction of these four theses. The refutation of each of them leads to the refutation of the thesis of first-person authority.

On this construal the first three theses concern the relation between mental states and belief or report of the first-person. The fourth, the incorrigibility thesis, concerns the relation between belief or report and the higher-order belief of the first-person or between the belief or report and the belief of the third-person.

The thesis of self-intimation entails the infallibility thesis. The former thesis declares that I am conscious of every mental state I have. The infallibility thesis is weaker, since it does not imply the consciousness of every mental state, but it declares that whenever I believe that I am in a particular mental state, I cannot be mistaken about what I believe. The indubitability thesis is weaker than the former two – it does not commit us to the infallibility, but it states instead that doubt concerning my beliefs of my current mental state is unjustified.

The infallibility thesis commits one to the incorrigibility thesis, for if one's mental state beliefs or reports are always true, then nobody is justified to correct them. The converse is not true – the incorrigibility thesis is weaker than the infallibility thesis, since the falseness of report does not imply corrigibility, for the correction may be unjustified in some cases. Given the fact that the report has been corrected, one can infer that this particular report was fallible. Accordingly, if one could refute the incorrigibility thesis, one has refuted the infallibility thesis at the same time.

Objections to the thesis of first-person authority

The first step in analysing the thesis is to define a set of mental states about which it is sound to formulate the thesis at all. There turns out to be no intuitively acceptable criteria of the mental. Rorty (1970) has suggested the incorrigibility as the mark of the mental, but his formulation of incorrigibility as the lack of accepted procedure for overriding first-person reports, does not individuate the states we are used to call the mental. In the present context, this criterion begs the question, since it entails that no corrigible report is the report of a mental state. It would appear that it is not possible to give the common criteria to the mental. There could be a kind of family resemblance between the entities called mental, but there is no individuating property for the mental states. The mind is a notion with blurred edges.

Starting with the objections to the thesis of first-person authority, it should be noted that the unconscious and subliminal states of mind belong to the class of mental states. This fact refutes the thesis of self-intimation. Since the access to consciousness is controlled by the focus of attention, it is possible for us to have

mental states without being conscious of them. The attention may be conscious (voluntary) or unconscious (automatic).

The thesis of omniscience leads to an infinite regress (Ryle 1949:156–158). If I am simultaneously aware of my mental states and beliefs about the states, then I must be aware of beliefs about beliefs or be aware of my being aware. As it is logically possible to follow this line infinitely, the holder of the thesis of omniscience must delimit the possible number of synchronous acts of attention that are directed to mental states. Hence there must be mental states inaccessible to consciousness, namely, those that incorporate more synchronous acts of attention than possible. Then the proponent of the thesis of omniscience must explain how one could know about mental states that are inaccessible to introspection, unless he admits either that this knowledge is not based on introspection or that one does not possess this knowledge.

Partly on the influence of the above-mentioned critique, it is nowadays hard to find a philosopher who endorses the thesis of first-person authority concerning every state of mind. Usually the distinction is made between dispositional mental states and occurrent mental events. The former include beliefs, moods, wants, intentions, emotions, motives, desires and purposes – they are such that we can have them without accompanying belief about having them. Occurrent mental events are experientially detectable: for a person to have them is to experience something. Occurrent mental events include pains, after-images and occurrent, datable thoughts (Rorty 1970; Audi 1974).⁵ The thesis of first-person authority is defended concerning the latter, not the former states of mind. In case of occurrent mental events, the argument from the infinite regress does not apply, for neither believing nor knowing is an occurrent event (Audi 1974:264).

It is possible to refute first-person authority concerning occurrent mental events.⁶ The argument is following: the knowledge of a fact presupposes the categorisation of that fact, which in turn presupposes the implicit categorical or theoretical framework. To believe or report that I have pain, I need to know the meaning of the word "pain". If I do not command language, my reports are constantly mistaken. Another, closely related argument concerns the vagueness of mental state terms. There are cases when mental states are hard to classify – for example, I may be in trouble considering whether this particular feeling is a pain or a strong itch or something in between.⁷ This means that our categorical framework (the ordinary language) is not applicable in every situation. Although I

⁵ Although I find this distinction insufficiently based, for the sake of argument it is indispensable to follow this usage. Suffice it to say that the treatment of self-knowledge as a theoretical achievement is not committed to this distinction.

⁶ Analogous arguments are analysed in Malcolm (1967), Alston (1971), Audi (1974) and Doppelt (1978) to take a more or less random selection.

⁷ I shall analyse a similar question in the next section of the paper. But in case of non-linguistic awareness, the solution is obviously different, since without language, or inside the private conceptual scheme, the doubt seems out of place.

have no possibility for checking the correctness of my report in such a case, the knowledge of the vagueness of mental state terms gives me the reason to doubt in the correctness of my reports. To remove this perplexity, I could appeal to the previous experience or to the tacit knowledge of causes of my sensation. This appeal, however, might not resolve my doubts concerning the classificatory reliability of my reports, since the possibility of a mistake could not be excluded. But this is just the important fact in repudiating Cartesianism – the banishment of metaphysical certainty guarantees the logical possibility of rational challenge to our beliefs about our own mental states (Cf. Aune 1967:45).

It might be objected that the preceding arguments may cast doubt on the first-person authority with regard to the properties of occurrent mental events, but nothing has been said about our knowledge of the existence of them. After all, I may report that I feel something, but cannot say exactly what it is that I am feeling. Actually, there are some objections to the infallibility of the reports about the existence of occurrent mental events to which I shall turn presently.

Firstly, there is the "distinct existences" argument, endorsed by Armstrong (1968:106) and Smart (1970:108) – the introspective awareness of one's own mental states is the effect of the mental states that are objects of awareness. Causes and effects must be distinct existences, since they are only contingently conjoined. Following these considerations, it is logically possible to be aware of pain without being in pain (the refutation of infallibility thesis) and to have pain without being aware of that pain (this refutes the thesis of self-intimation). This argument is open to considerable criticism, and for that reason it does not have much force against the thesis of first-person authority. In employing the observational model of introspection, it implies the materialist version of the Cartesian theatre (Shoemaker 1994). The second trouble with the argument is that it has gone astray from the definition of occurrent mental state. It is said that the concept of pain is such that it is impossible to have pain without feeling pain (Kripke 1972, Lewis 1972).

I suggest a version of "distinct existences" argument, which is free from the above-mentioned defects. It is not possible to have pain without feeling pain, but it by no means follows from this grammatical fact that it is impossible to feel pain without being conscious of it or to believe that one has pain without feeling it. Granting that "having" a sensation means the same as "feeling" a sensation, I do not grant that "having a sensation" means the same as "believing that one has a sensation" (Cf. Malcolm 1967:132–134). Recent analysis of the concept of consciousness by Ned Block (1995) gives support to the preceding considerations. Block distinguished access-consciousness (the content of the state is representational, poised for rational control of speech and action) and phenomenal consciousness (the content is experiential and not necessarily representational). The awareness is not limited to what lies in the focus of attention according to Block's model. He illustrates the point: one is engaged in intense conversation when suddenly one realises that outside the window there has been "a deafening

pneumatic drill digging up the street". Block (1995:234) states that one was P-conscious of the noise all the time, but through the act of realisation one became both P-conscious and A-conscious of it. In accordance with this model, it is possible for me to have pain, i.e. to be P-consciously aware of pain, without believing that I have pain, since I am not A-conscious of pain, and believing is an A-conscious process. Since it is possible for me not to believe that I have pain, when I am in pain, the thesis of omniscience is refuted.⁸ As for the thesis of infallibility, it may be said that since P-consciousness need not be logically prior to the A-consciousness, it is possible to believe that I am in pain without feeling the phenomenological quality characteristic of pain. Although this is not a very rational thing to do, there are commonly familiar cases of self-deception along this line.⁹

There is another objection to the infallibility thesis inspired by Armstrong (1968:104–105). The mistake is at least logically possible due to the time gap between the report and the state reported. It is not possible for a mental event and introspective report to occur at the same instant of time. Every introspective report is actually retrospective (see also Ryle 1949). The reliability of retrospective reports depends on the reliability of the short-term memory. If this line of thought is sound, then the infallibility of reports must entail the infallible memory.¹⁰ The converse is not true. Even granting the possibility of infallible memory does not lead us closer to the infallibility of reports, since reporting presupposes the verbalisation of the material to be reported. The verbalisation entails the possibility of a mistake, especially when the experimental subject is under a high cognitive load (Ericsson & Simon 1980). The process of verbalisation is necessarily accompanied by categorisation, and this process involves the application of general concepts to particular facts and hence is essentially liable to error.

If the preceding arguments are sound, the thesis of omniscience and the infallibility and indubitability theses are not true. As regards the incorrigibility thesis, it has actually two meanings, depending on the possible or impossible corrector: the impossibility of the correction by first-person or by third-person.

⁸ It is a curious thing about the nature of philosophical argumentation that on certain construal the model implies the weak version of the same thesis it refutes. If we substitute the expression "believe that" by "is aware of", then we have the weak form of the thesis of omniscience – mental states are such that if I am in state x, then I am aware of that state. This thesis is so weak that the above-mentioned objections do not apply to it, since the arguments concern the states of which we are A-consciously aware. But as P-consciousness is not enough for the belief of report, by the same token it is not enough to guarantee first-person authority.

⁹ As regards the rationality condition of the A-conscious state, I follow Block (1995: 231) in thinking of it as a sufficient, but not as a necessary condition.

¹⁰ This view is endorsed by Uus (1994: 65) to bring an example. Though he writes about "the private non-physical memory of qualities", his approval of the infallibility thesis (Uus 1994: 51–52), commits him to the postulation of the infallible memory.

The former is the view I shall presently turn to. A. J. Ayer, one of the proponents of the first-personal corrigibility and an opponent of the corrigibility by the third-person case, sets his position clearly in the following passage:

He may not be infallible, but still his word is sovereign. If he is not infallible, others may be right when he is wrong. Even so their testimony is subordinate to his in the same way and for the same reason as the testimony of the clairvoyants is subordinate to that of eye-witnesses. If his reports are corrigible it must be that he himself is ready to correct them, not after an interval of time in which a lapse of memory would rob him of his authority, but as it were in the same breath. [...] The logic of these statements that a person makes about himself is such that if others were to contradict him we should not be entitled to say that they were right so long as he honestly maintained his stand against them (Ayer 1964:73).

In saying that the testimony of the first-person is to be preferred to the evidence the third-person may have, Ayer compares this with the preference of the eye-witness' testimony to the testimony of the clairvoyants. His reason lies in the fact that the testimony of the clairvoyants is credible only through an agreement with the eye-witness' evidence.

In some reading, Ayer's view involves certain difficulties especially if we take it to presume the infallibility of the second report by which the person corrects his former report. Otherwise his correction would lead to an infinite regress. In light of the former critique of infallibility, any reason for accepting the infallibility of the second report remains invisible. Given that neither report is infallible, we have two reports – the first that is under correction and the second that corrects the former. What rationale could the bystanders have for preferring the latter to the former except for the reason that it was his last word? He may correct himself again and again after all. To revive some Wittgensteinian intuitions, it may seem difficult to take such a correction seriously, since the person cannot check the correctness of his report, he cannot compare the report with the reported mental state, because he is not in that particular state any more. It is just like buying a new copy of the same morning paper in order to check whether the first was telling the truth (Wittgenstein 1995: § 265).

This does not mean that it is impossible for the person to correct his own beliefs or reports. On the contrary, we correct them all the time. This correction forms a part of our folk psychological theory. It is a common-sense platitude that other people and I tend to correct our reports if they conflict with our expectations. The correction may seem impossible if we cling to the verificationist theory of truth in case of mental states instead of the coherence theory. Wittgenstein was right that one cannot verify one's reports by comparing them with the paradigm of the mental state reported. But there is a way open to justify the correction – the report is unacceptable if it is not coherent with the background knowledge. This kind of justification works both in case of the first-person and the third-person. In default condition the justification is unnecessary,

but the need arises in case of incoherence. The third-person has a rationale to accept the last report of a first-person if it is coherent to infer that the first-person corrected himself in the light of his background knowledge. To the Wittgensteinian this may simply seem as begging the question, for it could be possible that the person does not know what kind of state he wants to weigh against his background knowledge, for he is not in that state anymore. But this Wittgensteinian intuition overlooks one possibility. Granting the fallibility of memory, there is still a justification by the coherence of the memory reports applicable to the first-person (Ayer 1954; Cornman 1968). Concerning Ayer's point about the resemblance between the third-person and the clairvoyant, it is sufficient to notice that background generalisations and expectations are not established solely on the basis of first-person experience, for one cannot form a coherent body of knowledge by inducing from a single case. It can be concluded that the acceptability of the report depends on its conformity with the overall pattern of background knowledge, and the first-person does not always have the final say in it.

Now to the incorrigibility by the third-person. The first objection to it is known as the argument from the cerebroscope. Suppose that a successful neuroscientist has established the necessary and sufficient physiological correlates of mental states. Imagine now the subject of experiment reporting to have a sensation *S* and the neuroscientist discovering that the subject's brain is not in the state *S** being the necessary and sufficient correlate for the sensation *S* to occur. In such a case the experimenter is justified to correct the report (Armstrong 1968; Rorty 1970).

I do not regard the argument from the cerebroscope to have much force, since it presupposes the neuropsychological theory we do not have yet. At the present state of science, this argument belongs more to sci-fi stories than to the laboratory of neuroscience. But the argument could be developed in a modified form. At the present state of psychology, the "inverted cerebroscope" is clearly possible. The idea is the following: the psychologist fixes the brain state *S** and conditions the subject to discriminate the corresponding mental state *S*. If the subject does not discriminate the state *S*, then the psychologist justifiably corrects the subject.¹¹ Still, the problem with arguments of this sort is that they imply physicalism from the beginning and hence beg the question against those philosophers who, relying on first-person authority, attempt to refute physicalism. Another problem with this argument concerns the vagueness of mental state terms. There is no reason to suppose the one-to-one correspondence between folk psychological and neuroscientific classifications. From the neuroscientific point of view, it may appear that the common-sense distinction between wanting and desiring or between knowing and believing, for example, is not justified, i.e., their physiological realisations are not distinct.

I suggest another argument against the incorrigibility thesis that does not rely on the possible developments of science and is closely related to everyday

¹¹ I am grateful to Jüri Allik for suggesting this possibility to me.

practices.¹² Although the connection between mental state and behaviour is only contingent, the meaning of mental state terms is nevertheless partly determined by observable behaviour. It may be said that the background knowledge is a criterion for applying certain mental predicates to other persons. The mental state reports form only one part of the background knowledge, by no means are they the sole criteria of mental states. In saying that there is contingent connection between reports and mental states and between reports and background knowledge, I do not cast doubt on their criteriality. Two things need not be necessarily connected in order to be criteria for each other (see Dancy 1985:72). If we command language and know the relevant network of facts about the person (medical history, character traits, emotional and motivational state, etc.), then by observing his behaviour we are inclined to doubt his reports and even correct them if they contradict the facts we know about him. This tacit knowledge forms a part of our folk psychological theory, which states that usually people tell the truth, but sometimes they are mistaken, deceive oneself or simply lie. One does not believe the complaints of pain, etc. by neurotics or ignored children, if one has reason to believe that actually they seek attention (Cf. Audi 1974:257). Lacking relevant knowledge, one may be unable to decide. In such a case one either accepts the report or searches for the relevant knowledge.

If the preceding lines of thought are sound, then the thesis of first-person authority is refuted. It is false that one has first-person authority in all cases, hence there are cases when one does not enjoy first-person authority. Still, clearly there are situations when we do not make mistakes, do not doubt our mental states and we are not being corrected all the time. The explanation of these situations will take place in the last section of this article. In this section, my objective was to undermine the belief in the certainty of first-person knowledge. Metaphysics needs an absolute basis in order to be metaphysics. It is not reasonable to build a system (the subjective-scientific methodology in Uus (1994) for example) on those shaky foundations. The fact that the thesis of first-person authority is not true also cuts off the possibility of the refutation of philosophical theories by the appeal to first-person authority.

Ineffable awareness does not guarantee first-person authority

There may still be one way open to the proponent of the first-person authority thesis. Inasmuch as the above-mentioned arguments rely on the vagueness of mental state terms, the proponent of first-person authority may argue that our awareness of our own mental states need not be mediated by language. Some experiences of mine are ineffable, but this does not mean that I do not have them. The limits of my language do not mark the limits of my world. Why could not my awareness of the ineffable experience be incorrigible, infallible and indubitable? Other people cannot correct this experience, because they do not know which

¹² Parsons (1970) and Doppelt (1978) have developed the arguments of a similar kind.

experience it was; the ineffable experience does not have a place in the language game. I could not be mistaken in the existence of this experience, since having committed a mistake indicates that we had means to elucidate what the mistake consists of. But in case of infallible knowledge, I could not compare my experience with the paradigm of the experiences, since I do not have this paradigm. To paraphrase the words of Wittgenstein: I cannot even buy a new copy of the morning paper, for I cannot say which morning paper it is I should buy. In this respect I cannot doubt my ineffable experience as well.

It is possible that this view is correct. But first-person authority construed in that way would serve its purpose to a very limited extent. First-person authority has sense only in the potential context of communication, only when someone has the right to say the last word. About the ineffable experience nothing can be said. The elusiveness of experiences gives rise to the situation where the person loses the continuity of his experiences; he could not know that his experience in the next instant of time is of the same kind as it was an instant before. One cannot even say that it is the same kind of *experience*, for it would mean that the act of classification has been made. One cannot say about the ineffable that it is experience. To say simply that it is "the same" is meaningless without mentioning what it is that is identified, for the sameness presumes the criterion of identity (Geach 1971:69).

In order not to lose his ground, the person may name his ineffable sensation by pointing to it and saying: 'This sensation I will call *A*'. Even this may not be possible as the following private language argument purports to show. The word presupposes the rules in order to be used correctly. These rules cannot be private, for if at most only one person can know the rules for the correct use of *A*, then no one can distinguish the difference between someone obeying and merely thinking that he is obeying the rules. In that case, no one can know whether someone obeys the rule, hence no one could know how to use the expression *A* correctly. Such an expression is meaningless. For the expression to be meaningful the rules for its correct use must be public. Rule-following is a public activity in the language community that guarantees common understanding. If the language is ever used for communication, its words cannot obtain their meaning from private objects (Wittgenstein 1995).

It may seem that the argument has little force against the solipsist who is not interested in talking about his ineffable experiences to others. There is, however, a possibility to adapt this argument to show that the solipsist is not able to use the word *A* to communicate with his "later self" (Dancy 1985). It is enough to imagine his experience constantly changing as a result of which there is no criterion of sameness for it.

One may ask: what does it mean to say that the awareness of the ineffable experience is infallible? Ordinarily I am infallible only if it is possible to verify that no mistake has been made. Both concepts of fallibility and infallibility presuppose the criteria for correctness. If we cannot draw a distinction between

erring and not erring, then to say that I am infallible is merely to pay a meaningless compliment to oneself (Cf. Rorty 1965). If it is not appropriate to say about something that it is corrigible, fallible or dubitable, then it also cannot be said that it is infallible, incorrigible or indubitable. The same point may be put differently and even more convincingly for some philosophers of a different training. The infallible report is the true report. To doubt is to bring the truth value of the report into question. The correction means a change in the truth value of the report. The concept of truth value belongs to semantics and presupposes the propositional form. In case of a lack of the propositional form (the ineffable experience), we do not have semantics, hence the ineffable experience does not have truth value. To say that ineffable experience is incorrigible, infallible or indubitable is to say something meaningless.

In some reading of Wittgenstein he held the similar view about all sensations – it is meaningless to doubt the correctness of the first-person sincere avowals, since being not independent of what they express, these could not be regarded as independent reports (Malcolm 1963). This might lead to the question whether the generalisation of the private language argument to all sensation reports will undermine the objections in the preceding section and would show that there is blatant inconsistency in the present argument. It may be also said that in the case of sensations, some form of verificationism is inevitable, since it is pointless to talk about mistakes if no one is able to verify whether the mistakes have been made.

I should point out that I do not follow Wittgenstein in that point. Saying that it is meaningless to doubt the avowals, Wittgenstein (1995: §§ 281–288) postulates a non-contingent connection between the inner processes and outward criteria in order to avoid the scepticism of other minds. The arguments in the preceding section show that the connection between reports as outer behaviour and mental states is only contingent. Hence the doubt in the correctness of reports is in some cases justified. In order to accept the report under discussion, there must be a kind of coherence between the report and our background knowledge. If the report conflicts with the background knowledge, we must assume the report to be false in order to preserve the rationality condition. In the case of ineffable knowledge, no such background is available and correction is not possible.

First-person authority as theoretical principle of folk psychology

The notion of first-person authority is not vacuous. From the fact that the thesis of first-person authority could be refuted does not follow that there could not be cases when we do not make mistakes or do not doubt our state of mind.

We have a network of beliefs about the world which can be called the implicit folk psychological theory. The folk psychological theory is the only realm where we all must be experts, for otherwise communication would be impossible. In the light of this theory, we interpret our behaviour as the behaviour and our reports as the reports. One could ascribe intentions, beliefs and sensations to oneself only by

the medium of this implicit theory of interpretation. The ascriptions of mental states are theoretical by its nature. In this regard, there is no point in distinguishing between the ascription of dispositional states and occurrent mental events. I cannot ascribe jealousy or happiness to myself or others if I do not know what "jealousy" and "happiness" are. Similarly, I cannot attribute pain to myself or others if I do not know what "pain" is, and this knowledge presupposes the knowledge of various facts external to my sensations.

First-person authority belongs to this network of common-sense understanding. We believe that the speaker knows what he is talking about, otherwise there would not be anything to interpret (Davidson 1984). We believe that persons around us are (minimally) rational (Cf. Dennett 1971, Shoemaker 1990, 1994, 1995). It means that we believe that if someone reports that he is in pain, then he is in pain and that he believes that he is in pain. This is just another way of saying that, provided that people are rational, they must enjoy first-person authority. The same thing, we believe, must be true in the first-person case also. According to folk psychology, the first-personal beliefs about one's own mental states normally require no justification.

Provided that the concept of first-person authority is closely tied up with rationality, it is normative. The rationality of human agent, however, is not exhausted by first-person authority. There is also a notion of rationality that stresses the coherence between beliefs, behaviour and background knowledge. We ascribe certain beliefs and other mental states to ourselves, since relying on the background knowledge we believe that it would be rational for us to have such mental states. In case of inconsistency, we do not set the agent's overall rationality in doubt, but we distrust instead the rationality of the particular act. Sometimes this involves overriding the first-person authority. Normally one is not aware of that tacit comparison with background knowledge, since these inconsistencies are not very common.

Our folk psychology is embedded into ordinary language.¹³ If somebody reports that he is in pain, we believe him and it seems senseless to ask: "How do you know that?" Even if somebody asks this question, he nevertheless presumes first-person authority, since otherwise the question would be baseless. One might repeat the same question again and again to every answer the person gives. Another example concerns Moore-paradoxical sentences. Sentences like "It rains, but I do not believe it is raining", are odd from the point of view of folk psychology and ordinary language. At the same time, it is not a logical contradiction, for the conjunction of true sentences is true and both conjuncts could be true (Shoemaker 1995). Moore-paradoxical sentences are odd, because our folk

¹³ This may be the reason why it is so hard for some ordinary language philosophers to follow the views that run contrary to common-sense intuitions. It seems to them that these views involve transformations of ordinary language-usage. They are right, of course, but it does not follow that such views are nonsensical.

psychological theory does not allow us to believe that x and to believe that not x at the same time.

According to folk psychology, the knowledge of one's own mind is direct. We believe that this knowledge comes directly from experience, whereas in a third-person case the knowledge must be indirect. This thesis requires qualification. The belief that knowledge is direct must be distinguished from the belief that one enjoys epistemic privilege *vis-à-vis* one's own mind as explicated in the thesis of first-person authority. I would argue that the first belief is plausible but wrong and that common-sense does not hold the second belief. In the literature concerning folk psychology, the extent to which the common sense is Cartesian is sometimes exaggerated. Although we commonly believe the knowledge of one's own mind to be privileged, it does not follow that we subscribe to the thesis of first-person authority. The first-person authority that common-sense attributes to itself is not an absolute authority – we acknowledge that our beliefs and reports fall from time to time under correction. Sometimes others know more about my mind than I do, sometimes they know less, sometimes I know as little as others, and it is not the case that I always know better. There are cases when we apply our background knowledge to correct others and ourselves, since their/our reports do not cohere with our background expectations.

The application of background generalisations need not be a conscious process. It may be compared with the application of grammar of the native language in which we all are experts. Such an expertise gives some credit to our belief that one's knowledge of one's own mind is direct (Gopnik 1993). The process may become conscious if there are not enough cues in the inner and outer environment to apply it to or if the tacit knowledge is inadequate to justify its application. In such cases the common sense gets in trouble and either suspends judgement or seeks professional help. These are the cases when the first-person is in the same position as the third-person (Cf. Bem 1972:5).

The plausibility of the common-sense belief that the knowledge of one's own mind is direct may also stem from the fact that in the first-person case, the experiential content forms a part of the background knowledge that directs the mental state ascriptions. No such cues are available in the third-person case.

The folk psychological view about first-person authority is not the Cartesian view as explicated above, it is more similar to the sociological model. According to it, the first-person authority is not a peculiar property of mental states. It is a property belonging to reports and it is possible to report about any sort of entity or situation (Cf. Rorty 1979:106). The report does not acquire its privileged status from the entity reported. Actually, there is no privileged kind of reports that would guarantee first-person authority. In this sense the infallibility or incorrigibility of the report is rather sociological than ontological. The report is incorrigible for me if I have no need to correct it, because it is true or if I do not have enough knowledge to correct it, but it does not mean that it is incorrigible in principle. The statement that there are 77 922 pets in Tartu is incorrigible for the

second reason. I suggest that one usually enjoys first-person authority in a similar sense. Sometimes, however, one does not enjoy any first-person authority, and these are the cases in which one makes mistakes and is being corrected.

At this point, the last argument against the Cartesian model of first-person authority can be presented, which states that its explanatory power is more restricted than that of the sociological model. If one assumes that the sociological model seems intuitively acceptable to you and if the intuitions of the Cartesian are just opposite, relying on his model, the latter cannot say that your intuitions are wrong, he cannot correct them, since it would contradict his incorrigibility thesis. The Cartesian cannot even explain how it is possible that our intuitions about the same thing (the mind) are different, for according to his model these intuitions arise from the very nature of a thing. He cannot even state that you do not have such intuitions, since according to his model he is not in a position to know that and you cannot be wrong in reporting about your intuitions. Relying on the sociological model of first-person authority, it is possible to explain both that sometimes our intuitions may turn out to be wrong and that different people may have different intuitions.

We do not enjoy Cartesian first-person authority about our own mind – the four theses explicated above proved to be wrong. This result, however, should not confront our common-sense intuitions. If one gives up the Cartesian presupposition that first-person authority is a necessary property of mental states, and if one thinks of it instead as the contingent property of reports, one gets the view of folk psychology. Thus-construed first-person authority does not ascribe absolute infallibility, incorrigibility, indubitability and omniscience to us, but states simply that normally we consider ourselves being infallible and incorrigible, while sometimes we do not consider ourselves to enjoy such epistemic privilege. These ascriptions are always open to revision. First-person authority, belonging to folk psychology, does not have ontological implications. Drawing upon it, one cannot build such metaphysical systems as has been done based on the Cartesian first-person authority.

Address:

Bruno Mölder
Mõisavahe 26–15
EE2400 Tartu
Estonia

Phone: 27/486230

E-mail: bruno@physic.ut.ee

References

- Alston, William (1971) "Varieties of Privileged Access". *American Philosophical Quarterly* 8, 3, 223–241.
- Armstrong, David M. (1968) *A Materialist Theory of the Mind*. London: Routledge & Kegan Paul.

- Audi, Robert (1974) "The Limits of Self-Knowledge". *Canadian Journal of Philosophy* 4, 2, 253–267.
- Aune, Bruce (1967) *Knowledge, Mind and Nature*. NY: Random House.
- Ayer, Alfred J. (1954) "Can There Be a Private Language?" *Supplementary Proceedings of the Aristotelian Society* 28, 63–76.
- Ayer, Alfred J. (1964) *The Concept of a Person and Other Essays*. London: Macmillan.
- Bem, Daryl J. (1972) "Self-Perception Theory." In *Advances in Experimental Social Psychology*, vol. VI. L. Berkowitz, ed. 1–62, New York & London: Academic Press.
- Block, Ned (1995) "On a Confusion About a Function of Consciousness". *Behavioral and Brain Sciences* 18, 2, 227–247.
- Cornman, James W. (1968) "Private Languages and Private Entities". *Australasian Journal of Philosophy* 46, 2, 117–126.
- Dancy, Jonathan (1985) *An Introduction to Contemporary Epistemology*. Oxford: Basil Blackwell.
- Davidson, Donald (1984) "First Person Authority". *Dialectica* 38, 2–3, 101–111.
- Dennett, Daniel C. (1971). "Intentional Systems". *Journal of Philosophy* 68, 4, 87–106.
- Dennett, Daniel C. (1991) *Consciousness Explained*. Boston: Little Brown.
- Doppelt, Gerald (1978) "Incorrigibility and the Mental". *Australasian Journal of Philosophy* 56, 1, 3–20.
- Ericsson, K. Anders & Herbert A. Simon (1980) "Verbal Reports as Data". *Psychological Review* 87, 3, 215–251.
- Geach, Peter (1971) *Mental Acts. Their Content and their Objects*. London: Routledge & Kegan Paul.
- Gopnik, Alison (1993) "How We Know Our Minds: The Illusion of First-Person Knowledge of Intentionality". *Behavioral and Brain Sciences* 16, 1, 1–14.
- Kenny, Anthony (1973) "Cartesian Privacy". In *The Anatomy of Soul*, 113–128. Oxford: Basil Blackwell.
- Kripke, Saul (1972) *Naming and Necessity*. Cambridge, Ma.:Harvard University Press.
- Lewis, David (1972) "Psychophysical and Theoretical Identifications". *Australasian Journal of Philosophy* 50, 3, 249–258.
- Malcolm, Norman (1963) *Knowledge and Certainty: Essays and Lectures*. Englewood Cliffs: Prentice Hall.
- Malcolm, Norman (1967) "The Privacy of Experience". In *Epistemology: New Essays in the Theory of Knowledge*. A. Stroll, ed. 129–158, New York, London: Harper & Row.
- Parsons, Kathryn P. (1970) "Mistaking Sensations". *The Philosophical Review* 79, 2, 201–213.
- Rorty, Richard (1965) "Mind-Body Identity, Privacy and Categories". *Review of Metaphysics* 19, 1, 24–54.
- Rorty, Richard (1970) "Incorrigibility as the Mark of the Mental". *Journal of Philosophy* 67, 12, 399–424.
- Rorty, Richard (1979) *Philosophy and the Mirror of Nature*. Princeton, NJ: Princeton University Press.
- Ryle, Gilbert (1949) *The Concept of Mind*. London: Hutchinson.
- Shoemaker, Sydney (1990) "First-Person Access". In *Philosophical Perspectives, Action Theory and Philosophy of Mind*. J. Tomberlin, ed. 187–214, Atascadero, CA: Ridgeview.
- Shoemaker, Sydney (1994) "Self-Knowledge and 'Inner Sense'". *Philosophy and Phenomenological Research* 54, 2, 249–314.
- Shoemaker, Sydney (1995) "Moore's Paradox and Self-Knowledge". *Philosophical Studies* 77, 2–3, 211–228.
- Smart, Jack J. C. (1970) "Brain Processes and Incorrigibility". In *The Mind-Brain Identity Theory*. C. V. Borst, ed. 107–109, London: Macmillan.
- Uus, Undo (1994) *Blindness of Modern Science*. Tartu: Tartu Observatory.
- Wittgenstein, Ludwig (1955) *Philosophical Investigations*. 3ed. G.E.M. Anscombe, ed. Oxford: Blackwell.