

Tero Vaaja

# The Problem of Other Minds: Themes from Wittgenstein



Tero Vaaja

The Problem of Other Minds:  
Themes from Wittgenstein

Esitetään Jyväskylän yliopiston yhteiskuntatieteellisen tiedekunnan suostumuksella julkisesti tarkastettavaksi yliopiston yliopiston Historica-rakennuksen salissa H320 syyskuun 19. päivänä 2015 kello 12.

Academic dissertation to be publicly discussed, by permission of the Faculty of Social Sciences of the University of Jyväskylä, in building Historica, hall H320, on September 19, 2015 at 12 o'clock noon.



UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2015

The Problem of Other Minds:  
Themes from Wittgenstein

JYVÄSKYLÄ STUDIES IN EDUCATION, PSYCHOLOGY AND SOCIAL RESEARCH 531

Tero Vaaja

The Problem of Other Minds:  
Themes from Wittgenstein



UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2015

Editors

Olli-Pekka Moisio

Department of Social Sciences and Philosophy, University of Jyväskylä

Pekka Olsbo, Sini Tuikka

Publishing Unit, University Library of Jyväskylä

URN:ISBN:978-951-39-6264-7

ISBN 978-951-39-6264-7 (PDF)

ISBN 978-951-39-6263-0 (nid.)

ISSN 0075-4625

Copyright © 2015, by University of Jyväskylä

Jyväskylä University Printing House, Jyväskylä 2015

## ABSTRACT

Vaaja, Tero

The Problem of Other Minds: Themes from Wittgenstein

Jyväskylä: University of Jyväskylä, 2015, 60 p.

(Jyväskylä Studies in Education, Psychology and Social Research

ISSN 0075-4625; 531)

ISBN 978-951-39-6263-0 (nid.)

ISBN 978-951-39-6264-7 (PDF)

The topic of this dissertation is the problem of other minds from the viewpoint of Wittgenstein's later philosophy. I distinguish themes from Wittgenstein's writings that involve first-person authority, expression, and perceiving others, and use these themes as tools to both acknowledge and intellectually handle the problem of other minds as a deeply human, living problem. I identify epistemic asymmetry, i.e. the assumption that direct experience provides each subject with unique and privileged knowledge of her own mental states, as the root issue of the problem of other minds. Understood this way, the problem of other minds is not merely a skeptical thought experiment, but a fundamental question about communicating and being with others. Article I identifies three themes in Wittgenstein's later thought, each of which deals with epistemic asymmetry in some form: 1) first-person authority and the contrasting uncertainty about the mental lives of others, 2) ineffability of immediate experience, and 3) perceiving the mentality of another person through her bodily behavior. In each of these themes, Wittgenstein can be seen as denying that the legitimate asymmetry between the subject of experience and others precludes us from knowing the experiences of others. Articles II through IV are developments of one or more of these themes, connected explicitly to Wittgenstein's philosophy to a varying extent. Article II deals with the so-called zombie argument in the context of physicalism/anti-physicalism debate in contemporary philosophy of mind, which is an important occurrence of the problematics of other minds in recent discussion. Article III elaborates the connection between expression and self-knowledge. Article IV explores Wittgenstein notion of criterion, asking whether perceiving bodily criteria of another person's mental states in Wittgenstein's sense can amount to perceiving the mental states themselves.

Keywords: other minds, subjectivity, intersubjectivity, consciousness, Wittgenstein

**Author's address** Tero Vaaja  
Department of Social Sciences and Philosophy  
University of Jyväskylä

**Supervisors** Mikko Yrjönsuuri  
Department of Social Sciences and Philosophy  
University of Jyväskylä

Sami Pihlström  
Faculty of Theology  
University of Helsinki

**Reviewers** Søren Overgaard  
Philosophy Section, Department of Media, Cognition  
and Communication  
University of Copenhagen

Olli Lagerspetz  
Department of Philosophy  
Åbo Akademi

**Opponent** Søren Overgaard  
Philosophy Section, Department of Media, Cognition  
and Communication  
University of Copenhagen

## ACKNOWLEDGEMENTS

I have been helped and influenced in writing this dissertation by more people than it is possible to enumerate. All the students and staff at the Department of Social Sciences and Philosophy I have had the pleasure of knowing would deserve a mention; here I can list only a selection of a few very important people.

My very special thanks go to Mikko Yrjönsuuri, in whose office I have always felt welcome, and who has trusted in my abilities even at moments when I did not trust them myself;

To Sami Pihlström, who has been a model of professionalism for me, and whose influence on me is probably greater than he can guess;

To Olli-Pekka Moisio, whose support and encouragement in innumerable discussions went far beyond the confines of usual academic advising, into a level that is possible only in true friendship;

To Antti Heikinheimo, my co-author in the Article II of this dissertation, the driving force behind our discussions on the philosophy of consciousness, and with his razor-sharp analytical wit, a shining example for me since our undergraduate years,

To Jarno Hietalahti, Onni Hirvonen, and Juhana Toivanen, who are both great colleagues and the kind of friends every person should have;

To a great number of friends, colleagues, tutors, and acquaintances who have helped in numerous ways over the years; I cannot possibly do justice to all of them, but my special thanks and greetings go to Jari Kaukua, Sofia Miguens, Rui Vieira da Cunha, João Santos, Tommaso Piazza, and Klaas Hellinga.

My work has been supported by grants from Finnish Cultural Foundation and its Central Finland Regional fund. Research and Innovation Office of the University of Jyväskylä allowed me to spend time abroad by a mobility grant, and University of Porto kindly hosted me. I want to express my deepest gratitude for this support.

Jyväskylä 26.8.2015

Tero Vaaja

## LIST OF ORIGINAL PUBLICATIONS

- I Vaaja, Tero. 2013. Wittgenstein's "Inner and Outer": Overcoming Epistemic Asymmetry. *Nordic Wittgenstein Review*, vol. 2, no. 1, 107-129.
- II Heikinheimo, Antti & Vaaja, Tero. 2013. The Redundancy of the Zombie Argument in *The Conscious Mind*. *Journal of Consciousness Studies*, vol. 20, no. 5-6, 6-26.
- III Vaaja, Tero. 2014. Expressivism, Self-Knowledge, and Describing One's Experiences. *Journal of Mind and Behavior*, vol. 35, no. 3, 151-166.
- IV Vaaja, Tero. Wittgensteinian Criteria and Perceiving Other Minds. To be submitted for publication.

## FIGURES

FIGURE 1 The structure of the dissertation as a flowchart .....	37
FIGURE 2 The structure of the dissertation as a pie chart.....	38
FIGURE 3 Avowals and first-person authority.....	48

## ABBREVIATIONS

CSM II = *The Philosophical Writings of Descartes, vol. II*. Translated by J. Cottingham, R. Stoothoff and D. Murdoch. Cambridge: The Cambridge University Press (1984).

CV = Wittgenstein, Ludwig (1998): *Culture and Value*. Edited by G.H. von Wright in collaboration with Heikki Nyman. Revised edition of the text by Alois Pichler. Translated by Peter Winch. Oxford: Blackwell.

*Essay* = Locke, John (1690/2001): *An Essay Concerning Human Understanding*. Kitchener: Batoche Books.

*Intellectual Powers I* = Reid, Thomas (1786): *Essays on the Intellectual Powers of Man, vol. I*. Dublin: Printed for L. White. Retrieved from archive.org [17.2.2015].

*Intellectual Powers II* = Reid, Thomas (1786): *Essays on the Intellectual Powers of Man, vol. II*. Dublin: Printed for L. White. Retrieved from archive.org [17.2.2015].

LW II = Wittgenstein, Ludwig (1992): *Last Writings on the Philosophy of Psychology: volume II*. Edited by G.H. von Wright and H. Nyman. Translated by C.G. Luckhardt and M.A.E. Aue. Oxford: Blackwell.

LWL = *Wittgenstein's Lectures, Cambridge 1930-1932*. From the notes of John King and Desmond Lee. Edited by Desmond Lee. Chicago: The University of Chicago Press (1980).

M = Moore, G.E. (1993): "Wittgenstein's Lectures in 1930-1933". In *Ludwig Wittgenstein: Philosophical Occasions 1912-1951*, edited by James C. Klagge and Alfred Nordmann, Indianapolis: Hackett Publishing Company, 46-114.

NFL = Wittgenstein, Ludwig (1993): *Notes for Lectures on "Private Experience" and "Sense Data"*. In *Ludwig Wittgenstein: Philosophical Occasions 1912-1951*, edited by James C. Klagge and Alfred Nordmann, Indianapolis: Hackett Publishing Company, 202-288.

OC = Wittgenstein, Ludwig (1969): *On Certainty*. Edited by G.E.M. Anscombe and G.H. von Wright. Translated by Denis Paul and G.E.M. Anscombe. Oxford: Blackwell.

PI = Wittgenstein, Ludwig (1953/2001): *Philosophical Investigations*. Third edition. Translated by G.E.M. Anscombe. Oxford: Blackwell.

*Principles* = Berkeley, George (1710/1996): *Principles of Human Knowledge*. In *Principles of Human Knowledge and Three Dialogues*, edited by H. Robinson, Oxford: Oxford University Press, 1-95.

RPP I = Wittgenstein, Ludwig (1980): *Remarks on the Philosophy of Psychology: volume I*. Edited by G.E.M. Anscombe and G.H. von Wright. Translated by G.E.M. Anscombe. Oxford: Blackwell.

*Treatise* = Hume, David (1739/1999): *A Treatise of Human Nature*. Kitchener: Batoche Books.

Z = Wittgenstein, Ludwig (1967): *Zettel*. Second Edition. Edited by G.E.M. Anscombe and G.H. von Wright. Translated by G.E.M. Anscombe. Oxford: Blackwell.

## CONTENTS

ABSTRACT

ACKNOWLEDGEMENTS

LIST OF ORIGINAL PUBLICATIONS

FIGURES

ABBREVIATIONS

CONTENTS

1	INTRODUCTION .....	13
	1.1 For the Reader .....	13
	1.2 The Overarching Research Task.....	15
2	BACKGROUND .....	18
	2.1 Historical overview .....	21
	2.2 A Wittgensteinian approach .....	32
3	DISSERTATION STRUCTURE .....	37
4	RESULTS .....	40
	4.1 Article I: "Wittgenstein's 'Inner and Outer': Overcoming Epistemic Asymmetry" .....	40
	4.2 Article II: "The Redundancy of the Zombie Argument in <i>The Conscious Mind</i> " .....	43
	4.3 Article III: "Expressivism, Self-Knowledge, and Describing One's Experiences" .....	46
	4.4 Article IV: "Wittgensteinian Criteria and Perceiving Other Minds" .....	49
5	DISCUSSION .....	52
	YHTEENVETO (FINNISH SUMMARY) .....	60

# 1 INTRODUCTION

## 1.1 For the Reader

The present work results from a combination of two interests: the later philosophy of Ludwig Wittgenstein (1889-1951), and the problem of other minds. It would be convenient to state that one of those is my subject matter, while the other one provides a background or a viewpoint for my treatment of the subject matter. Unfortunately, I cannot make that statement just like that, before some necessary explaining is done.

In philosophy, there are no other instruments of research than the investigator him- or herself. So, it seems appropriate and important to take some time to describe the human being who has been both the instrument and its operator – not for the sake of self-aggrandizement, but for the sake of giving an account of the tools by which the present project has been carried out.

As a remark of personal history, reading Wittgenstein's *On Certainty* at a certain point during my undergraduate studies sparked an interest in Wittgenstein in me, probably as much because of the feel of unfinished and living thinking that permeates the text (and Wittgenstein's later philosophy in general) as because of the substantial contents of *On Certainty*. It was also the first occasion, and so far more or less the only occasion, when I developed an interest in a single thinker, as opposed to more vaguely delineated "themes" or "problems". There has been an elusive and amorphous question that has been a basic philosophical problem for me; it could be put like: "How to express or communicate a feeling? Something inner, something that seems inescapably personal, subjective, and unique?" I am not sure when that question captured the most of my attention, but surely reading Wittgenstein's later works helped to fuel it. Maybe it is then not too hard to see how I ended up writing my Master's Thesis about the critique of solipsism in Wittgenstein's later writings<sup>1</sup>; I

---

<sup>1</sup> Originally, I suggested a comparison between Wittgenstein's treatment of solipsism in his early and later writings as the topic of my Master's Thesis. After being kindly informed by Professor Petri Ylikoski that the topic is a good one, but a good one for a PhD dissertation and not for a Master's Thesis, I narrowed it down. A sensible person would have, of course,

am not sure whether I decided on it before or after I learned that Sami Pihlström had written a monograph on solipsism. I wrote the Master's Thesis, and received a fairly good grade from the examiners, who were probably relieved that I succeeded in writing something remotely intelligible about the subject. An essay summarizing the main points of that work in Finnish was later published as a part of an anthology.

When I ended up planning a PhD dissertation, I guess I was craving for at least a bit less obscure subject, and I did not feel it was a good idea to define myself simply as a Wittgenstein researcher – I had just enough common sense to realize that the latter is a crowded niche. After a process that involved much more pure chance than conscious planning, I picked up the topic of other minds. Over a period of several years, I had wildly different conceptions of what the work would eventually look like, at some point planning a historical look into the problem and tracking different variations of it, relegating Wittgenstein to the background, and initially dropping his name from the title of my project. That proved futile; the way I realized its futility was by contemplating what objectivity could mean in the context of philosophical discussion. It could not mean basing one's work only on results confirmed by a majority of trustworthy colleagues; there never seemed to be enough agreement among philosophers to confidently take anything at a face value. Nor could it mean accepting only non-subjective, established facts, because there seems to be an endless controversy over what counts as a non-subjective established fact, depending on one's viewpoint. The most objective thing to do, apparently, is to acknowledge that one *has* a viewpoint.

In the end, it seems that all my thinking so far has happened in a Wittgensteinian context, in one way or another. When contemplation of virtually any interesting topic brings to one's mind passages from *Philosophical Investigations*, and when the first half-consciously posed question for any new text is "What would a Wittgensteinian say about this?", it is evident that one has assumed something like a theoretical viewpoint in philosophy. That is what is undeniably the case with me, although necessary qualifications about the "theoretical viewpoint" may be in order, given that Wittgenstein is a philosopher who is famously being thought to *not* offer any theories about anything. Maybe it is a methodological viewpoint rather than a theoretical viewpoint, if that phrase is not too vague to be informative. I like to think the offerings of the later Wittgenstein simply as a set of thinking tools. Then, admitting that one has a viewpoint, the minimal requirement of objectivity is to be explicit about the viewpoint; and the long-term goal must be to gain a comparative, critically reflective stance toward the viewpoint. At this stage, I hope I can demonstrate that I have at least filled the minimal requirement. I have (I think) also travelled some way in order to reach the long-term goal, but it is hard to say exactly how far.

According to one characterization, the skill of writing is to be able to create a context in which other people can think.<sup>2</sup> Wittgenstein has provided me with just

---

at least considered adopting the original idea as a perfect follow-up topic for the eventual PhD dissertation. I wonder what it is like to be a sensible person.

<sup>2</sup> This phrase has been attributed to Edwin Schlossberg in online sources (see e.g. [http://www.goodreads.com/author/quotes/30546.Edwin\\_Schlossberg](http://www.goodreads.com/author/quotes/30546.Edwin_Schlossberg) [4.12.2014]).

that – a context for thinking. One should recall Wittgenstein’s outspoken words in the preface of his *Philosophical Investigations*:

I should not like my writing to spare other people the trouble of thinking. But, if possible, to stimulate someone to thoughts of his own.<sup>3</sup>

While parts of this work (in particular Article I) consist of elucidating Wittgenstein’s thoughts, they are employed for the purpose of distinguishing themes for further development and discussion. I like to think that I am thus using Wittgenstein’s grand posthumous work in the way its author wished it to be used, regardless of whether the historical Wittgenstein would consider my developments to be worth anything.

Another major thing to note about the background of this dissertation is that although the later Wittgenstein can be characterized as standing in between the “analytic” and “continental” traditions, influencing both and drawing interest in both, the majority of the work on this thesis has happened in the context of analytic philosophy, widely conceived. This is a practical explanation (although probably not a *justification*) for the fact that the questions of mind-body dualism and the nature of self-knowledge receive much attention, but the rich discussion of intersubjectivity and empathy in the phenomenological tradition receives less. I am aware that this is a limitation.

So, is this dissertation simply about the problem of other minds as such, with Wittgenstein’s philosophy playing only the role of a background framework that provides a respectable intellectual heritage? Not quite; too little in this dissertation is explicitly about the problem of other minds as a thought-historical phenomenon, and too much of it is explicitly about Wittgenstein, for that claim to be plausible without qualification. The issue is that it is not clear what would count as “the problem of other minds as such” divorced from any particular way of looking at it – in my case, divorced from looking at it through Wittgenstein’s writings. The thinking tools employed here have the power to change one’s conception of the subject matter. After reading Wittgenstein, the problem of other minds does not look the same as before reading Wittgenstein – and that, it seems to me, is his essential contribution here. Maybe inevitably, writing about the problem of other minds from a Wittgensteinian perspective is to a large extent writing about Wittgenstein. At least it is so in this dissertation, whose research task I explicate in the next section.

## 1.2 The Overarching Research Task

The problem of other minds has the potential to address a concern that is pervasive in any human life that is at least minimally reflective. Human life involves fundamental practical questions about one’s knowledge about others and one’s relation to them. Even if we all normally take it for granted that other human beings have

---

<sup>3</sup> PI, x

minds, and happily contend that skepticism about them is an exercise only relevant in a philosophy class, there is a host of questions about the minds of others that cannot be brushed aside so easily.

- a) I know what I am thinking and feeling; in a way, I understand myself perfectly. I can see why I act in the way I do. But can I ever understand another person in the same way?
- b) There is a “world” behind my words and expressions. I know what it is like to be me; I know how the things I say and do flow out of my experiential world. How can I share that world with another?
- c) Others have sensations and feelings, but are they surely similar in me and in another person? Some differences we can detect, but could there not be undetectable ones?
- d) Human beings have minds, and we are quite confident that many other members of the animal kingdom have something more or less similar as well, but just how many do have it? How do we know how far down toward simpler life forms we have to ascend to find beings that are not “minded” anymore?
- e) Once a mind has existed, when can we be sure that it has ceased to exist (in a dying or a comatose patient)?
- f) I know what I mean by my words; I know what my words are intended to convey to others when I speak. But still others can wildly misinterpret my words, or fail to grasp any of their meaning, or the thought I tried to convey. Is all communication ultimately just guesswork of hidden intentions?

Sometimes, as in Anita Avramides’s *Other Minds*, the problem of other minds is introduced, and interest to it motivated, by citing questions like the aforementioned ones, which can occur to any intellectually curious person. But then, it is assumed that there is a further, distinct or special question, which is the philosophical problem of other minds:

What sort of problem is the problem of other minds? One very general response to this question is the following: The problem of other minds is a distinctively *philosophical* problem. [...] Where the non-philosopher assumes a community of minds who share a common world and then is struck by the question whether this community extends beyond the human sphere, the philosopher is struck by the possibility that even this human community is an illusion.<sup>4</sup>

When an attempt is made in philosophy to take an intellectual hold of a complex issue by refining it into a manageable sharp question (like the straightforward skept-

---

<sup>4</sup> Avramides, A. 2001. *Other minds*. London: Routledge, 1. Emphasis in the original.

tical question “How can we know that there are other minds, *at all?*”), one can get the feeling that the question that is being addressed is no more any of those questions that pervades human life. The connection to the everyman’s questions is lost, and thereby is lost the motivating force of those questions to evoke interest in a philosophical treatment of the problem of other minds. The everyman’s questions characterize a whole field of very concrete life problems of subjectivity and intersubjectivity; and it seems arbitrary to “refine” the philosophical problem of other minds to encompass anything less than that whole field.<sup>5</sup>

Wittgenstein’s later philosophy takes place in that large field of subjectivity and intersubjectivity. While acknowledging that reading Wittgenstein in such a broad context is by no means a neglected endeavor, I still find it worth stressing that Wittgenstein’s contribution to the theme of other minds goes beyond offering “a criterial solution” to the problem of other minds, or alternatively condemning the problem as “confused” from the start, or suggesting that the confidence in the existence of other minds belongs to the “basic certainties” explored in the remarks of *On Certainty*. Through Wittgenstein, it is possible to approach a host of questions about what it is like to be a human subject, who has both a unique first-person view on herself, and a modest, one-out-of-many interpersonal view on herself.<sup>6</sup> My task in this thesis is to give the problem of other minds, through Wittgenstein, a treatment that widens the understanding of both Wittgenstein and the problem.

The primary purpose of this dissertation is not to be a contribution to Wittgenstein exegesis. I hope I am developing a way to read Wittgenstein in a way that focuses on what is applicable, thought-inspiring and pragmatically interesting in his writings. The value of scholarly and exegetical work on Wittgenstein is immense; if the impressive commentary literature on Wittgenstein did not exist, it would have been impossible for me to do my own work. From my viewpoint, however, excessive attention to textual interpretation and locating Wittgenstein in a network of different “-isms” can make one blind to the charm, creative potential, and even existential value of his thinking. That is why this dissertation, ultimately, is not a thesis about Wittgenstein, but a thesis that picks up themes from Wittgenstein.

---

<sup>5</sup> In case it seems that I quote Avramides very unfavorably above, it has to be stressed that her account is by no means a crude one. She assumes that the non-philosopher can be made to see that it is worth taking the short step to posing a philosophical question about other minds. According to her, that philosophical question will eventually turn out to be not fundamentally an epistemological question, but a question about what a mind is, intimately connected with how we naturally respond to the behavior of other beings as subject-involving action. But her final verdict seems to be that, seen in the right light, the philosophical problem of other minds ceases to be a relevant question. And now, if it is assumed that the philosophical question is more or less independent of the non-philosophers’ questions, and the only goal of the former is to make itself vanish, then it will look like, from the perspective of a non-philosopher, just one of those esoteric questions that is of no interest to anyone outside a philosophy department. Or then it can be assumed that the vanishing of the philosophical question makes the non-philosophers’ questions vanish too; but this seems to me quite implausible, considering the way how the latter are involved with practical, moral and existential worries.

<sup>6</sup> Throughout this introduction and in the articles of this dissertation, my use of personal pronouns is intended to be gender-neutral. In the place of a gender-neutral pronoun, I sometimes use “he” and sometimes “she”.

## 2 BACKGROUND

In philosophy, the question of how it is possible to access another living being's experiential world finds expression in the so-called problem of other minds. The problem of other minds is an established name for a certain topic of discussion, but there hardly is a standard way of articulating it as a well-defined problem. A series of examples will serve:

The problem of other minds is the problem of how to justify the almost universal belief that others have minds very like our own.<sup>7</sup>

Even when we know everything physical about other creatures, we do not know for certain that they are conscious, or what their experiences are (although we may have good reason to believe that they are).<sup>8</sup>

[E]ach of us believes that he is not alone in the universe – that there are other beings who think and reason, hold beliefs, have sensations and feelings. And while a person can observe another's behavior and circumstances, he cannot perceive another's mental states.<sup>9</sup>

There is the doubtful, as they say, "privilege" of "direct access" to one's own mind, but the existence of that of the therefore possibly underprivileged other is supposed to remain another matter. This "other minds" is a confusion, a confluence of questions, but there are principally these: I have a mind: do I know whether others do and if I do how do I?<sup>10</sup>

What is the nature of the inferences that we all so commonly, and rightly, make from certain behavioral evidence to the mental lives of other people?<sup>11</sup>

[O]ne problem of other minds, at least, can be put like this. Being psychologically  $\psi$  is precisely what is recognized as instantiated in a person's own subjective experience of being  $\psi$ . This is the most basic source of her conception of what being  $\psi$  consists in, hence of

---

<sup>7</sup> Hyslop, A. 2014. Other Minds. The Stanford Encyclopedia of Philosophy (Spring 2014 Edition)

<sup>8</sup> Chalmers, D. J. 1997. The conscious mind: In search of a fundamental theory. Oxford University Press, USA, 102

<sup>9</sup> Plantinga, A. 1967. God and Other Minds. A Study of the Rational Justification of Belief in God. New York: Cornell University Press, 188

<sup>10</sup> Ziff, P. 1965. The simplicity of other minds. The Journal of Philosophy 62 (20), 575-584, 575

<sup>11</sup> Pargetter, R. 1984. The scientific inference to other minds. Australasian Journal of Philosophy 62 (2), 158-163, 158

what she means by 'being  $\psi$ ', hence of what it is to be  $\psi$  insofar as she understands this. So how could what it is to be  $\psi$  - again, insofar as she grasps what this is - possibly be detached entirely from her subjective experience? It must, rather, be tied in some way to that very experience. But then how is she supposed to make any sense at all of another's being  $\psi$ ?<sup>12</sup>

Traditionally, there are two problems of other minds: one epistemological, one conceptual. The epistemological problem asks how you can know, or how you can be justified in believing, that another person has a mind at all: that there exist other subjects of experience. The conceptual problem asks how you can so much as understand that there could exist other minds or subjects of experience: how you can have the concept of another's mind or experience.<sup>13</sup>

[T]he so-called "problem of other minds" is generally expressed in the question, "How do I know that other people have mental states *like my own*?"<sup>14</sup>

This set of examples is enough to illustrate the following:

- 1) There are at least two versions, or dimensions, of the problem. One **epistemological**, having to do with how we can know things (justifiably, certainly) about the minds of others; and one **conceptual**, having to do with how it is possible to come up with the idea of others as possible minded subjects in the first place,
- 2) The epistemological problem can be put as either the question whether we can know **if others have minds at all**, or as the question whether we can know **what their mental states are like**, how they feel from the "inside"; in particular, whether their mental states are like our own, and
- 3) What makes attaining such knowledge problematic is the assumed **lack of "direct access"** to the minds of others and/or the **impossibility to have direct perceptual knowledge** about the minds of others.

There is a wider issue which can, I think, be claimed to give rise to the problem in all these variations. I call it **epistemic asymmetry**. Knowing others appears problematic, because we take that there is an ideal default way of knowing about a state of mind, and that is our first-person knowledge about mental phenomena as they occur in us. Whatever knowledge we take ourselves to have of the minds of others, it is inferior in comparison with our knowledge of our own minds. Each of us is supposedly aware of the best possible way of knowing a mental phenomenon: experiencing that mental phenomenon oneself.

Descartes is a personification of such asymmetry. His *cogito* installed the indubitable knowledge of the existence of my thinking consciousness as the foundation

<sup>12</sup> Brewer, B. 2002. Emotion and other minds. *Understanding Emotions: Mind and Morals*. Aldershot: Ashgate, 23-36

<sup>13</sup> Pickard, H. 2003. V. Emotions and the Problem of Other Minds. *Royal Institute of Philosophy Supplement* 52 (1), 87-103, 87

<sup>14</sup> Carruthers, P. 2013. *The opacity of mind: an integrative theory of self-knowledge*. Oxford University Press, xiii

of knowledge in general. Even if its foundational status can be denied, it is still the paradigm of knowledge. The value of the *cogito* for Descartes was that it was the firm point that could stand even in the face of the most extreme attack of skepticism imaginable, an attack that throws into doubt the existence of the world with all its fellow creatures, and eventually allows restoring confidence in that world by a rational argument. If one agrees with the posterity that the point of the *cogito* is a good one, even if the argument for restoring confidence in the external world is not, then the existence of other minds, however firmly we believe in it, appears as embarrassingly badly justified. If minds are simply thinking things or consciousnesses, only contingently united with bodies in a living human being, then the best possible empirical experience is completely silent about the existence or qualities of other minds. This shortcoming is more striking, because it has a clear point of comparison: my certainty of the existence and qualities of my own mind.

Alec Hyslop has stated that the problem of other minds stems precisely from asymmetry in respect of knowledge: We have direct knowledge of our own mind, and we do not have the same kind of knowledge of the minds of others.<sup>15</sup> That is why it seems to require a special argument to show that our beliefs about the minds of others can be justified. But the asymmetry does not *only* raise the challenge of philosophical skepticism about other minds. As I see it, the asymmetry gives rise to a number of questions about the relations of subjects to others, questions that are not essentially connected with skeptical problems. They are (at least potentially) in place for someone who is unmoved by skeptical thought experiments, or holds that some satisfactory anti-skeptical argument can be given. In one way or another, all these questions stem from the idea that the first person perspective on a mind of a person is the privileged perspective. It is this wider field of problems of intersubjectivity, problems that are – I think – part of the human condition as people live it every day, that are the interesting subject matter of the problematics of other minds. This wider field of problems is what Wittgenstein is directing us to, and it is what informs the articles of this dissertation. Next, I will articulate what I take this wider field of problems to be like.

The statements of a subject about her own mental states have a peculiar authority. J.L. Austin wrote that whenever we are confronted with a question about what somebody feels, a “unique place is reserved for [the person’s own statement] in the summary of the facts of the case”, and in unusual cases when a person’s own statement about her feelings contradicts with what we were inclined to say about the matter, we cannot reject the person’s testimony without “feeling some uneasiness”.<sup>16</sup> Self-deception about one’s feeling is possible, especially in complex cases, but there are also cases where attributing error to a subject about her mental state seems next to senseless, for example, in the case of feeling pain. What is the nature of such first person authority, and in what kinds of cases can it be challenged?

Also, it seems inescapably intuitive to say that in some sense the subject of a feeling, sensation, or a phenomenal experience in general, knows it better than any-

---

<sup>15</sup> Hyslop, A. 1995. *Other minds*. Dordrecht: Kluwer Academic, see especially chapter one.

<sup>16</sup> Austin, J. L. 1961. *Philosophical papers*. Edited by Urmson, J. O. & Warnock, G. J. Oxford: Clarendon Press

one else can. Can one ever communicate a particular experience to others by words or other symbols, considering that the very usefulness of such symbols depends on them being abstract and general, while the subjective experiences of a person are unique and particular?

In cases where we engage in some kind of a scrutiny to find out what is in our mind, we say we introspect. So it seems that we have a privileged means of observing our own mind, which we lack in the case of others. In their case, we have to rely on external evidence, that is, on their testimony and their bodily conduct. But it may seem just an unfortunate contingency that this is the case. It is a recurring theme in fantasy fiction to imagine worlds where there would not be such reliance on external evidence, that is, worlds where magic or unknown technology could give us something like telepathic ability. What would it be like to be a telepath, and would it not be an enormously more fortunate situation than our actual one, in which we are denied access to the minds of others?

In the wide sense in which I approach the problem of other minds, it is a problem about this asymmetry. It is not a problem only for philosophers in a classroom. It is an intrinsic problem in being a human subject, and it finds expression equally in everyman's questions like those mentioned in section 1.1 above, and in philosophical thought-experiments. I find in the later Wittgenstein a philosopher who addresses that problem. His later thought is, among many other things, an attempt to face the asymmetry between a subject and others, and to put it in its proper place. But because the problem of other minds is also a classical topic of philosophical conversation that has a history, it is appropriate to summarize some of that history before elaborating on the Wittgensteinian approach. In Wittgenstein's hands, the problem of other minds transforms from a skeptical problem into a conceptual problem, and into a rich discussion of subjectivity, oneself and others. But to appreciate all that, it must be seen how the traditional skeptical problem was formed, and then see how Wittgenstein's thought will look like against that background.

## 2.1 Historical overview

A passage of Descartes's *Second Meditation* is often referred to as the entrance of the problem of other minds to the stage of philosophy.<sup>17</sup> There, Descartes initially discusses how we know a body such as a piece of wax, asking whether it is by means of the senses that we know the wax to be staying the same wax throughout the radical changes it undergoes when heated. Ordinary language, according to Descartes, suggests that we know such things by seeing them. Then,

[...] this might lead me to conclude without more ado that knowledge of the wax comes from what the eye sees, and not from the scrutiny of the mind alone. But then if I look out of the window and see men crossing the square, as I just happen to have done, I normally say that I see the men themselves, just as I say that I see the wax. Yet do I see any more

---

<sup>17</sup> This section (2.1) is much indebted to Avramides's *Other Minds* (2001), the best historical exposition of the problem I know.

than hats and coats which could conceal automatons? I *judge* that they are men. And so something which I thought I was seeing with my eyes is in fact grasped solely by the faculty of judgement which is in my mind.<sup>18</sup>

The passage is only a detail in a long argument, and its reference to human beings seems quite accidental. Anyway, what Descartes requires in this stage of the *Second Meditation* is an example of a situation where we need a judgment that goes beyond what is available to senses to tell that something is the case, even though in ordinary language we carelessly say that we *see* it to be the case. The example he comes up with is that of judging that there are men on the street while seeing hats and coats from his window. The passage by itself does not say that we need an act of the faculty of judgment, something that goes beyond what is available to senses, also when we see human bodies without hats and coats, in order to tell that they are not automatic machines. But the latter claim seems inevitable, given the distinction between mind and body introduced by Descartes's dualism.<sup>19</sup>

Notwithstanding the prominence of skepticism in the ancient Greek philosophy, there is very little in the Greek tradition that could be taken as a formulation of the problem of other minds. Voula Tsouna assesses some candidate examples, and concludes that they cannot be taken as anticipations of the problem of other minds in its modern form.<sup>20</sup> Although the idea that perceptions or thoughts of another are unknowable finds expression in the antiquity, it always seems to be stated against the background contention that others *have* thoughts and perceptions, though these are private to each subject. According to Sextus Empiricus, the Cyrenaics held that perceptions are private to each subject and cannot be accessed by another person, but this features as a part of the rationale for withholding judgment about the nature of objects. Because it is impossible to intersubjectively compare perceptions to assess their quality, no person's perceptions have a better claim to truth than those of any other person, and it is best to withhold judgment about what the objects of the world are really like. Far from doubting the existence of the perceptions of others, this skeptical argument *requires* that there is a multiplicity of perceiving subjects.

One of the reasons given by Tsouna for why the existence of other minds did not fall inside the scope of ancient skepticism is that the Greeks did not subscribe to a mind-body divide in a form in which it is known to modern philosophy. Furthermore, even though the view that a person knows his own subjective states infallibly was familiar to the Greeks, the certainty about one's subjective states was not yet juxtaposed with any contrastive uncertainty about the existence of the world outside oneself. Descartes brought these views onto the stage of modern philosophy, and because of that, he can be considered an originator of the problem of other minds. The passage about the hats and coats by itself is quite incidental. More important is that in seeking a conclusive answer to the skeptical arguments inherited from antiquity, he finds the starting point of self-evident truth in the fact of his own subjective

---

<sup>18</sup> CSM II, 21

<sup>19</sup> For discussion, see Avramides, A. 2001. *Other minds*. London: Routledge, Avramides, A. 2010. *Knowledge of other minds*. *The Routledge Companion to Epistemology*, 433

<sup>20</sup> Tsouna, V. 1998. Remarks about Other Minds in Greek Philosophy. *Phronesis* 43 (3), 245-263

experience. Now the contrast between the knowledge of one's own thinking mind, which is self-evident, and knowledge of everything outside it, which is open to doubt until it is reconstructed by rational argument, becomes important. And, because for Descartes the conceivability of the thinking mind apart from matter showed it to be separable from body, even restoring our faith in the reliability of our sense perceptions of material bodies will not give us knowledge of the minds of others. The latter are not reachable by the senses.

Descartes himself did not recognize a special problem in knowing other minds, apparently satisfied with practical certainty that God would not fail to unite a thinking mind with each and every living human body. It was Thomas Reid who pointed it out as a problem engendered by Cartesian philosophy. In the words of James Somerville, "the traditional problem virtually begins with [Reid] ... The very words 'other minds' first occur extensively in his pages"<sup>21</sup>. Reid was concerned with the threat of skepticism whose source he identified in an unquestioned assumption by the preceding philosophy from Aristotle to Descartes to empiricists: the assumption that knowledge of the world around us and of ourselves is mediated by ideas.

Reid saw it as a deep error in Descartes that he allowed no other ultimate legitimate foundation of knowledge than what is self-evidently given to consciousness. Where for Descartes the reliability of the senses was something that needed to be proved by a rational argument, Reid proposed that the senses are a legitimate source of knowledge in their own right, and that their objects are the worldly objects themselves, not ideas of worldly objects. According to Reid, it belongs to the first principles of thought, belonging to the natural constitution of human beings, that distinctly perceiving objects in one's surroundings makes the existence of those objects self-evident, unless there is a special reason to distrust the senses in a particular case.

The need of laborious proof for the reality of the external world as it is presented to us via ideas is one part of the problematic legacy of Cartesianism. But another part of that legacy is that we seem unable to have as much as ideas of the minds of others. The view that the minds of others are "invisible" and out of the reach of perception was taken for granted after Descartes by Locke<sup>22</sup>, Berkeley<sup>23</sup>, Hume<sup>24</sup>, and

---

<sup>21</sup> Somerville, J. 1989. Making out the signatures: Reid's account of the knowledge of other minds. In *The Philosophy of Thomas Reid*, edited by Dalgarno and Matthews, Dordrecht: Kluwer, 249-273

<sup>22</sup> "Man, though he have great variety of thoughts, and such from which others as well as himself might receive profit and delight; yet they are all within his own breast, invisible and hidden from others, nor can of themselves be made to appear. The comfort and advantage of society not being to be had without communication of thoughts, it was necessary that man should find out some external sensible signs, whereof those invisible ideas, which his thoughts are made up of, might be made known to others." *Essay*, III, 2, 1, p. 328

<sup>23</sup> "A spirit is one simple, undivided, active being: as it perceives ideas, it is called the *understanding*, and as it produces or otherwise operates about them, it is called the *will*. Hence there can be no idea formed of a soul or spirit [...] Such is the nature of *spirit* or that which acts, that it cannot be of itself perceived, but only by the effects which it produceth. [...] [T]he words *will*, *soul*, *spirit*, do not stand for different ideas, or in truth, for any idea at all, but for something which is very different from ideas, and which being an agent cannot be like unto, or represented by, any idea whatsoever." *Principles*, sect. 27, p. 35

<sup>24</sup> "When any affection is infus'd by sympathy, it is at first known only by its effects, and by those external signs in the countenance and conversation, which convey an idea of it." *Treatise*, 2.1.11.3, p. 220

also Reid<sup>25</sup>, although the latter of course did not speak in terms of ideas. Berkeley thought his philosophy could eradicate skepticism about external world by denying that the supposition of a world of material substance “external” to mind makes sense in the first place. Reid, however, observed that Berkeley’s way of securing the existence of tables, chairs and houses came at a price of losing something even more crucial. Of Berkeley’s philosophy, Reid says that he (Reid) “once believed this doctrine of ideas so firmly, as to embrace the whole of Berkeley’s system in consequence of it; till, finding other consequences to follow from it, which gave me more uneasiness than the want of a material world”<sup>26</sup>. At least one of the grave consequences Reid saw was that other subjects become detached and unknowable:

[T]here is one uncomfortable consequence of [Berkeley’s] system, which he seems not to have attended to, and from which it will be found difficult, if at all possible, to guard it. The consequence, I mean, is this, that, although it leaves us sufficient evidence of a supreme intelligent mind, it seems to take away all the evidence we have of other intelligent beings like ourselves. What I call a father, a brother, or a friend, is only a parcel of ideas in my own mind; and being ideas in my mind, they cannot possibly have that relation to another mind which they have to mine, any more than the pain felt by me can be the individual pain felt by another. I can find no principle in Berkeley’s system, which affords me even probable grounds to conclude, that there are other intelligent beings, like myself, in the relations of father, brother, friend, or fellow-citizen. I am left alone, as the only creature of God in the universe, in that forlorn state of *egoism*, into which it is said some of the disciples of Des Cartes were brought by his philosophy.<sup>27</sup>

Reid realizes that this problem does not arise only as an uninteresting corollary to skepticism about external world in general. It arises as a separate problem for Berkeley, whose philosophy was geared to remove philosophical doubt about an external world.

To avoid the skeptical absurdities that Reid took to follow from the philosophies of his predecessors, Reid thought it necessary to deny the fundamental assumption that he identifies in them: the assumption that knowledge is founded on ideas, which are the immediate objects of perception. Instead, he proposed that the legitimate basis of knowledge is much broader, and the senses as such, directly perceiving the outside world, have just as obvious (although fallible) claim to be sources of knowledge as the deliverances of ideas in one’s consciousness. Among his first principles of knowledge, “propositions which are no sooner understood than they are believed” and admitting of no proof in terms of more original convictions than

---

“No passion of another discovers itself immediately to the mind. We are only sensible of its causes or effects. From these we infer the passion: And consequently these give rise to our sympathy.” *Treatise*, 3.3.1.7, p. 386

<sup>25</sup> “When we see the sign, and see the thing signified always conjoined with it, experience may be the instructor, and teach us how that sign is to be interpreted. But how shall experience instruct us when we see the sign only, when the thing signified is invisible? Now this is the case here; the thoughts and passions of the mind, as well as the mind itself, are invisible, and therefore their connection with any sensible sign cannot be first discovered by experience; there must be some earlier source of this knowledge. [...] [W]hen I see the features of an expressive face, I see only figure and colour variously modified. But, by the constitution of my nature, the visible object brings along with it the conception and belief of a certain passion or sentiment in the mind of the person.” *Intellectual Powers* II, essay VI, ch. 5, pp. 289-290

<sup>26</sup> *Intellectual Powers* I, essay II, ch. 10, p. 189

<sup>27</sup> *Intellectual Powers* I, essay II, ch. 10, pp. 197-198

themselves<sup>28</sup>, he includes the principles that “there is life and intelligence in our fellow-men with whom we converse”<sup>29</sup> and that “certain features of the countenance, sounds of the voice, and gestures of the body, indicate certain thoughts and dispositions of mind”<sup>30</sup>. These convictions, Reid observes, are part of a child’s natural disposition long before the child is capable of reasoning, and they are a precondition of being introduced into the practice of reasoning in the first place. So, he contends, the belief in the minds of others “stands upon another foundation than that of reasoning”<sup>31</sup>, and qualifies as a first principle.

The so-called argument from analogy, around which much of the philosophical discussion about other minds revolves, was famously put forward by John Stuart Mill, whose passages in his *An Examination of Sir William Hamilton’s Philosophy* is often cited as the *locus classicus* of that argument.<sup>32</sup> For Mill, the argument was a reaction to Reid’s conclusions. The empiricist Mill held a version of the view that all knowledge stems from what is directly given to a subject in perception. He refers to Reid’s protest that such a view makes it unintelligible how we can know anything about other minds, and insists that Reid’s worry is unfounded. According to Mill, the existence of other minds can well be proven in his system, which postulates minds only as series of states of consciousness, and the notion of matter amounting only to the “permanent possibility” of sensation.

I conclude that other human beings have feelings like me, because, first, they have bodies like me, which I know, in my own case, to be the antecedent condition of feelings; and because, secondly, they exhibit the acts, and other outward signs, which in my own case I know by experience to be caused by feelings. I am conscious in myself of a series of facts connected by a uniform sequence, of which the beginning is modifications of my body, the middle is feelings, the end is outward demeanor. In the case of other human beings I have the evidence of my senses for the first and last links of the series, but not for the intermediate link. I find, however, that the sequence between the first and last is as regular and constant in those other cases as it is in mine. In my own case I know that the first link produces the last through the intermediate link, and could not produce it without. Experience, therefore, obliges me to conclude that there must be an intermediate link; which must either be the same in others as in myself, or a different one: I must either believe them to be alive, or to be automatons: and by believing them to be alive, that is, by supposing the link to be of the same nature as in the case of which I have experience, and which is in all other respects similar, I bring other human beings, as phenomena, under the same generalizations which I know by experience to be the true theory of my own existence. [...] We know the existence of other beings by generalization from the knowledge of our own: the generalization merely postulates that what experience shows to be a mark of the existence of something within the sphere of consciousness, may be concluded to be a mark of the same thing beyond that sphere.<sup>33</sup>

The existence of the external world of material bodies is not a problem Mill sees necessary to address. Minds are series of states of consciousness, and Mill takes himself to have explained enough when he has explained how our notion of external materi-

<sup>28</sup> *Intellectual Powers* II, essay VI, ch. 4, p. 238

<sup>29</sup> *Intellectual Powers* II, essay VI, ch. 5, p. 284

<sup>30</sup> *Intellectual Powers* II, essay VI, ch. 5, p. 287

<sup>31</sup> *Intellectual Powers* II, essay VI, ch. 5, p. 285

<sup>32</sup> E.g. by Malcolm, N. 1958. Knowledge of Other Minds. *The Journal of Philosophy* 55 (23), pp. 969-978

<sup>33</sup> Mill, J. 1872. *An Examination of Sir William Hamilton’s Philosophy*. Fourth edition. London: Longmans, Green, Reader, and Dyer, Retrieved from archive.org [17.2.2015], 243-244

al bodies is generated on the basis of those states of consciousness, namely, as the notion of permanent, systematic possibilities of that kind of conscious states that we call perceptions of external objects. But how can we know that there are other series of states of consciousness, that is, other minds? This is something that Mill at least sees necessary to address, and he goes on to claim, by the argument from analogy, that the “real externality” of other minds is capable of proof.<sup>34</sup>

More important than the argument’s merits in Mill’s context is the way it has become a standard answer to the problem of other minds, undoubtedly because it is in a sense highly intuitive. We have bodies, and from our own case we know how modifications of our bodies are followed by sensations, feelings and other mental phenomena. Also, we know from our own case how our bodily behavior follows from those mental phenomena. We observe other bodies in the world similar to our own, and we observe in them similar regular patterns of bodily modifications and subsequent behavior as in our own case. We cannot observe the middle link – the mental phenomena – in others, but it is perfectly reasonable and plausible to argue that there middle link must be present in their case also, given that it is constantly and regularly present in my own case. The argument from analogy provides as good a rational reason for our beliefs about other minds as can ever be attained, and that should be enough.

Dissatisfaction is often voiced about the argument from analogy on the grounds that it is so far removed from both the phenomenology of encountering other people, and implausible as an account of how interpersonal understanding actually comes about. In the 18<sup>th</sup> century, Reid noted the fact that infants are socially engaged with their caretakers from a very early age, and deems incredible the claim that this would be something they have to learn by reasoning that the things happening in the body of the caretaker probably indicate workings of a mind in them. Reid claims it to be a part of the natural constitution of a human being that we perceive other human bodies as minded. In the early 20<sup>th</sup> century, C.D. Broad ridiculed the view that a child reaches her beliefs about the minds of others by analogical argument, deeming the view “too silly to need refutation”, adding that “[i]f the belief in other minds and other mental events were reached in this way, it might perhaps be entertained as a bold speculative opinion by a few exceptionally ingenious and observant persons at the ripe age of thirty-five”<sup>35</sup>. And in contemporary philosophy, David Chalmers, as an example of a philosopher who subscribes to something like the argument from analogy<sup>36</sup>, admits that the argument is primarily a way to rationally justify our beliefs in other minds, not necessarily something that is actually employed in our intersubjective relations<sup>37</sup>.

---

<sup>34</sup> Mill, J. 1872. *An Examination of Sir William Hamilton’s Philosophy*. Fourth edition. London: Longmans, Green, Reader, and Dyer, Retrieved from archive.org [17.2.2015], 238, footnote \* (See also Avramides 2001, 164-171)

<sup>35</sup> Broad, C. D. 1925. *The Mind and its Place in Nature*. London: Kegan Paul, Trench, Trubner & Co, 324-325

<sup>36</sup> See Article II, 15-17.

<sup>37</sup> Chalmers, D. J. 1997. *The conscious mind: In search of a fundamental theory*. Oxford University Press, USA, 246

However, the argument can be interpreted charitably as a philosophical argument offering (just) a rational justification for beliefs about other minds, instead of a psychological account of what happens in the minds of subjects when they encounter others. As such, it became a commonplace in philosophy, particularly in the discussions of influential authors around the mid-1900's; not in the sense that everyone would have accepted it, but in the sense that it tends to get mentioned as the basic position or a starting point of critical discussions. It still has that role in encyclopedia articles and overviews, and one of the rare relatively recent analytical monographs dedicated to the problem of other minds, Alec Hyslop's *Other Minds*<sup>38</sup>, is an extended defense of the argument.

As a standard argument, the argument from analogy also has standard objections. First, it (arguably) is a generalization based on a single case; I can only verify the link between mental phenomena and bodily behavior in my own case. Second, it is an inference to an uncheckable conclusion; the kind of verification that I have for the link in my own case cannot be had in the case of others.<sup>39</sup> These objections make explicit two background assumptions that can be called the Cartesian assumptions giving rise to the problem of other minds. According to the first Cartesian assumption, we know mental phenomena primarily from our own case, that is, from the way they appear to us in consciousness. According to the second Cartesian assumption, the minds of others cannot be perceived, except in a mediate way, inferred from the bodily behavior of others. Given these assumptions, the argument from analogy suggests itself as the most satisfactory way of justifying our beliefs in other minds. The argument says we know the minds of others by an analogy with our own mind, of which we are taken to have unproblematic prior knowledge. Also, it is assumed we know our own states of consciousness to be the causes of our bodily behavior, and the argument allows us to infer that the same kind of causes very probably operate unperceived in others, insofar as their bodies display the same regularities of behavior as our own. The manifest fact that the argument of analogy can at best establish that others very *probably* have minds – even though the probability may be exceedingly high – can be thought to be one more cause of dissatisfaction in its own right. It may be true that no argument can ever refute a skeptical argument with absolute certainty anyway, but in the case of other minds this shortcoming is particularly shocking. Given the constitutive role that being with others has for all human life, it may well seem unacceptable that our beliefs about others as thinking and feeling beings could have in the last analysis no other fundamental justification than a probabilistic inference.<sup>40</sup> As F.H. Bradley put it, “we don't want inferred friends who are mere hypotheses to explain physical phenomena”<sup>41</sup>.

Later discussions, unsurprisingly, have usually proceeded by suspecting that there is something fundamentally wrong with the way the problem is stated. It is an attraction of behaviorism, in general, that it gives voice for the anti-Cartesian intuition that mental phenomena are not mysterious hidden entities, but rather publicly

---

<sup>38</sup> Hyslop, A. 1995. *Other minds*. Dordrecht: Kluwer Academic

<sup>39</sup> The problematics around the argument from analogy are referred to in Article II, 16-17.

<sup>40</sup> See also Article II, 16.

<sup>41</sup> Quoted through Avramides, A. 2001. *Other minds*. London: Routledge, 307n101

observable. For logical positivism, traditional metaphysical questions about whether a given type of entity “really exists” were to be discarded and replaced by analysis of language; the problem of the existence of other minds would be one such dispensable pseudo-problem. If claims about other minds have any meaning at all, their meaning must reside in the empirical content of such claims, that is, on the way in which they can (in principle) be translated into empirically testable claims about behavior. The conclusion of the argument of analogy, that there “really exist” other minds which, however, lie beyond the reach of any observation from the outside, is not a meaningful statement, and so the argument that leads to it becomes irrelevant.

However, a position that tries to condemn the problem of other minds as senseless on the grounds that the *meanings* of statements about the mental phenomena of others are exhausted by the ways in which we confirm and deny such statements based on observable behavior has to answer to a problem. If we accept that statements about other minds are ultimately reducible to observation statements about their behavior (in the manner of Carnap’s reduction of the “heteropsychological” to the “autopsychological”), what is the status of those statements by which we report our observations, that is, our own sense experience? The same analysis does not apply to them, so it seems that the meaning in which statements about mental phenomena are employed in the case of others is radically different from the meaning in which they are employed in the case of oneself.

This is connected to the first Cartesian assumption motivating the argument from analogy, that is, the assumption that we know mental phenomena primarily from how they appear to us in consciousness in our own case. Another way of arguing that the argument from analogy is faulty in its very starting point is to claim that once this assumption is in place, the terms of the argument become unintelligible. As long as the only way in which I know mental phenomena is to encounter them in my direct subjective experience, what gives rise to the hypothesis that there might be mental phenomena that are *not* parts of my direct subjective experience, that is, the hypothesis that the argument from analogy is meant to confirm? The argument assumes that we justify our beliefs about others based on an analogy with our own case; but this requires that the arguer already has a conception of mental phenomena as something that can occur without her consciousness. Prior to an understanding that there is a range of potential subjects of experience in the world, it is unclear what could give a subject a conception of her states of consciousness as something she is the subject of, and thus an understanding of “her own case”. Thus, to be intelligible, the argument from analogy has to presuppose what it attempts to prove.

Many of Wittgenstein’s remarks in his later philosophy emphasize that what is crucial about the argument from analogy is not its epistemic persuasiveness, but the question whether the starting point of the argument makes sense.

What gives us *so much as the idea* that beings, things, can feel?<sup>42</sup>

If one has to imagine someone else’s pain on the model of one’s own, this is none too easy a thing to do: for I have to imagine pain which I do not feel on the model of the pain which I do feel. That is, what I have to do is not simply to make a transition in imagination from

---

<sup>42</sup> PI §283

one place of pain to another. As, from pain in the hand to pain in the arm. For I am not to imagine that I feel pain in some region of his body.<sup>43</sup>

“But if I suppose that someone has a pain, then I am simply supposing that he has just the same as I have so often had.” — That gets us no further. It is as if I were to say: “You surely know what ‘It is 5 o’clock here’ means; so you also know what ‘It’s 5 o’clock on the sun’ means. It means simply that it is just the same there as it is here when it is 5 o’clock.” — The explanation by means of identity does not work here. For I know well enough that one can call 5 o’clock here and 5 o’clock there “the same time”, but what I do not know is in what cases one is to speak of its being the same time here and there.<sup>44</sup>

With this development, the focus is turned from the traditional skeptical problem into a conceptual problem of other minds. The latter is not primarily about justifying uncertain beliefs about other minds, but about the preconditions of being able to form such beliefs in the first place. The Cartesian assumptions, insofar as something like the argument from analogy is what they ultimately lead to, make it questionable how an intersubjective conception of mind should be possible at all. Now the question is also linked to the related issue of solipsism. There is a prior question to skeptical doubts about other minds: What accounts for the fact that we can as much as formulate the doubts?

At one time, when his views were interacting with those developed in the Vienna Circle, Wittgenstein held a view that propositions pertaining to one’s own states of mind were verified by direct experience, but that experience itself was not properly described as *subjective*, or subject-involving. Direct experience as such should be thought as in a way neutral. It does not come with a possessor; sensations and feelings are not “had” by a subject in a way coins or matchboxes are had by persons, although this point is obscured by our using locutions like “I have a pain” or “I have a red sensation”. It is a peculiarity of a proposition like “I have a pain” that I need no other verification for it than the direct experience, and in this sense such propositions have a special status. But the subject-indicating “I” in such a proposition is redundant; the function of the proposition is not to speak about a particular person, but merely to indicate direct experience. “I have a pain”, which mentions a subject, is only an alternative form of expression for “It hurts”, which does not. This stands in contrast with propositions like “I have a matchbox”, where the function of the proposition is to speak of a particular person, and to denote an owner. The first person pronoun, therefore, should be said to have two different uses, one where its proper function is to mention one particular person among many, and one where it is a logically dispensable part of a direct experience report. Wittgenstein’s condemnation of metaphysical solipsism as not merely false but senseless was partly based on these considerations. It suggests a direction out from the conceptual problem generated by the Cartesian assumptions: It is not the case that I am primitively acquainted with *my* subjective experience and my subjective experience only. I am acquainted with direct experience as such, without being primitively acquainted with it as *my* experience. Direct experience as such is neutral regarding a subject, and my conception of myself as an individual – the conception of “my own case” – must thus

---

<sup>43</sup> PI §302

<sup>44</sup> PI §350

be dependent rather on the ways in which I identify my physical body as one body among others.

According to the verificationist view shared by Wittgenstein and logical positivists at the time, the meaning of a proposition is given by its method of verification. Thus, a proposition like "N.N. feels pain" is understood insofar as we know what counts as a confirmation or a refutation of it, and these confirmations or refutations are given in terms of observable bodily states and verbal reports of the other, and cannot intelligibly be given in any other way. Questions about mental phenomena in others are not ultimately unanswerable, as skeptical arguments threaten to show; rather, their very meaningfulness requires them being answerable. But this line of thought comes back to a problem mentioned earlier: It implies that ascriptions of experience, when they are done in the first person, have a different meaning from corresponding ascriptions when experience is ascribed to someone else. This is because the method of verification is different in the two cases, which implies a difference in meaning. Ascriptions of experience to N.N. can be ultimately reduced to empirical statements about N.N.'s behavior, but when N.N. herself reports an experience, she is not reporting behavior, but indicating direct experience. The method of verification in the third-person case, which requires identifying a certain body as N.N.'s body and observing what that body does, is out of place in the first-person case. "I feel pain" needs no verification apart from the direct experience itself. If this is so, then, according to the verificationist principle, "I feel pain" said by N.N. and "N.N. feels pain" said by another person do not share a meaning. But presumably, when I say of N.N. that she feels pain, what I want to say of her is exactly what she says when she reports her pain. I would not accept the claim that only she can talk directly about this instance of pain experience, while I am talking about her verbal and bodily behavior. P.F. Strawson in his *Individuals* put the conundrum thus:

[Consciousness-ascribing] phrases are used in just the same sense when the subject is another as when the subject is oneself. Of course the thought that this is so gives no trouble to the non-philosopher: the thought, for example, that "in pain" means the same whether one says "I am in pain" or "He is in pain". The dictionaries do not give two sets of meanings for every expression which describes a state of consciousness: a first-person meaning and a second-and-third person meaning. But to the philosopher this thought has given trouble. How could the sense be the same when the method of verification was so different in the two cases — or, rather, when there was a method of verification in the one case (the case of others) and not, properly speaking, in the other case (the case of oneself)?<sup>45</sup>

Strawson's suggestion was to recognize that there is a primitive concept of *person*. As the concept was used by Strawson, a person is an entity who serves as the subject of both consciousness-ascribing predicates and predicates that apply to physical bodies. In our thought and language, Strawson holds, it is manifest that we say, for example, "I have a burned hand" and "I feel pain", and in both propositions we refer by "I" to a person. What makes the concept of person important is that by it we understand ourselves and others as subjects of both physical and mental qualities unambiguously. In *Individuals*, Strawson took it only as a secondary task to suggest what in the

---

<sup>45</sup> Strawson, P. F. 1959. *Individuals. An Essay in Descriptive Metaphysics*. London and New York: Routledge, 99-100

natural facts gives rise to the concept of person; more important was its logical primitiveness. Strawson argues in a subtle way for the claim that the concept of person cannot be analyzed in terms of something more fundamental, in particular, not in terms of a Cartesian ego as a subject of consciousness-ascribing predicates, and a physical body as a subject of physical predicates. Our ability to think about ourselves as pure thinking subjects and physical bodies is dependent on our first having the concept of person, not the other way round. Strawson is in effect addressing the conceptual problem of other minds. He states that his argument is not directed to refuting skepticism about other minds, but its conclusion is a necessary condition for the skeptical problem to be intelligible. If one considers minds separable from bodies, a probably insoluble problem of other minds may ensue, but in order to have a conception of oneself as a thinking mind, one has to first have a concept of person that does not recognize the mind-body divide.

Where for Strawson the primary entity that has states of consciousness (in both the first person and third person case) is a person, Wittgenstein of *Philosophical Investigations* talks about *human being* in a similar context.

[O]nly of a living human being and what resembles (behaves like) a living human being can one say: it has sensations; it sees; is blind; hears; is deaf; is conscious or unconscious.<sup>46</sup>

What gives us *so much as the idea* that beings, things, can feel? [...] Only of what behaves like a human being can one say that it *has* pains.<sup>47</sup>

Suppose I say of a friend: "He isn't an automaton". – What information is conveyed by this, and to whom would it be information? To a *human being* who meets him in ordinary circumstances? What information *could* it give him? (At the very most that this man always behaves like a human being, and not occasionally like a machine.) [...] My attitude towards [the friend] is an attitude towards a soul. I am not of the *opinion* that he has a soul.<sup>48</sup>

The latter passage is soon followed in *Philosophical Investigations* II, iv, by an aphorism which occurs in Wittgenstein's writings in two variants:

The human body is the best picture of the human soul.<sup>49</sup>

The human being is the best picture of the human soul.<sup>50</sup>

The concepts having to do with consciousness and mentality are such that they are correctly applied to creatures that behave more or less like human beings; that is, beings who are manifestly like us. The most fundamental question can be said to be: How do we manage to relate to others as thinking and feeling beings? And the later Wittgenstein's way of dealing with this problem was not to give an epistemological argument, but to carefully attend to how concepts and statements in this domain of human life are actually used. Wittgenstein's remarks suggest a general point that the uses of those concepts and statements cannot be based on anything else than human

---

<sup>46</sup> PI §281

<sup>47</sup> PI §283

<sup>48</sup> PI II, iv, 152.

<sup>49</sup> PI II, iv, 152

<sup>50</sup> RPP I, §281; CV, 56

behavior, and an attitude that separates our reactions to living to human beings from those to objects or corpses.

It has been said that behaviorism resolves the problem of other minds “by fiat”<sup>51</sup>, that is, by defining mental phenomena so that mental state ascriptions are simply entailed by appropriate ascriptions of observable behavior. The general problem with this approach is that it ignores an asymmetry that is quite manifestly in place: When a subject talks about her own mental states, she is not informing others about how she behaves. Behaviorism discards the subjective use of mental state concepts in favor of a purely third-personal account suitable for scientific observation. Wittgenstein took pains to distance himself from such an approach. In his later philosophy, one can find an extended attempt to show that a human being does not include a hidden and necessary invisible realm of mental objects beyond her body, but to do this in a way that does not fall into the subjectivity-ignoring error of behaviorism.

## 2.2 A Wittgensteinian approach

Subjective experience, its immediacy for the subject, and its alleged hiddenness from others is a major theme in Wittgenstein’s later philosophy. This is referred to as the theme of “inner and outer” in his writings. As indicated in the Article I, I read Wittgenstein’s writings about the inner and outer as a wide-ranging critique of epistemic asymmetry in its different forms. If I am right in saying that it is the epistemic asymmetry, based on the Cartesian assumptions about the primacy of knowledge about one’s own mind and the inferential nature of knowledge about other minds, that gives rise to the different varieties of the problem of other minds, then Wittgenstein is addressing what deserves to be called the root issue of the problem.

What I suggest that we should find from Wittgenstein is not a solution to a single problem of other minds, in particular to the traditional skeptical problem: “How can we *know* that others have minds?” All of his contribution is not captured in a single philosophical argument or treatment, like, say, insisting that it is misguided to look for epistemic grounds where we have simply pre-reflective, primitive trust, as may be suggested based on the ideas of *On Certainty*. Instead, he delves deep into the central issue of asymmetry between oneself and other: the asymmetry (and sometimes symmetry!) of knowing one’s own thoughts and feelings and knowing those of others; and of expressing oneself and facing the expressions of others. More than solving or diagnosing the problem of other minds, what Wittgenstein does is re-shaping the problem and appreciating it as an actual relevant issue in the human condition, rather than doing away with it. The articles of this dissertation reflect this way of understanding the problem of other minds and Wittgenstein’s handling of it. A Wittgensteinian take on the problem of other minds, as I have attempted such a

---

<sup>51</sup> Hyslop, A. 2014. Other Minds. The Stanford Encyclopedia of Philosophy (Spring 2014 Edition), Reynolds, J. 2010. Problems of Other Minds: Solutions and Dissolutions in Analytic and Continental Philosophy. *Philosophy Compass* 5 (4), 326-335

project, is not a linear treatment of one neat problem, but a group of interconnected insights about knowing oneself and knowing others.

Wittgenstein's discussions around the so-called private language argument, behaviorism, and sensation-language are dense and very much debated. Here is one attempt to articulate a central insight of those discussions: If we want to understand what we mean when we ascribe feelings and experiences to ourselves and others, it is misguided to think that we are simply talking about "inner" processes or objects, hidden in the consciousness of each subject.

It shows a fundamental misunderstanding, if I am inclined to study the headache I have now in order to get clear about the philosophical problem of sensation.<sup>52</sup>

A word only gains a meaning as a part of a public linguistic practice. Naming a certain sensation "headache" would not mean anything as such, if there were not interpersonal standards on when and how to talk about headaches: ways of describing their quality, location and intensity, and ways to distinguish headaches from other kinds of sensations. Such standards can only be established in public communication, and they have to be implicitly learned by a child in order to learn the use of the word "headache". In the light of Wittgenstein's assimilation of the meaning of a word with its use, such standards are also constitutive to the meaning of "headache". It is a mistake to think that a subject's knowledge about her own mental states could be independent of community and learning.

Wittgenstein was acutely aware that a reader could take him to deny the significance of subjective experience, and he repeatedly fights off an imagined interlocutor who accuses him of behaviorism.<sup>53</sup> Denying the significance of subjective experience would be startling from someone who, like Wittgenstein, in the early 1930's used the term "proposition" to refer only to what was verified or falsified by direct experience,<sup>54</sup> and around the same time is reported to have said:

The world we live in is the world of sense-data; but the world we talk about is the world of physical objects.<sup>55</sup>

According to Jaakko Hintikka, it can even be claimed that Wittgenstein never abandoned this fundamental view, and always remained at bottom a philosopher of immediate experience.<sup>56</sup> Surely it would be a mistake to say that for Wittgenstein, subjective experiences are irrelevant for mental state ascriptions, and our talk about sensations and feelings are only an abstraction of a complex practice of talking about behavior. It seems more appropriate to say that his problem was how communicating subjective experiences is possible by means of language.

In a series of remarks in *Philosophical Investigations*, Wittgenstein poses a question: How do words refer to sensations? His starting point is not one where a subject

---

<sup>52</sup> PI §314

<sup>53</sup> E.g. PI §281, 296, 304-308

<sup>54</sup> See M, 55-60.

<sup>55</sup> LWL, 82

<sup>56</sup> Kusch, M. & Hintikka, J. 1988. *Kieli ja maailma*. Oulu: Pohjoinen. Prometheus ISSN, 123-139

could automatically have, in virtue of being acquainted with her own mind, a conception of what “pains”, “feelings” or “sensations” are, and could proceed to pose skeptical questions about the existence of the “same things” in others. Wittgenstein’s question is a more fundamental one: How can sensations become the subject matter of a shared language in the first place?

How do words *refer* to sensations?—There doesn’t seem to be any problem here; don’t we talk about sensations every day, and give them names? But how is the connexion between the name and the thing named set up? This question is the same as: how does a human being learn the meaning of the names of sensations?—of the word “pain” for example. Here is one possibility: words are connected with the primitive, the natural, expressions of the sensation and used in their place. A child has hurt himself and he cries; and then adults talk to him and teach him exclamations and, later, sentences. They teach the child new pain-behaviour. “So you are saying that the word ‘pain’ really means crying?”—On the contrary: the verbal expression of pain replaces crying and does not describe it.<sup>57</sup>

Wittgenstein suggests a beginning of an answer. The use of the word “pain” is importantly connected with expressing pain. The use is an extension of, or a development out of, natural and unreflective expressions of pain. If pain seems like an incommunicable private experience, maybe this is only because we assume that pain-talk always has to serve a descriptive function, and overlook its expressive function.

What we might be inclined to call “privacy of sensation” is rightly seen as a feature of the way such mental phenomena are spoken of. It is just part of the use of experience-words that there is a legitimate asymmetry in that use. The legitimate asymmetry, insofar as it can be in any way satisfactorily summarized in a phrase, comes to something like this: A person is not entitled to speak of another person’s experience in the capacity of its subject. Wittgenstein continues shortly after the previous remark, having an imaginary debate with an interlocutor:

In what sense are my sensations *private*?—Well, only I can know whether I am really in pain; another person can only surmise it.—In one way this is wrong, and in another nonsense. If we are using the word “to know” as it is normally used (and how else are we to use it?), then other people very often know when I am in pain.—Yes, but all the same not with the certainty with which I know it myself!—It can’t be said of me at all (except perhaps as a joke) that I *know* I am in pain. What is it supposed to mean—except perhaps that I *am* in pain?

Other people cannot be said to learn of my sensations *only* from my behaviour,—for I cannot be said to learn of them. I *have* them.

The truth is: it makes sense to say about other people that they doubt whether I am in pain; but not to say it about myself.

“Only you can know if you had that intention.” One might tell someone this when one was explaining the meaning of the word “intention” to him. For then it means: *that* is how we use it. [...]

The proposition “Sensations are private” is comparable to: “One plays patience by oneself.”<sup>58</sup>

---

<sup>57</sup> PI §244

<sup>58</sup> PI §§246-248

If a game is played with more than one player, it is no more what we call patience. Comparably, if someone fails to respect the authority of each subject to express her own states of mind, and claim he can speak of the subject's experiential world with the same authority as the subject herself, then he is no more using the vocabulary of personal experience in the usual way – the way which, in fact, can be said to be an essential part of our conceiving each other as an individual subject. An utterance like “I cannot have another person's pains” looks like a statement of fact, a description of an unfortunate limitation of human capacities. But by Wittgenstein's lights this is confusion. The utterance rather illustrates a part of the practice in which we take part when being with others, by delimiting what is admissible to say about others, in contrast to oneself.

The asymmetry between a subject's relation to her own mind and her relation to the minds of others is not an asymmetry between certain knowledge and uncertain beliefs. It is an asymmetry between experiencing, which is not in its own right an epistemic relation, and knowing.<sup>59</sup> When a person expresses pain, such an utterance is not subject to scrutiny by epistemic standards. It would be out of place to ask for a justification or supporting evidence for an expressive utterance. Doubting one's own pain appears rightly senseless, because doubting does not have an application in this context. When epistemic standards are not used, both the concepts of doubting and knowing are equally inapplicable. This leads Wittgenstein to say that it is strictly speaking wrong (or at least redundant) to say that I *know* I am in pain, as opposed to simply saying that I *have* a pain – which is an expressive statement.

So, what looks like a paradigm of certain knowledge – my acquaintance with my own pain – is not really straightforwardly a case of knowledge at all. This means also that other people's knowledge of my subjective life does not pale in comparison with that paradigm, because the paradigm was confused to begin with. Let us look again at a passage already quoted above:

If we are using the word “to know” as it is normally used (and how else are we to use it?), then other people very often know when I am in pain. – Yes, but all the same not with the certainty with which I know it myself! – It can't be said of me at all (except perhaps as a joke) that I *know* I am in pain. What is it supposed to mean – except perhaps that I *am* in pain?

Whatever the motivation might be for us to say that knowing the minds of others is problematic, such a worry does not normally permeate the whole of human life. As Wittgenstein writes, “I can be as *certain* of someone else's sensations as of any fact”<sup>60</sup>. Of course, being certain is not a guarantee that one is right, but given that most, if not all, of human knowledge is fallible anyway, the mere fact that we can be in error about the subjective experiences of others is hardly a reason to deem knowledge im-

---

<sup>59</sup> However, Wittgenstein's remark that I have “an attitude towards a soul” regarding other people, rather than an opinion that they have a soul (PI II, iv, 152), should also be kept in mind. We can know things about the thoughts, feelings and sentiments of others, but the primitive fact that we see each other as subjects of thoughts, feelings and sentiments in the first place is not grounded on an opinion. What makes our epistemic relation to others possible is not itself an epistemic relation, but an attitude. (I thank Sami Pihlström for emphasizing this point for me.)

<sup>60</sup> PI II, xi: 190

possible or inferior in that domain of life. The only real motivation to say that “we cannot *really know* the subjective experiences of others”, it seems, is that we contrast it with some other domain where we supposedly have infallible knowledge, or at least more firmly grounded knowledge. The obvious object of contrast is the alleged privileged and direct knowledge of our own mental states – but Wittgenstein has just given us reason to think that such privileged type of knowledge is a confusion. Another object of contrast that Wittgenstein considers is mathematical certainty<sup>61</sup>, but it would be odd to say that intersubjective knowledge (or any other type of knowledge) is disqualified because it does not allow for definite methods and proofs like those used in mathematics.

However, what Wittgenstein admits is that knowing the mental lives of other people is an extremely nuanced matter, involves subtle evidence that he characterizes as “imponderable”, and always leaves room, at least in principle, for disagreement that cannot be solved by rational argument. But this is just acknowledgement of the important fact that human beings are not transparent to one another; our knowledge of each other requires co-operation and goodwill from the part of both the subject and the other. These admissions do *not* mean that we are fundamentally right when we despairingly say “In the end, one can never really know what goes on in another person’s mind”.

There is an asymmetry between the first and the second-or-third person that is an internal part of human life. Only I am in the position to express my own state of mind. But there is not a metaphysically grounded epistemic asymmetry that would make our knowledge of other minds defective. The unfavorable comparison of third-person knowledge to first-person knowledge is illegitimate. The articles of this dissertation deal with three different themes in which Wittgenstein can help us to this insight.

---

<sup>61</sup> See Article I, 114

### 3 DISSERTATION STRUCTURE

Article I distinguishes three different ways in which Wittgenstein's later philosophy is a critique of epistemic asymmetry. It takes up three themes: **First-person authority and uncertainty about others**; **ineffability and describing one's experiences**; and **perceiving mental states by perceiving the body of the other**. It is the task of Article I to distinguish these themes and elucidate them in the context of Wittgenstein's philosophy. Articles II through IV, in turn, are developments of one or more of these themes. So, Article I, with its explicit focus on Wittgenstein's thought, forms the basis for the rest of the work, as pictured in Figure 1.

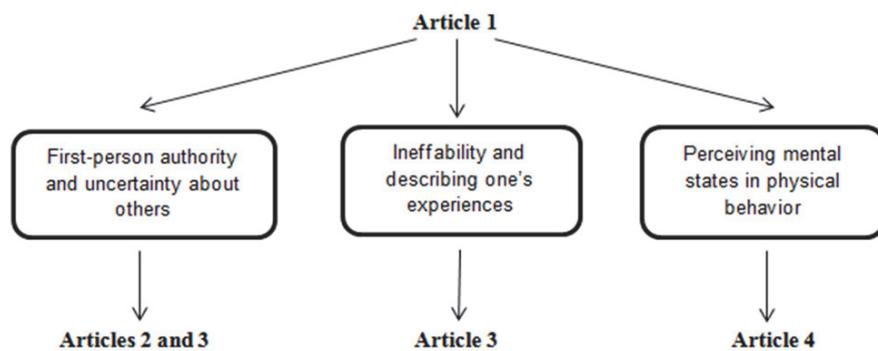


FIGURE 1 The structure of the dissertation as a flowchart

Even though Figure 1 displays the relationships between the individual articles satisfactorily, it is maybe good to present an alternative illustration too, because a flowchart-like diagram may give an undue impression of linear progress toward a specific solution. The three themes are not clear-cut questions waiting for a conclusive answer. Rather, they are areas in which Wittgenstein saw the threat of what I have called epistemic asymmetry; the assumption that our capacity to know or ac-

cess the “inner lives” of others is fundamentally defective compared to our unproblematic knowledge of our own mind. Wittgenstein’s way of addressing these threats was not so much to treat them as problems that have to be solved, but to offer alternative ways of thinking about the asymmetry between a subject and others that is an essential part of the human condition, so that the asymmetry might cease to appear intangible and epistemically threatening. This exemplifies the therapeutic nature of his philosophy. In the same spirit, it is best to understand the treatments of articles II through IV as developments rather than solutions of the problems posed by the three themes. This point is better served by the illustration of Figure 2.

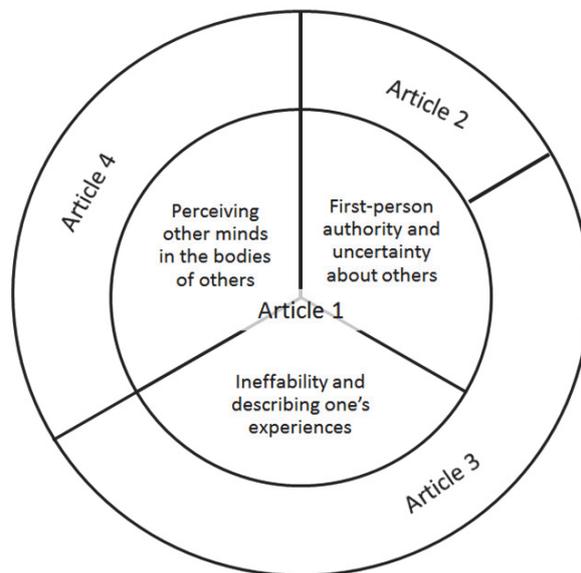


FIGURE 2 The structure of the dissertation as a pie chart

Obviously, this illustration has its flaws too: It should not be read as implying that Article III carries 50% of the substantial content of articles II through IV taken together. It merely implies that Article III spans two of the core themes, whereas Article II is concerned with its core theme to a lesser extent than the others, its main contribution being to a discussion whose connection to the objectives of this dissertation is more indirect than it is in the case of the articles III and IV.

While the rest of this thesis is a work of an individual researcher, Article II is a joint effort. Both authors contributed equally (50%-50%) to the process. More specifically, Mr. Heikinheimo produced most of the section 2, I produced most of the section 3, and the rest of the text was produced by collaborative writing. Considering the role of a PhD thesis as a work that prepares the student for a career in research, the pedagogical value of writing a collaborative article was remarkable. The process

of writing a philosophy paper with another person differs markedly from the process of writing alone. It became evident that in order to produce a coherent high-quality paper as a joint effort, it was needed not only to coordinate the authors' efforts in sketching, writing and editing stretches of text, but also to do some intensive work in explaining and elucidating one's views and ideas to the other. My view is that the present work was significantly more valuable as a learning process thanks to one of its articles being co-authored.

## 4 RESULTS

This section will summarize each of the four articles in turn, providing some commentary and clarification, as well as seeking to make clear their interconnectedness.

### 4.1 Article I: “Wittgenstein’s ‘Inner and Outer’: Overcoming Epistemic Asymmetry”

Article I pursues the insight that Wittgenstein’s contribution to the problem of other minds is best understood as his battling against what I have called epistemic asymmetry. His battles are taking place on more than one front. In Article I, I distinguish three such fronts, thus striving for a balanced account of his most important and inspiring thoughts in this area.

The interest of Article I is not primarily in its scholarly interpretation of Wittgenstein. It does not advance any radical or novel thesis about his thinking or its development. Its primary aim is rather in giving a sympathetic account of Wittgenstein’s central convictions and motivations in a certain area of his thinking, and doing it with as much clarity of expression as possible.

As indicated above in section 3, the three fronts – three forms of epistemic asymmetry that Wittgenstein addresses – I distinguish are:

- First person authority and the contrasting uncertainty about others
- Ineffability and describing subjective experiences
- The alleged indirectness of learning about the mental states of others through their bodily behavior

Each of these gets its own section in Article I. In its sub-section 3.1, devoted to first person authority, I first point out that Wittgenstein considers at least some honest first person ascriptions of mental states (in particular, self-ascriptions of pain) to be exempt from doubt. Furthermore, I point out that his claim is “grammatical” or “log-

ical" in character, meaning that doubt about my self-ascriptions of pain is unviable not because it is exceedingly improbable that I could err about my own sensations, but because doubt about them would be "senseless" or "idle". I link this immunity to doubt of first-person experience-talk to Wittgenstein's considerations in his so-called private language argument. The core of that argument is that there are no criteria for telling the difference between a correct recognition and a misrecognition of a subjective experience, if such an attempt of recognition is conceived in abstraction from public, interpersonal discourse. Wittgenstein contends, however, that a subject can rightfully apply a concept like "pain" to her subjective experience without relying on any kind of criteria. When criteria are absent, standards for correctness are absent also. The import of section 3.1 is to show that Wittgenstein granted a certain kind of a special secure status to self-ascriptions of experience, but this admission is not grounded in a notion of special epistemic access. Rather, it is grounded in the constitutive features of first-person experience-talk as a use of language. The importance of this lies in the way it makes it less tempting to construe the subject's relation to her experiences as a paradigmatically good epistemic relation, to which the subject's relation to the experiences of others should be unfavorably compared.

The sub-section 3.2 addresses the other side of the coin, namely, the alleged uncertainty of all ascriptions of mental states to other subjects. The position spelled out and attributed to Wittgenstein in 3.1 was that honest first-person avowals of at least some experiences are in a special "logical" way exempt from doubt. The topic of 3.2 is that it seems impossible to eliminate doubt completely when ascribing mental states to others from a third-person perspective. The possibility of pretense, that is, the possibility of an avowal not being honest, seems to be at least in principle present in any case of an avowal of another person. In general, conclusive verification of any assumption about the subjective experiences of another person may be thought to be impossible. Sub-section 3.2 points out Wittgenstein's twofold reaction to these assumptions. On the one hand, he denied that other-ascriptions of subjective experiences are always tainted with some amount of incurable uncertainty; he emphasized that there is an abundance of everyday situations where we ascribe experiences to others with unwavering, non-reflective certainty. Admittedly, this can be taken as merely a quite uninteresting observation of common sense psychology. But more interestingly, Wittgenstein accepted the idea that some kind of ambiguity, or epistemic "open-endedness", can be taken to be a constitutive element of the practice of ascribing experiences to others. The most important insight, according to my reading, is that the threat of epistemic asymmetry results from a misguided comparison between the immunity to doubt enjoyed by some avowals in the first person, and the ambiguity or open-endedness of third-person mental state ascriptions. Wittgenstein's insight is to treat these as two constitutively different kinds of language use, rather than a parallel pair of cases, one case enjoying a privileged access to a domain of inner experiences that is lacking in the other case.

Section 4 addresses the second form of epistemic asymmetry, summed under the title "Ineffability". This is the intuition that the phenomenal qualities of subjective experiences cannot be directly or ideally revealed to others, and subsequently, that each subject knows more about her experiences than she can ever communicate

to others by reports. I point out a number of elements in Wittgenstein's thought that serve to undermine this intuition. First, I note Wittgenstein's opposition to the assumption that all first-person talk about subjective experiences is descriptive in character. I point out that Wittgenstein, not succumbing to the "descriptive fallacy"<sup>62</sup>, recognized and emphasized in his later thinking that not everything that superficially looks like a description of a state of affairs actually functions as a part of fact-stating discourse. In particular, he saw the first-person subjective experience talk as importantly connected to primitive, natural expressions (in particular in the case of his favorite example of pain). If there is a philosophical problem in the inadequacy of reports of subjective experiences, this can become seen as less problematic through the observation that much of first-person subjective experience talk is anyway more properly assessed by its expressive function than by considerations of truth and accuracy. Article III will later pick up this point, and elaborate further the possible ways in which expressive and descriptive talk are related in avowals.

Another point is that subjective experiences only seem ineffable if the actual *undergoing* of a given experience is construed as the ideal and perfect way of being epistemically in touch with that type of experience, and experience-reports are construed as vain attempts to reproduce the experience by means of language. Wittgenstein, I argue, inspires one to view descriptions of experiences as linguistic instruments that are used for a variety of purposes and in a variety of conversational contexts, and in no context does their point need to be assimilated to that of attempting to give a *sample* of the subjective experience to another. Finally, there are, of course, real and disquieting situations in human life when we feel that another person's experiential world remains "hidden" and intangible to us, but these need not be seen as indications of a deep metaphysical ineffability of subjective experiences. They are better seen as breakdowns in interpersonal understanding, resulting from a failure to share a background of sufficiently similar stock of past experiences, or some other kind of failure to share a common context of communication.

Section 5 addresses the alleged asymmetry between a subject's "direct" knowledge of her own mental states and her "indirect" knowledge of the mental states of others based on their bodily behavior. I identify as Wittgenstein's target the preconception that since we, in general, do not resort to observing our own behavior in order to learn about the goings-on of our mental life, there is a "direct" and privileged way of learning about them that is available in our own case but unavailable in the case of others. This leaves us with only observing behavior in the case of others, as a kind of a second-best option. The sub-section 5.1 spells out Wittgenstein's criticism of construing introspection as a source of evidence, comparable to a kind of inner observation. The therapeutic consequence of this position, I suggest, is that coming to know of the minds of others based on behavioral evidence no more needs to be considered inferior compared to any other possible type of evidence. Sub-section 5.2 points out a connection between the theme of seeing patterns of bodily behavior as manifestations of mental states and Wittgenstein's discussions of aspect-

---

<sup>62</sup> See Austin, J. L. 1975. *How to do things with words : the William James lectures delivered at Harvard University in 1955*. Edited by Urmson and Sbisà, second edition, Oxford: Clarendon Press

seeing. The important analogy is that just as in seeing an ambiguous picture, one can see the aspect “in” the picture but not as a thing separate from the picture, it is possible to see a mental state (in particular, a feeling or emotion) literally embodied in its physical expression. Article I ends with concluding remarks in section 6.

## 4.2 Article II: “The Redundancy of the Zombie Argument in *The Conscious Mind*”

Article II, co-authored with Antti Heikinheimo, presents an argument (or, so to speak, a meta-argument) in the context of the debate of physicalism against anti-physicalism in contemporary philosophy of mind. Its critical focus is on the so-called “zombie argument” and its use in arguing for an anti-physicalist position, in particular by David Chalmers<sup>63</sup>. The zombie argument is maybe the place where the standard skeptical problem of other minds features most prominently in contemporary discussions. The crux of the zombie argument as presented by Chalmers is that the conjunction of all true physical propositions about the world does not logically necessitate any proposition about phenomenal consciousness, which is taken (again, by Chalmers) to be a denial of physicalism. The zombie argument, in the form in which it is targeted in Article II, argues for this conclusion by assuming a link between conceivability and logical possibility. We lay out the argument as follows. Let P be the conjunction of all true propositions about fundamental physical facts, and let Q be any contingently true proposition about conscious experience:

- |           |  |
|-----------|--|
|           | (1) It is conceivable that P and not-Q   |
|           | (2) For any proposition R, if it is conceivable that R, then it is logically possible that R |
| Therefore | (3) It is logically possible that P and not-Q  |
| Therefore | (4) Physicalism is false. <sup>64</sup>  |

The first premise says that it is possible to conceive of a situation where all the fundamental physical facts of our world are as they are now, but where the facts about conscious experience are different to an arbitrary detail. The second premise says that conceivability entails logical possibility. The conclusion (3) says that a situation where all the fundamental physical facts of our world are as they are now, but where the facts about conscious experience are different to an arbitrary detail, is logically possible. This, in Chalmers’s argumentation, entails the further conclusion (4) that physicalism is false, given his reading of physicalism as the thesis that facts about conscious experience are logically supervenient on the totality of fundamental physical facts.

---

<sup>63</sup> Chalmers, D. J. 1997. *The conscious mind: In search of a fundamental theory*. Oxford University Press, USA

<sup>64</sup> Article II, 7

In section 2 of Article II, we spell out why the argument needs additional support in order to be convincing. The premise (2), linking conceivability and logical possibility, can be called into question, because it seems plausible that we can in some sense conceive of at least some things that are not logically possible. Our example is Goldbach's conjecture, an unsolved problem in number theory. It is possible to try to prove either the truth or falsity of Goldbach's conjecture, which makes it plausible to say that both options are conceivable for a mathematician. Now, if mathematical truths are logically necessary, then regardless of whether Goldbach's conjecture is actually true or false, we are able to conceive of something that is not logically possible. In other words, conceivability is not a reliable indicator of logical possibility. Chalmers takes this into account, and holds that only a strong form of conceivability, termed "ideal and positive conceivability", implies logical possibility. We argue, however, that this move does not help the argument's persuasive power, because it makes its premise (1) far less evident. Facts about ideal and positive conceivability are not easily accessible for us, and because of this, those discussing the argument need to resort to idiosyncratic intuitions about what is conceivable in the required sense and what is not.

Chalmers presents in his *The Conscious Mind* a group of five arguments against physicalism. The zombie argument is the first of these, and the one that has received the most attention. We go on to investigate whether some of the other arguments is suitable to lend the additional support that the zombie argument needs, concentrating on two options.

In section 3, we consider the option called "argument from epistemic asymmetry". Here, Chalmers argues that conscious experience poses a genuine epistemic problem of other minds. What Chalmers means by epistemic asymmetry is that no amount of familiarity with the physical facts of the universe can by itself endow us with the knowledge that other living things have conscious experiences. Our grounds for belief in consciousness, Chalmers writes, "derive solely from our own experience of it"<sup>65</sup>. We acknowledge that this epistemic thesis can be used to support the premise that zombies are ideally and positively conceivable. However, we point out that this leads Chalmers to account for our knowledge of consciousness in others in terms of an inference to the best explanation/argument from analogy, and this position has uncomfortable consequences. First, if epistemic asymmetry is accepted, then best explanation/analogy arguments seem inadequate to fully answer the problem of other minds generated by the asymmetry. Second, the position is vulnerable to the standard objection that the argument from analogy is an instance of an inductive generalization based on only one case. Third, the position overlooks the conceptual problem of other minds, famously spelled out by Wittgenstein.

In section 4, we turn to "the argument from absence of analysis", which we view as Chalmers's fundamental argument. The argument states that the concept of conscious experience cannot be analyzed in terms of physical properties. There is a reason why the concept of consciousness cannot be defined, even in an approximate way, by reference to physical processes that play causal-functional roles in an organ-

---

<sup>65</sup> Chalmers, D. J. 1997. *The conscious mind: In search of a fundamental theory*. Oxford University Press, USA

ism. If such a conceptual analysis is not available, then facts about conscious experience are not logically supervenient on fundamental physical facts, refuting physicalism in the sense targeted by Chalmers. We identify what seems to us to be Chalmers's most plausible way of motivating the claim that such a conceptual analysis is not available, namely, the argument that a sufficient explanation of phenomena of conscious experience cannot be given in terms of physical properties, even in principle. Chalmers argues that physical properties are ultimately defined in terms of spatio-temporal structure and causal dynamics, and they are able to explain only phenomena that are themselves about such structure and dynamics. The phenomena of conscious experience, according to Chalmers, are not (only) about structure and dynamics, and so explanations of them in terms of physical properties are necessarily insufficient. We demonstrate how this argument, termed "the argument from explanatory insufficiency", combines with the argument from absence of analysis, to generate a consistent argument against physicalism. We point out that the latter argument can be used to support the premise (1) of the zombie argument, but because the latter argument reaches the conclusion of the zombie argument by itself, without having to invoke any controversial claims about conceivability and possibility, it makes the zombie argument redundant. We sum up these conclusions and make some disclaimers about them in the latter part of section 4 and in the short concluding section 5. It is worth noting that even though we argue against the zombie argument's usefulness in motivating an anti-physicalist position, we are not ourselves arguing either for anti-physicalism or physicalism; we remain neutral on the matter. We do not pass a judgment on whether or not the argument from absence of analysis is fully compelling; our point concerns merely the order of primacy of the relevant arguments.

Section 3 is the place where Article II most directly engages with argumentation about the details of the other minds problem, but the zombie thought experiment in general is a notable appearance of the problem in the contemporary philosophy of mind. The zombie thought experiment addresses one of the core themes of this dissertation: the possibly ineliminable amount of uncertainty about the minds of others. The attention received by the zombie argument testifies that a significant number of authors deem the zombie thought experiment intuitively tempting, at least enough so that the issue of whether or not it depicts a possible state of affairs ("logically" or otherwise possible) deserves serious comment. Even if practically nobody harbors a real, practical suspicion that some inhabitants of the natural world might be zombies, even its status as a thought-possibility might tell something interesting about how contemporary people (and philosophers) think about intersubjectivity.

### 4.3 Article III: “Expressivism, Self-Knowledge, and Describing One’s Experiences”

Article III spans two of the themes discussed in Article I: first-person authority, and describing subjective experiences. I pointed out in Article I the Wittgenstein-inspired idea that communicating one’s subjective experiences to another person only seems philosophically problematic if one conceives communication as an attempt to transfer one’s token of experience into the mind of another. The remedying conception is to view describing subjective experiences as “a special application of language, for special purposes”<sup>66</sup>. One objective of Article III is to make a small contribution to elucidating the way this “special application of language” may work. A central theme of Article III is the interconnectedness of expressive and descriptive ways of talking in the context of avowals<sup>67</sup>. The partly (but *only* partly) expressive character of avowals, mentioned in Article I, plays a key part in how I conceive the nature and role of first-person authority and incorrigibility of avowals in Article III. In this way, Article III addresses also the theme of first-person authority.

The discussion in Article III is conducted in terms of self-knowledge, thus bringing in one more interrelated concept. I take as my starting point two accounts of the authoritative knowledge we have of our own mental states, as those accounts are distinguished by David Finkelstein<sup>68</sup>. According to the first account, *detectivism*, our knowledge of our own mental states is grounded in some process akin to inward perception, that is, introspective detection. This view has obvious affinities with Wittgenstein’s target in his critique of private objects, as Finkelstein makes clear. The second account, *constitutivism*, sees the verdicts we make about our own mental states as more akin to decisions or declarations than to perceptual reports. In sections dedicated to both of these positions respectively, I offer reasons why neither of the positions is satisfactory as an account of first-person authority or self-knowledge.

Following the lead of Finkelstein and others, I examine the point suggested by Wittgenstein that there is a connection between verbal avowals and natural, pre-linguistic expressions. I note that the related account of *expressivism*, in its simple form, explains first-person authority in a deflationary way, by stating that avowals are actually not truth-evaluable statements, but only look superficially like such. Instead, they are cultured expressive behavior, engaged in tasks like requesting, eliciting emotions, threatening or pleading. I locate in Gilbert Ryle a view that saw the “primary application” of avowals like this. Because according to simple expressivism, avowals are not truth-evaluable, they cannot be meaningfully corrected by another person for the trivial reason that they do not have the kind of content on which meaningful disagreement is possible. For the same reason, however, simple expressivism cannot view avowals as instances of self-knowledge, any more than grunts and gestures can be viewed as instances of self-knowledge. Also, the central idea that avowals simply are not descriptive statements is implausible, because it seems

---

<sup>66</sup> PI §609

<sup>67</sup> As I point out in Article III, footnote 1, my use of the term “avowal” is broad.

<sup>68</sup> Finkelstein, D. H. 2003. *Expression and the Inner*. Cambridge: Harvard University Press

unavoidable to admit that in many contexts they function like fact-stating talk, and can, for example, be contradictory. More sophisticated developments of the expressivist point embrace this.

I identify in Wittgenstein, in particular in the part II of *Philosophical Investigations*, not a simple expressivist view, but a view that acknowledges that avowals can play the role of both expressions and descriptions, with a range of intermediate cases. In the remainder of the paper, I apply this construal of avowals as moving on an expressive-descriptive scale, and attempt to give outlines of an account of self-knowledge and first-person authority that gives a central place for expression, but also incorporates the valid insights of both detectivism and constitutivism.

A case of avowal that functions simply as a verbal replacement of a primitive expression, as in Wittgenstein's example in his PI §244, I call a "primitive avowal". Avowals done from a completely detached perspective (as if from a third-person position) I call an "intellectual self-ascription". The cases on a scale between these two extremes are my main object of interest. These latter I call "deliberations", adopting the word from its slightly different application by Richard Moran<sup>69</sup>.

I illustrate a deliberative avowal by an example where a subject is required, in a context of a medical check, to describe his pain experience by using a scale of 1 to 10. The example is meant to be representative of a common and salient kind of describing an experience. It is also meant to be maximally plausible as a situation where the mental state in question is suitable to be expressed (a pain experience), but where the purposes and surroundings of the avowal are specifically those of describing. Furthermore, having the avowal carried out by giving a number on a scale makes it unambiguous that the possible avowals stand in logical relations to each other; that is, they can unambiguously contradict each other. I propose that such an avowal is plausibly seen as a speech act that is both an act of describing and that of expressing: a request for others to accept the pain-avowal as a valid description of the subject's pain. I go on to make two main points based on my considerations. First, the authority of avowals is crucially based on their (partly) expressive nature. Second, the expressive and descriptive aspects of a deliberative avowal are interdependent in such a way that the expression is made by putting forward a description, but on the other hand, the successfulness of the description as a speech act is dependent on its expressive quality. This interdependence is why I purposefully do not say that a deliberative avowal has two "components" or parts. Figure 3 below is an illustration (completely heuristic) of a possible way in which the different modes of avowals, their expressive and descriptive aspects, and first-person authority may be related to each other:

---

<sup>69</sup> Moran, R. 2001. *Authority and estrangement: An essay on self-knowledge*. Princeton (N.J.): Princeton University Press

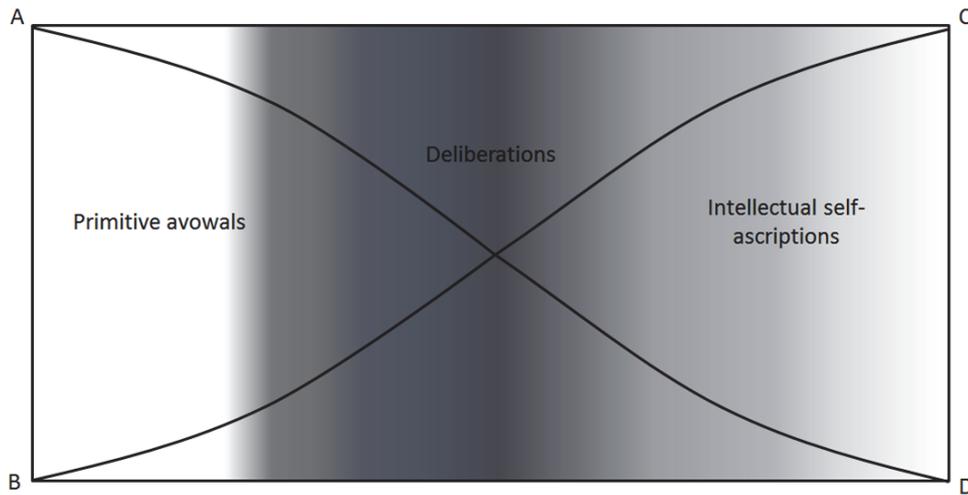


FIGURE 3 Avowals and first-person authority.

Color gradient = the strength of first-person authority

AD = The expressive aspect of avowal

BC = The descriptive aspect of avowal

Here, first-person authority is taken to not apply to primitive avowals, insofar as they are put forward, and heard, in the manner of Ryle's "primary application" of avowals, that is, as requests, demands, and other non-descriptive utterances. Moving toward intellectual self-ascriptions on the scale, first-person authority is gradually weakening, as avowals become increasingly detached and "detectivistic". In the middle, I propose that first-person authority can be seen to be the strongest in deliberative cases where the avowals are already intended as self-descriptions, but put forward in an expressive manner. In such cases, they can be somewhat likened to declarations or decisions, and their authority can be explained via that role, as suggested by the constitutivist position.

I conclude the paper by characterizing a sense in which self-knowledge is a process whereby a subject becomes an adept and reflective user of the expressive and descriptive aspects of avowals. This adeptness involves acknowledging that a subject has both a "subjective", first-person viewpoint on himself, and an "objective", third-person viewpoint on himself. As a final point, I note that first-person authority, as a part of human social or linguistic practice, is a way in which our subjective, expressive viewpoints on ourselves are allowed to play a role in our communal life, and I stress the ethical importance of this part of our practice - or, even though I do not talk explicitly in these terms in Article III, this part of the "language-game" of avowals.

#### 4.4 Article IV: “Wittgensteinian Criteria and Perceiving Other Minds”

Article IV deals with the notion of criteria found in Wittgenstein, seeking to clarify its contribution to the problem of other minds, in particular connecting it with the theme of perceiving the mental states of another person in the expressions of the other person. In this way, Article IV picks up the theme singled out in the section 5 of Article I, elaborating the idea that it is possible to perceive the mental life of others “directly” by perceiving their bodily behavior, without an intermediate process of inference or assumption.

I start Article IV with a brief background section, introducing how the notion of criteria and the theme of “inner and outer” feature in Wittgenstein’s later thought. Especially the idea that some kinds of bodily behavior serve as criteria of mental states in others, and not merely as signs of them or inductive evidence for them, has been thought to mark a novel Wittgensteinian solution or “dissolution” of the problem of other minds. One of my leading thoughts in Article IV is that the way in which Wittgenstein contributes to the problem of other minds is by criticizing the distinction between the “inner and outer”, and that his remarks on criteria are interesting against that background.

Wittgenstein made remarks of the form “Y is a criterion for X” in a variety of contexts, not making it very clear what kind of entities they are that stand in criterial relations to each other. I take criteria to be, quite broadly construed, ways of telling whether something counts as an X. This is consistent with Glock’s judgment that “[t]he basic point [of Wittgenstein’s talk about criteria] is that certain phenomena or facts license the application of certain words”<sup>70</sup>.

In section 2, after spelling out the above broad construal of criteria, I move on to distinguish the relevant features of the notion of criteria. My analysis, although neither uniquely exhaustive nor the only one possible, is still meant to be adequate in spelling out Wittgenstein’s usage and the subsequent interpretations of commentators, in enough detail to serve the purposes of Article IV. I distinguish four relevant features:

- a) Criteria stand in an intrinsic (“grammatical”) relation to what they are criteria for,
- b) Criteria are context-dependent,
- c) Criteria can be multiple, and
- d) Criteria can be defeasible.

I point out that, when we consider the case of types of bodily behavior as criteria for mental states, a) and d) indicate a problematic tension. First, a) suggests that meanings of mental state terms can be explained or defined at least partly by reference to behavior, and it becomes crucial to distinguish this stance from behaviorism. Distance to behaviorism is provided by d). Making behavioral criteria defeasible means acknowledging that even when all criteria for another person being in a certain men-

<sup>70</sup> Glock, H. 1996. *A Wittgenstein dictionary*. Oxford: Blackwell

tal state are satisfied, it is still at least in principle possible that the person is not in fact in that mental state; pain-behavior alone does not make it the case that a person is in pain, for example. But if pain and pain-behavior are ultimately distinguishable in this way, what is the import of the claim that there is, however, an intrinsic relation between them? I then draw attention to the element in Wittgenstein's conception of intersubjectivity addressed in Article I's section 5, namely, the possibility of immediately perceiving a mental state, in particular a feeling or an emotion, by perceiving the body of the other. This is an important aspect of Wittgenstein's critique of the inner and outer. In the rest of Article IV, I attempt to clarify how there can be said to be behavioral criteria for the mental states of others consistent with the features a)-d) (although I will have important qualifications with d)), and how the notion of criteria fits together with the insight that the minds of others can be immediately perceived in bodily terms.

In section 3 I adopt a reading, following McDowell in his essay "Criteria, Defeasibility, and Knowledge"<sup>71</sup>, according to which criteria should be thought as constitutive of the thing they are criteria of. This means that its being the case that X consists in (some of) the criteria for X being satisfied, and when X obtains, perceiving the criteria of X amounts to perceiving X. So it is acknowledged that there is an intrinsic connection between X and the criteria of X. What it is to be an X is explained or defined by explicating the criteria of X.

McDowell suggests that the possibility of deceptive cases, that is, cases where we falsely judge a case as an X based on perceiving the usual criteria of X, should not prevent us from saying that in *good* cases, perceiving the criteria of X amounts to perceiving X. The ever present possibility of deceptive cases indicates that criteria are always in a sense "defeasible", but they are defeasible in a specific sense that stems from their context-dependence and multiplicity. In varying circumstances, criteria can gain or lose their status as criteria. Encountering a problematic or a downright deceptive case involves acknowledging that criteria that are normally used for telling whether something is an X are not obviously useful in this case; that is, they are not (or not obviously) criteria in *these* circumstances. But in any given situation, *if* we judge that something is a criterion of X, *then* we cannot consistently say in that same situation that the criteria for X are satisfied but it is still not the case that X.

This reading emphasizes that the way we classify phenomena by criteria and signs is dynamic, and although what it is to be an X is to be spelled out via some (typically diverse) set of criteria, it need not be conclusively defined by reference to any fixed set of criteria. This, furthermore, does not prevent us from saying that in specific unproblematic cases, its being the case that X consists in some of X's criteria being satisfied, and perceiving the criteria in those circumstances amounts to perceiving X.

In the context of other minds, then, it can be said that there is a diverse and open-ended set of bodily criteria for, say, anger, and these criteria belong to the "grammar" of the concept of anger: they belong to an explanation of what anger is.

---

<sup>71</sup> McDowell, J. 1998. *Meaning, knowledge, and reality*. Cambridge (Mass.): Harvard University Press

Any of these criteria can fail in specific circumstances; it is always possible to encounter a deceptive case where the usual criteria are not applicable. But in good cases, the bodily or behavioral criteria of anger – frowning, huffing, cursing, or the like – are constitutive of the other person’s anger in that instance, and perceiving them does not need to, following McDowell’s phrase, “fall short of the fact”<sup>72</sup> of the other person’s anger.

The obvious worry, however, is that this way of thinking does not depart far enough from behaviorism. Assimilating instances of bodily behavior with instances of mental states, even with qualifications, seems to violate an important intuition that mental states, even if in some sense perceptible, are not perceptible in quite the same way as ordinary objects are perceptible. The convenient way of putting the matter is that we perceive *expressions* of the mental states of others on their faces and bodies, rather than mental states as such. The notion of expression seems to involve a distinction between the expression and the expressed. In section 4, I suggest that the distinction is not absolute, and that there is a sense of “expression” that allows us to conceive some expressive bodily events as literal embodiments of mental states, without thereby being guilty of an illegitimate “objectifying” of the mental.

---

<sup>72</sup> McDowell, J. 1998. *Meaning, knowledge, and reality*. Cambridge (Mass.): Harvard University Press

## 5 DISCUSSION

I will finish this Introduction with a section that attempts to elaborate the overall picture this work as a whole intends to present. I will explain how Wittgenstein's philosophy guides one to rethink the problem of other minds, by questioning the assumption that there is a way knowing the minds of other people that is forever denied from each subject: namely, the assumed privileged way in which each of us knows our own mind. The most general moral I draw from Wittgenstein is this: There is genuine *asymmetry* between my first-person viewpoint of myself and my viewpoint on the subjective lives of others (that is why any unsophisticated behaviorism appears so self-evidently wrongheaded), but there is no *epistemic asymmetry* between myself and others regarding my mental states, at least not such that should give rise to the notions of privileged access and private inner realm of consciousness.

What does this moral amount to? Let us take the negative claim first. There is no epistemic asymmetry. Does that mean that another person can be in an epistemically equal position regarding my subjective experiences as I? There is something counterintuitive in this, although a flat denial of the claim seems wrong too. On the one hand, very mundane experiences of living with others seem to testify that the claim is at least in some sense true. There are cases where other people can really offer me an insight about what I am thinking and feeling that is better than my own judgment. Sometimes I genuinely need others to make sense of what is really going on in my thoughts and feelings. If this were not true, many forms of therapy, counseling, or just plain consulting a friend about my emotional worries, would not work. But on the other hand, we rightly feel that there is something preposterous about the claim that another person can know what I think equally well, or better, than I know it myself. The claim seems to violate something fundamental about our very subjectivity and individuality.

What it seems to violate is first-person authority, which is the motivating topic of the discussion of self-knowledge and expressivism in Article III. First-person authority is the assumption that each subject's sincere statements about her own mental states have a unique claim to be true (an idea obviously connected with that of privileged access). This is not only a disinterested theoretical assumption. It also conveys a certain form of respect, marking a line that should not be crossed lightly, if

we are to treat each other as autonomous, full-blown individuals. But according to the view I read from Wittgenstein, denying epistemic asymmetry does not negate first-person authority, understood in the right way. This is because the crucial difference between my avowals and the judgments of others about my mental states is not based on privileged *knowledge* that I have of my mental states, and that others lack. The difference is rather that the *relation* of my avowals to their subject matter (my mental states) is categorically different than that of the judgments of others. My avowals have the capacity to be *expressions* of my mental states. This is also what the positive claim above, the positive part of the general moral, comes to. There is a genuine (non-epistemic) asymmetry between the first-person and the second-or-third person viewpoint, because only I can express my subjective states. Others can make judgments about them, inquire about them, doubt them, or sympathize with them, but they cannot express them. This “cannot”, Wittgenstein would obviously insist, does not mark any contingent natural limitation, but is what he would call a “grammatical” remark.

If I have a pain, it is I who can express my pain. The false picture of the “inner” that Wittgenstein is battling against tries to explain this by insisting that it is because there is a private, phenomenal pain-object in me that only I have access to. But even though this false picture should be denied, it is obvious that the pain I feel has *something* to do with my expression of pain. Wittgenstein is clearly not denying this. He is not downplaying subjective experience. He is only warning against construing it as a parade of ready-made, private inner objects. Thus the remark in *Philosophical Investigations*:

“But you will surely admit that there is a difference between pain-behaviour accompanied by pain and pain-behaviour without any pain?” – Admit it? What greater difference could there be? – “And yet you again and again reach the conclusion that the sensation itself is a *nothing*.” – Not at all. It is not a *something*, but not a *nothing* either!<sup>73</sup>

The subjective experience, the feeling, does not exist as a determinate something prior to the beginning of the language-game of expressing one’s feelings. Still it is obviously there – it is not a “nothing”. But a pain is only identified as a pain, enters the language-game as a pain, so to speak, when I express it as pain. My subjective experience is not a totality of transparent, given objects that my avowals seek to accurately convey. It is rather the possibly quite amorphous and chaotic *subject matter* of my expressive acts.

So, in a certain sense I *am* in a special, unique position in any inquiry about my mental states. But this special position is not due to an epistemic vantage point. It is due to my being the one who expresses my mental states, and due to each subject’s expressive avowals being the ground level of discussions about that subject’s mental states. They are the ground level in the sense that questions of truth or error are not in normal circumstances raised about them. They are the foundation on which inquiries about a subject’s experiential world rest. Such inquiries can be constructively critical, and they can lead to changes in the way the subject sees her own experiences; they can make it so that the subject wants to start expressing herself differently. But

---

<sup>73</sup> PI §304

no matter how deep and sophisticated reflection about someone's mental states we find ourselves in, the subject's expressive avowals retain their special role in the practice. They are expressive rather than fact-stating, and so not subject to fact-checking either. This is a status that expressive avowals have in our life as subjects among other subjects. It is a status that is implicitly granted for them, not metaphysically possessed by them by default.

This Wittgensteinian insight is well voiced by Naomi Scheman. While analyzing a woman's experience of repressed anger, she writes:

[T]he theory of privileged access (the philosophical view that we are each the ultimate authority about our own emotions) can be seen less as a fact of epistemology than as a piece of social theory – a clue to what we care about in our interpretations of people. That we are inclined not to notice this, in part because of the emotions-as-inner-states picture of the mind, is typical of the workings of an ideology: matters of political choice come to seem to be matters of unchangeable fact. We think that emotions just *are* particular states of individuals, specifiable independently of social context. [...] We care most about our own view of ourselves since we are the ones who are allowed to determine how we are to be taken as feeling: privileged access functions as a sort of property right.<sup>74</sup>

First-person authority is seen in the right light if it is not seen as stemming from a mysterious privileged access, but as an ethical and social demand, a part of recognizing the other as a self-standing subject who has a unique viewpoint on her own experiential world – a subject's perspective.

What is that unique "subject's perspective", then? It would not be satisfactory to say that it comes down to the simple fact that only I am acquainted with my thoughts, sensations and other feelings – that only I can feel my pains, for example. For Wittgenstein, "Only I can feel my pains" is an example of a sentence that masquerades as a general, indubitable (metaphysical) fact, while what it really does is reflect something fundamental how we conceive of persons and pains, and how we talk about them. In Wittgenstein's terms, it illustrates a piece of grammar.

In actual practice, we very often talk about several people having the "same" pains, where "same" means "qualitatively identical". Two people suffering from the same spinal malformation, for example, can discover that they have the "same" back pains. But the picture of the hidden inner realm which Wittgenstein opposes insists that there is a further sense in which the two people could feel the "same" pain, but in fact they never do. Namely, they could have a *numerically identical* pain experience, the same pain-token; but in fact they never do, because each one only has access to one's own token. The picture of the hidden inner realm suggests that this is what the unique "subject's perspective" comes down to: the privileged access of each subject to his own array of pain-tokens. But this distorts the issue, and fails to make sense of the basis of first-person authority. Because if I am especially knowledgeable about my own feelings just by virtue of being lucky enough to be the only one who actually *feels* them, then why are we so sure that our feeling-reports are *accurate*? If my knowledge of my mental states is a matter of my being able to feel an array of private mental-state tokens, then what makes us so sure that we always feel them *right*,

---

<sup>74</sup> Scheman, N. 1980. Anger and the Politics of Naming. In *Engenderings. Constructions of Knowledge, Authority, and Privilege*, New York: Routledge.

that is, feel them as they really are? Our capacity of feeling the inner objects, it seems, should be as prone to error as any other cognitive capacity. Moreover, because nobody else can guide or correct us in the use of this capacity (at least not directly), this error-proneness should be somewhat worrying. If we want to keep hold of the conviction (as we usually do, I think) that in some sense a person's avowals of her own pains and other subjective experiences are especially resistant to error, then we will have to assume that our cognitive capacity of feeling the objects in our inner realm just happens to be amazingly reliable, as if by a metaphysical serendipity.

Wittgenstein improves matters by giving a better account of the "subject's perspective" and the way in which that perspective is guarded against error. The position of the subject is not special in virtue of fortunately having a perfect view on a gallery of inner objects. It is special because the subject talks and acts expressively; she has a kind of license for voicing her experiences that is not granted to others. When the subject talks expressively, she is not trying to produce a description of inner experience-objects she sees in her mind's eye. She is doing something constructive: putting her feelings out in the open, so that they could become concrete and identifiable for both herself and others.

It is the subject's expressions that have a special role, not her epistemic standing. Her expressions are not the end products of a game of first identifying an inner experience and then trying to translate it to words. They are the starting points of the game of naming and identifying one's subjective experiences.<sup>75</sup> One does not start with a ready-made cavalcade of inner experiences, facing only the problem of how to accurately communicate them outwards. One starts with a subjective reality, possibly quite messy and confused, that calls for expression; and understanding it even in first person requires working it out and conceptualizing it for others.

So getting clear about one's own inner experiences is a social affair. First-person authority is a kind of a right to determine and identify one's feelings. As such, it may (and does) become an object of social negotiation, even power games. Understanding this has potential to lead to a better self-understanding, a clearer picture of what we do when we talk to each other about how we feel. We are playing a public game, with lots of potential to learn more about what our experiences are like than we could ever learn merely from an isolated first-person stance. But if we find ourselves on the worse side of a power relation, there is also a danger that we may succumb to disregarding or belittling our own feelings and experiences, and sincerely thinking that we are right to do so.

The idea that "inner" feelings are phenomena of the public, shared world is enlightening. But the practice of talking about subjective feelings seems to need, in order to not become detached from the reality of lived experiences of some, maybe oppressed subjects, some kind of an anchoring point. It needs an objective basis that makes it illegitimate to claim that feelings and emotions are *solely* socially determined things, endlessly open to negotiation and re-negotiation. It is natural expression that provides the anchoring point. In a passage from *Notes for Lectures on "Pri-*

---

<sup>75</sup> See Wittgenstein's remark in PI §290: "What I do is not, of course, to identify my sensation by criteria: but to use the same expression again. But this is not the *end* of the language-game: it is the beginning."

*vate Experience*" and "Sense Data", Wittgenstein remarks that when we express our subjective experience, there are circumstances where the possibility of ingenuity or error is disregarded – not because ingenuity or error is improbable, but because the practice itself rules it out; the possibility "mustn't enter [the] game"<sup>76</sup>. He continues:

It is nonsense to say: the expression may always lie.

The language games with expressions of feelings (private experiences) are based on games with expressions of which we don't say that they may lie.<sup>77</sup>

Expressive actions have a constitutive role in our conceiving the thoughts and feelings of ourselves and others. This is connected with the Wittgensteinian idea that some outward, observable phenomena (crying, moaning, laughing, grimacing, smiling...) have a *critical* relation to the subjective experiences of sorrow, pain, joy, elation and so on, where this critical relation is stronger than that of being inductive evidence, but still not implying behaviorism. Investigating this idea has been the task of Article IV in this dissertation.

As pointed out in Article IV, an influential reading has interpreted the Wittgensteinian criteria for a given mental state (as an example, let us think of crying as a criterion for sadness) as a special kind of evidence. This reading recognizes that crying is more than empirically grounded inductive evidence for sadness, but keeps the idea that crying is external *evidence* for another thing – namely, the unobservable inner state of sadness. The reading claims that the difference between criteria and weaker forms of evidence is that the former are based on "grammar" rather than based on empirical discovery. At least roughly, this would put the relation between crying and sadness in the following way: What we *mean* by "sadness" is the kind of inner mental state that typically occurs in people when they cry. This is supposed to be grammatical elucidation, an explanation of what sadness is.

But I think this does not take Wittgenstein's critique of inner and outer far enough. In fact, the previous reading still conforms to the picture of private inner objects that Wittgenstein targets in the famous beetle-in-the-box passage.<sup>78</sup> The reading admits that sadness is, in the last analysis, an object in a person's inner world, although it is necessarily identified as such via crying and other kinds of critical evidence. But Wittgenstein's dialectic rather guides us to view sadness, and other such mental states, as phenomena that take place in a shared human life, in our interactions with others. Sometimes he poetically refers to such phenomena as "patterns in the weave of life".<sup>79</sup> Criteria for sadness, such as crying, are better thought as those kinds of things that can in suitable circumstances be (parts of) visible and concrete occurrences of sadness – the phenomenon of sadness embodied in a human being. Hence the attention given in Article IV for arguing that the relation between

---

<sup>76</sup> LWL, 245. The passage reads: "When I say that moaning is the expression of [toothache], then under certain circ[umstance]s the possibility of it being the expression without the feeling behind it mustn't enter my game."

<sup>77</sup> NFL, 245

<sup>78</sup> PI §293

<sup>79</sup> See e.g. PI II, i; LW II, 42, 61

Wittgensteinian criteria and what they are criteria for should be constitutive rather than evidential.

We learn to talk about sadness and to conceptualize it by responding to crying people. It is only on the basis of this that we subsequently learn to conceive of sadness as a private experience that can either be held back or expressed; and then we learn to refer to different kinds of embodied manifestations of sadness, both primitive and sophisticated, as expressions of sadness.

An expression cannot always lie, because there is no further court of appeal than the naturally occurring, primitive expressions of subjects when we assess the thoughts and feelings of ourselves and others. The picture that Wittgenstein opposes, the picture of the private inner, holds that the last court of appeal are the objects of subjective experience, considered ready-made and in isolation. The picture implies that our being able to express ourselves is dependent on our identifying our thoughts and feelings. But Wittgenstein's preferred view has it the other way round: Our identifying our thoughts and feelings is dependent on our ability to express ourselves. Basic human expressive acts and reactions – cries of pain, smiles of joy, tears of sadness – provide a kind of an anchoring point for understanding the feelings of ourselves and others; they are feelings in an intersubjectively observable form. They are parts of what pain, joy or sadness looks like when it occurs in human life. In certain circumstances, they make redundant the question of whether there is a mental phenomenon “behind” the expression, hidden in the person's mind, because they are *themselves* the phenomenon. Spontaneous expression brings the thoughts and feelings of its subject out into the open, into the publicly observable concrete world, in a manner that is in a certain way perfectly symmetrical. Namely, spontaneous, basic expressions like tears of sorrow or bursts of laughter are, even from the subject's own perspective, things that just objectively “happen”, without being intentionally produced or put forward for the information of others. My tears of sorrow can surprise me as well as they can surprise others. And it can even sometimes happen that I am quite unwilling to accept the testimony of my own expressive behavior. If someone cries without quite knowing why, she may be unwilling to ascribe sorrow to herself. This may be because, say, we take sorrow to always be an intentional state, directed to something, and to involve some kind of an evaluative judgment. And if the crying person sincerely cannot produce any object for her putative intentional state, or explicate the content of the evaluative judgment, then she may just judge herself to be confused, delusional or something like that, and judge that whatever she is feeling does not qualify as real, legitimate, sadness. But this is where – arguably – the intellectualizing of the experiential world should stop. When I cry, this should be an indication for both myself and others that I am somehow inside the scope of those mental phenomena that have crying among their criteria. Typically it is sadness, but of course people cry out of anger too, or because of some other type of being moved. Coming into terms with what my feeling actually is may require extended reflection (as well as help from others); and this coming into terms may be much more like construction work, or creating an autobiographical story, than an exercise in introspection. But the primitive fact that some real feeling is taking place in me in this situation should be removed from the domain of argumentation and

negotiation, and placed in the domain of objective facts. By the natural expressivity of crying, my feeling becomes real for me as well as for others.

Where does Wittgenstein's philosophy of inner and outer leave the traditional skeptical problem of other minds, namely, the ever-present possibility to doubt whether other people have thoughts and feelings at all? Maybe the closest Wittgenstein comes to addressing the problem in its traditional form is PI §420:

But can't I imagine that the people around me are automata, lack consciousness, even though they behave in the same way as usual? -- If I imagine it now - alone in my room - I see people with fixed looks (as in a trance) going about their business - the idea is perhaps a little uncanny. But just try to keep hold of this idea in the midst of your ordinary intercourse with others, in the street, say! Say to yourself, for example: "The children over there are mere automata; all their liveliness is mere automatism." And you will either find these words becoming quite meaningless; or you will produce in yourself some kind of uncanny feeling, or something of the sort.

Seeing a living human being as an automaton is analogous to seeing one figure as a limiting case or variant of another, the cross-pieces of a window as a swastika, for example.

Wittgenstein's reaction here is not to give an argument against the traditional problem, or even to deny that such skeptical imagination can, in some sense, be undertaken. It is rather to point out the idleness of this kind of doubt, its lack of pragmatic consequences. The later Wittgenstein's philosophy of mind has the effect of showing how relatively uninteresting the traditional skeptical problem of other minds is compared to the underlying picture of epistemic asymmetry and inner-outer divide, of which the traditional problem of other minds is just a part. The underlying picture, however, is still very much part of our philosophical thinking, and consequently the traditional problem of other minds continues to subtly appear in contemporary philosophy. Article II is all about one such case: David Chalmers's anti-physicalist "zombie argument". Chalmers holds it as evident that each subject's all knowledge about the existence of consciousness stems from one's own case; even complete knowledge of the objectively available physical facts of the universe could not demonstrate with certainty that other subjects have consciousness like I do. He explicitly uses this view of epistemic asymmetry to argue that facts about consciousness do not logically supervene on physical facts, and thus that a certain type of physicalism is false. In Article II, we argue that the idea of epistemic asymmetry is inherently so problematic that it does not help Chalmers's argument (although the argument may be helped by other means). Crucially, what is wrong with the idea that knowledge about consciousness is strictly first-personal is that this view cannot satisfactorily explain how we come to ascribe consciousness to others in the first place. Wittgenstein argues forcefully that conceiving myself as a subject of private experiences is dependent on being able to first recognize mental phenomena in a shared human life, through responding to the expressive actions of others. Chalmers's view has it that a subject begins with a primitive awareness of her own mental realm, noticing that "*I have this*", as if pointing to the subjective experience, and then doubting whether others have it too. But this overlooks the argument that our ability to conceive of ourselves having "*this*", a private experience, is dependent

on our more fundamental readiness to respond to other people's expressive behavior, which already involves treating them as on par with myself.

Wittgenstein's remarks in PI §420 suggest that taking other people as "zombies", or automata, is to *see* people in a certain way, rather than entertaining a meaningful hypothesis about them. One can perhaps draw a parallel to another series of passages, Z §§528-530, where Wittgenstein engages in a thought-experiment:

An auxiliary construction. A tribe that we want to enslave. The government and the scientists give it out that the people of this tribe have no souls; so they can be used for any arbitrary purpose. Naturally we are interested in their language nevertheless; for we certainly want to give them orders and to get reports from them. We also want to know what they say to one another, as this ties up with the rest of their behaviour. But we must also be interested in what corresponds in them to our '*psychological utterances*', since we want to keep them fit for work; for that reason their manifestations of pain, of being unwell, of depression, of joy in life, are important to us. We have even found that it has good results to use these people as experimental subjects in physiological and psychological laboratories, since their reactions - including their linguistic reactions - are quite those of mind-endowed human beings [...]<sup>80</sup>

The passage continues by picturing the tribesmen as behaving in the same manner as what in our own case suggests the presence of "inner" states, and by picturing us to be interested in the tribesmen's reports and reactions in the same way as those of regular human beings. It is not simple to interpret this passage, but it seems to suggest this: We would interact with the tribesmen in all the ways we interact with regular human beings, and in doing so we would neither contradict nor affirm our conviction that they "have no souls" - because that conviction does not have real consequences as a hypothesis. It is remarkable that Wittgenstein mentions in the thought-experiment that we want to enslave the tribe. He seems to be hinting that the officially held doctrine that these people have no souls could only serve the purpose of giving us the license to not morally care about the tribesmen - a license that ordinary life with the tribesmen could never support.

---

<sup>80</sup> Z §528. See also RPP I §§96-101.

## YHTEENVETO (FINNISH SUMMARY)

Tutkimukseni aihe on toisten mielen ongelma Ludwig Wittgensteinin myöhäisfilosofian näkökulmasta. Wittgenstein käsittelee ensimmäisen persoonan auktoriteetin ja ilmaisemisen teemoja sekä toisia ihmisiä havainnon kohteina. Erittelemällä näitä osia hänen filosofiastaan osoitan, että niiden kautta toisten mielen ongelma näyttää monimuotoisena ihmiselämän kysymyksenä, joka on laajempi kuin perinteinen skeptinen ongelma. Yksilöin toisten mielen ongelman peruslähtökohdaksi *episteemisen epäsymmetrian*: ajatuksen, että jokaisella subjektilla on suoraa ja virheetöntä tietoa omista mielentiloistaan mutta ei vastaavaa tietoa toisten subjektien mielistä. Meistä kullakin on erityinen subjektin näkökulma omiin mielentiloihimme sikäli, että voimme sanoilla ja teoilla ilmaista mielentilamme. Se, että meillä on tässä mielessä subjektin näkökulma ainoastaan omiin mielentiloihimme, ei Wittgensteinin mukaan kuitenkaan estä meitä tietämästä tai ymmärtämästä toisten mielentiloja. Artikkelit I yksilöi Wittgensteinin myöhäisfilosofiassa kolme teemaa, joiden yhteydessä hän argumentoi episteemistä epäsymmetriaa vastaan: 1) ensimmäisen persoonan auktoriteetti ja toisten ihmisten mielentilojen tietämisen epävarmuus, 2) välittömän kokemuksen sanallisen kuvaamisen vaikeus ja 3) toisen ihmisen mielentilan näkeminen tämän kehollisen käyttäytymisen kautta. Kunkin näiden teemojen kohdalla Wittgensteinin voidaan nähdä tunnustavan epäsymmetrian subjektin näkökulman ja toisen ihmisen näkökulman välillä, mutta kieltävän sen, että tämä epäsymmetria olisi luonteeltaan tiedollinen. Artikkelit II-IV käsittelevät yhdestä tai useammasta edellä mainitusta teemasta johdettuja kysymyksiä, liittäen jatkokysymykset aiheesta riippuen suoranaisesti tai välillisesti Wittgensteinin myöhäisfilosofiaan. Artikkelit II käsittelee mielenfilosofian nykykeskustelussa tunnettua niin kutsuttua zombiargumenttia, joka on toisten mielen ongelman huomattava esiintymismuoto viimeaikaisessa filosofisessa keskustelussa. Artikkelit III tarkastelee ilmaisemisen ja itsetuntemuksen (*self-knowledge*) välistä yhteyttä arvioimalla niin kutsuttua ekspressivististä itsetuntemuksen teoriaa. Artikkelit IV käsittelee Wittgensteinin myöhäisfilosofiassa käyttämää kriteerin käsitettä. Artikkelit selvittää, missä mielessä toisen ihmisen mielentilojen kehollisten kriteerien havaitseminen – Wittgensteinin tapaan ymmärrettynä – voi olla samalla hänen mielentilojensa havaitsemista.

## ORIGINAL PAPERS

### I

#### WITTGENSTEIN'S "INNER AND OUTER": OVERCOMING EPISTEMIC ASYMMETRY

by

Vaaja, Tero, 2013

Nordic Wittgenstein Review vol 2, issue 1, 107-130

Reproduced with kind permission by the publisher.

Tero Vaaja

## Wittgenstein's "Inner and Outer": Overcoming Epistemic Asymmetry

### ***Abstract***

In this article, I identify three ways in which Wittgenstein opposed an idea of epistemic asymmetry between the first person and the second-or-third person. Examining the questions of 1) absence of doubt about my own experience and uncertainty about the experiences of others, 2) ineffability of subjective experience and 3) immediacy of my knowledge of my own experience contrasted with my merely inferential knowledge about the experiences of others, I see Wittgenstein's remarks about "inner and outer" as a many-faceted denial of the claim that people's minds are in some deep way unknowable to others. These considerations also serve to clarify Wittgenstein's relation to behaviorism.

### ***1. Wittgenstein on other minds***

Wittgenstein is undoubtedly an important philosopher to consider when tracing the history of the so-called problem of other minds. There was a boom of writings on the topic from philosophers of the analytic tradition after the middle of the 20<sup>th</sup> century, probably largely inspired by Wittgenstein's remarks on solipsism and related matters. However, when Wittgenstein's contribution to the problem of other minds is explicitly assessed, the central attention is usually given to rather narrowly limited points in his writings. He is applauded for seeing a conceptual problem in the place of the

traditionally conceived skeptical problem of other minds (Avramides 2001). Sometimes his talk of “criteria” is seen as an attempt to answer a skeptic of other minds directly (Hyslop 1995). Michel ter Hark (1990, 1991) is an exception because he gives a detailed reading of Wittgenstein’s late philosophy of mind from the viewpoint of the other minds problem, and I generally agree with his account. However, I see Wittgenstein fighting against the idea that people’s minds are unknowable to others in several different contexts at once.

I suggest that Wittgenstein’s late writings about the theme of “inner and outer” are a holistic attempt to deny a significant *epistemic asymmetry* between first person and second-or-third person. By epistemic asymmetry I mean the assumption that each of us has a privileged epistemic access to his own mind (and to his own mind only), making our first-personal knowledge of our own minds better in quality, more immediate, or otherwise superior to any interpersonal knowledge.

I examine several contexts where Wittgenstein first identified a tendency to think that our knowledge of mental goings-on in other people is inferior to our first-personal knowledge, and then proceeded to criticize such a tendency. These contexts can be treated as variants of the problem of other minds. I also suggest how Wittgenstein’s late writings about psychology help to see his relation to behaviorism in the right light. Here I have drawn much inspiration from ter Hark (1991).

## **2. Forms of epistemic asymmetry**

I will identify three ways in which there seems to be an asymmetry between a subject’s first-personal knowledge of his own mind and his knowledge of the minds of others. In the main body of this paper, I examine Wittgenstein’s handling of each of these topics in turn.

First, there is *in corrigibility about my own experience, contrasted with some inevitable amount of uncertainty about the experiences of others*. In feeling pain or seeing a patch of red, there is no room for doubt or mistake in my own case. In contrast, it is always at least in principle possible that another person only seems to undergo a certain

experience, while actually he does not. I claim that Wittgenstein did not deny this; he admitted that at least some types of statements about our subjective experiences are incorrigible, but he attributed this incorrigibility to the logical role of those statements as avowals, not to a privileged access. He pointed out that certainty has different standards when we talk about the experiences of others, thus implying that a comparison between first-personal and third-personal knowledge here is misguided.

Second, there seems to be something *ineffable about the fundamental qualities of private experiences*. A characterization or description of an experience cannot fully disclose it to another person; the subject always knows more about his private experience than he can communicate. I claim that for Wittgenstein, such a view was based on an unrealistic view of what the human activity of “describing one’s experiences” is meant to accomplish, or needs to accomplish.

Third, each of us knows his own experience immediately, while we get to know the experiences of others through inferences based on their behavior. When one has an experience, the thing itself is present for the subject, but in observing someone else having an experience, what is present for observation is a piece of behavior which only suggests the presence of an “inner” experience. I claim that here Wittgenstein opposes the implication that seeing the experiences of others manifested in their bodily behavior is a second-best thing compared to some other, better way of being in touch with them.

### ***3. Doubt and certainty***

My exposition of Wittgenstein’s treatment of the first form of epistemic asymmetry has two parts. First, I examine his account of subjective experience statements and their authority. Second, I show how he characterizes our attributions of experience to others and what he has to say about certainty and uncertainty in that context.

### 3.1 Authority of the first person

Wittgenstein repeatedly states that doubt about one's own sense-experience is unintelligible. His favorite example is the impossibility of being wrong about whether one is in pain.

The truth is: it makes sense to say about other people that they doubt whether I am in pain; but not to say it about myself. (PI §246)

[Referring back to PI §283 (“Couldn’t I imagine having frightful pains and turning to stone while they lasted?”)]: Suppose I were in error and it was no longer *pain?* – But I can’t be in error here; it means nothing to doubt whether I am in pain! (PI §288)

The expressions used here are striking: it “makes no sense to say” that I doubt whether I am in pain; it “means nothing” to doubt my pains; “there is no such thing” as my being mistaken about them (LW II: 34-36, see also LW II: 30-31, 92). It is not merely the case that I never go wrong in attributing experiences to myself because I am in a better position to observe those experiences than anyone else. Rather, the claim is a stronger one: doubt is somehow “logically excluded” in this case. My subjective experience is an area where questions about my being right or wrong about the features of my experience are just not relevant; they are never raised and they need not be raised.

Even just in principle, why is doubt not possible? It can be thought that for Wittgenstein, it is a matter of what he calls “grammar”. This means that first-personal immunity to error is a norm upheld in our language. We are just not prepared to count anything as a mistake in the context of such first-personal statements, and we would consider first-personal expressions of doubt baffling:

[I]f anyone said ‘I do not know if what I have got is a pain or something else’, we should think something like, he does not know what the English word ‘pain’ means; and we should explain it to him. [...] If he now said, for example: ‘Oh, I know what “pain” means; what I don’t know is whether *this*, that I have now, is pain’ – we should merely shake our heads and be forced to regard his words as a queer reaction which we have no idea what to do with. (PI §288)

But it is clearly not an arbitrary norm, adopted just because we have decided to trust people on these things. Wittgenstein's private language argument gives reason to say that when someone applies a concept to her present subjective experience, there is no way, even for the subject herself, to make a distinction between a *correct* and an *incorrect* application of that concept in the absence of public and interpersonal criteria. The application of a concept like "pain" to one's subjective experience is rather like a primitive reaction, which is not aptly described as "recognizing one's sensation as one of pain", because there are no criteria for telling the difference between a correct recognition and a misrecognition, outside of what the subject feels appropriate to say in that situation. This, in effect, collapses the distinction between appearance and reality in the case of present first-personal experience. Most importantly, Wittgenstein recognizes that the immunity to error of my first-personal verbal expressions of pain is not the result of my superior epistemic access to my pains, but rather a fundamental feature of how first-personal experience-talk works. Such talk is essentially subjective, in the sense that it does not rely on naming one's sensations according to some objectively assessable criteria. Subjective criteria, on the other hand, are an oxymoron because "rules" that are only followed privately are not rules (as argued in Wittgenstein's discussions about rule-following). This implies that while discussing sense-experiences, I can rightfully apply concepts like "pain" to my subjective experience without relying on any kind of criteria (PI §289). Thus, the situation is as Wittgenstein sums it up in PI §258:

[W]hatever is going to seem right to me is right. And that only means that here we can't talk about 'right'.

This is the reason why error is excluded in principle in these cases: the necessary context required for making a difference between success and error is not there. Thus, there remains the fundamental fact that first-personal statements of one's experience are the undisputed starting point for the language-game of talking about subjective experiences (PI §290). Introducing objective criteria can make these statements revisable in some contexts, but they are always logically primary.

### ***3.2 Certainty and uncertainty in the third person***

Subjective experience is an area where “how things are” and “how things seem to me to be” are collapsed together. That simple point guarantees that whenever a person gives an honest report of their experience, we don’t have to (and we as a matter of fact do not) take into account the possibility that such a report might contain a mistake. But it is essential that the report must be *honest*. There is a huge family of cases where we see another person saying and doing things that suggest, for example, being in pain, but we entertain doubt as to whether she really is in pain. In those cases, we do not suspect the person to be in error about her pain, but we suspect her of being *insincere*. And this threatens to introduce an unbridgeable gulf between my first-personal case, where the possibility of error about my pain is peculiarly absent, and the case of other people, where there always seems to be the possibility of pain-behavior occurring without the person feeling pain: the possibilities of pretending or play-acting.

The problem is not only about the pains and other sensations of fellow human beings; it is a problem about a whole range of mental phenomena. As well as play-acted sensations, there are faked emotions, insincerely stated beliefs, hidden intentions and concealed desires. Regarding all these, our epistemic relation to the minds of others seems to be marred with incurable uncertainty: there is always the possibility of the appearance being different than the reality, and we are never able to check what the reality in the other person’s realm of consciousness is like. This incurable uncertainty about whether the appearance given by others matches the reality of their inner experience is in the strongest possible contrast with the utter lack of appearance-reality distinction in first-personal experience. Is this not a fundamental kind of asymmetry between knowing oneself and knowing others?

Wittgenstein’s reply to this problem is twofold. First, he argues that the incurable uncertainty has no real place in people’s lives in practice; it exists merely in philosophical reflection, where it gives rise to the idea of skepticism about other minds. Second, he argues that insofar as there is some fundamental element of uncertainty in our attributions of mental states to others, this uncertainty should

be viewed as an essential part of our relations to one another, and not as an epistemic shortcoming.

The basic point made by Wittgenstein about pretense is that pretending is a “complicated pattern” (e.g. LW II: 55) that has to be learned like any other sophisticated skill. There are natural expressions of sensations and other experiences that are in place long before anything like the ability to pretend makes an appearance. Apparently Wittgenstein is also here suggesting what he probably would call a “grammatical” point: mastering the concept of pretense, that is, being able to pretend and take some displays of behavior in others as cases of pretense, necessarily requires being able to take some displays of behavior in others as *genuine* expressions. Because pretense requires such a concept-mastery acquired through a complex interaction between human beings, the cases where pretense is even imaginable are actually rather limited. We never normally take into account the possibility that a newborn child might pretend, and for a good reason: not because we know newborn children to be honest, but because the prerequisites for anything to be called an act of pretense (or honesty) are not fulfilled in the case of the newborn (e.g. LW II: 39-40; PI II, xi: 194).

The idea that there is something blocking us from ever being certain of what goes on in others is connected with a false philosophical idea of “essentially inner events that no one else, in principle, could witness and which I am unable to reveal or describe to another person” (Moran 2001: 91). The truth is that “in countless cases” (LW II: 94) we are perfectly certain about the mental processes in someone else. This is clearly true whenever we recognize something as a universal natural expression of an experience occurring in normal circumstances. When seeing someone being burned by a flame and screaming, there is just no point in thinking: “of course there are always two possible cases; one of pain-behavior with pain-experience and one of pain-behavior without pain” (see LPE: 287). Doubting the authenticity of this situation would not have the normal consequences of doubt. It would rather be like trying to doubt that the future is connected with the present (see PI II, xi: 190).

So, in many occasions we are justifiably certain in our attributions of experiences to others. But the second part of Wittgenstein's reply to these worries is to point out that certainty and uncertainty regarding the experiences of others are of a special type, which gets misrepresented if it is contrasted with, for example, mathematical certainty:

In mathematics a particular kind of evidence that can be clearly presented leaves no doubt open. That is not the way it is when we know that someone was glad.

There can't be a long dispute in a court of law about whether a calculation has this or that result; but there certainly can be about whether someone was irritated or not.

But does it follow that one can know the one and not know the other? More likely what follows is that in the one case one almost always knows the decision, in the other, one *frequently* doesn't. (LW II: 85)

Rather than saying that knowledge of other minds is inferior to knowledge of some other things, Wittgenstein characterizes knowledge of other minds as being of a different *kind*. Mathematical certainty is generally achieved through a definite procedure that is not controversial. But there is no definite procedure for assessing people's reports and expressions of their sincerity; we cannot lay down anything resembling a proof here. In Wittgenstein's terms, the "language-game" played by experience-ascriptions is altogether different from those played by mathematical concepts; it does not include a determinate set of rules.

I can be as *certain* of someone else's sensations as of any fact. But this does not make the propositions 'He is much depressed', ' $25 \times 25 = 625$ ' and 'I am sixty years old' into similar instruments. The explanation suggests itself that the certainty is of a different *kind*. – This seems to point to a psychological difference. But the difference is logical. [...]

The kind of certainty is the kind of language-game. (PI II, xi: 190-191)

What is essential to the practice of judging the sincerity of the experience-reports of others is that it is based on evidence that is often extremely complicated and difficult to characterize; and that

in it nothing plays the role of conclusive evidence which everyone is forced to accept on pain of irrationality. Wittgenstein briefly remarks that it is partly based on “imponderable” (*unwägbar*) evidence (PI II, xi: 194; LW II: 95) that includes “subtleties of glance, of gesture [and] of tone” (PI II, xi: 194). Moreover, I might be “quite incapable of describing the difference” between such subtleties that for me make the difference between a genuine and a pretended expression in the other (PI II, xi: 194). If two people disagree on how to assess this subtle evidence, there is no universally valid procedure to solve such disagreements:

I am sure, *sure*, that he is not pretending; but some third person is not. Can I always convince him? And if not is there some mistake in his reasoning or observations? (PI II, xi: 193. See also Z §§554-556.)

These are clearly rhetorical questions that are meant to be answered in the negative. Anyway, Wittgenstein also remarks that there is such a thing as a better and worse judgment about the experiences of others; knowledge of people (*Menschenkenntnis*) is a skill that can be learned. But what one in this case learns is

...not a technique; one learns correct judgments. There are also rules, but they do not form a system, and only experienced people can apply them right. Unlike calculating-rules.

What is most difficult here is to put this indefiniteness, correctly and unfalsified, into words. (PI II, xi: 193)

A report or confession of one’s experience is the authoritative account of his experience to others, and when questions are raised about the sincerity of the report, these cannot be settled by any simple and stable set of criteria. Such questions are in a way open-ended. But it is a thoroughgoing motive in Wittgenstein’s discussions that this should not be seen as a philosophically significant *flaw* in our knowledge of other minds; it is just a constitutive difference. There is no good reason to compare our knowledge of other minds unfavorably to other areas of knowledge in this respect, and in particular not to the subject’s knowledge of her own mental states. Here Wittgenstein battles against epistemic asymmetry by not taking my first-personal accounts of my experience as the paradigm case of certain knowledge, and on the

other hand by respecting our knowledge of the mental states of others as its own type of knowledge, with its own peculiar characteristics.

#### **4. Ineffability**

The second form of epistemic asymmetry concerns the thought that it seems impossible to put the essential qualities of my first-personal sensations, feelings and experiences into words. There appears to be something “ineffable” about subjective experience. It is a familiar fact that sometimes, in the face of highly unusual and novel experiences, words fail to capture them. This is a relevant issue in dream research, for example (Revonsuo 2010: 84).

But it can be argued that it is not only unusual experiences but *conscious experiences in general* that have an ineffable element in them. It is one thing to point out that, as Hume in his *Treatise of Human Nature* noticed (T 1.1.1.9; SBN 5), spoken or written words cannot produce the taste experience of pineapple in someone who is not already familiar with it; however, this does not mean that we could not still describe the taste of pineapple in words. But it is a further thing to argue that even the words we legitimately use in such descriptions are not about the intrinsic qualities of the sensation as such, but describe them in a roundabout way, via metaphor and comparison with publicly accessible entities. David Chalmers writes:

We have no independent language for describing phenomenal qualities. [...] Although greenness is a distinct sort of sensation with a rich intrinsic character, there is very little that one can say about it other than that it is green. In talking about phenomenal qualities, we generally have to specify the qualities in question in terms of associated external properties, or in terms of associated causal roles.

[...] When we learn the term ‘green sensation’, it is effectively by ostension – we learn to apply it to the sort of experience caused by grass, trees and so on. (Chalmers 1996: 22)

I assume it to be uncontroversial that Wittgenstein would agree with Chalmers at least in this sense: sensations or feelings cannot be a basis for an independent language, separated from a public communicative context which gives experience-words their

normative properties. I will not go into any detail with this, but rather assume it as a given that this is Wittgenstein's position: he indeed recognizes that we can name and describe subjective experiences only with the help of publicly identifiable objects. What he denies, I argue, is that this poses any special problem for communicating them. Instead he recognizes the variety of situations in which and purposes with which people describe their subjective experiences to each other, and denies that there is any goal or purpose that communication about subjective experiences is constitutionally unable to achieve.

As Lagerspetz (2012) has observed, the tendency to think that there is some impossibility in principle of describing subjective experiences stems from failing to appreciate descriptions as actions in a communicative context. Wittgenstein emphasizes the *uses* in which descriptions of experiences are put in human life. As shown in e.g. PI §244, Wittgenstein saw first-personal experience-talk as importantly connected with primitive expressions. For reasons of space, I will not here go deeper into the possible interpretations and limitations of Wittgenstein's expressivism and his non-cognitive thesis of avowals (for discussion, see Rodriguez 2012; Hacker 2005; Bar-On & Long 2001; Macarthur 2010; Robjant 2012). I merely note the general point, made abundantly clear by the PI II section ix, that Wittgenstein saw utterances of first-personal subjective experience as capable of serving both expressive and descriptive roles, with mixed and intermediate cases.

The point that avowals sometimes are expressive in nature is enough to alleviate the problem of ineffability of subjective experience to some extent. As Macarthur (2010) explains, the later Wittgenstein is opposed to the assumption that language, in general, always serves some one simply definable function. In the particular case of first-personal experience-talk, he is opposed to the idea that all such talk is in the business of describing some "inner" event. Rather, in many contexts what is primarily at issue is not the descriptive accuracy of what the subject says, but its status as an avowal; an expressive utterance of the subject, which invites the hearers to attend to her. Moran (2001) elaborates, albeit primarily in the context of beliefs, the way in which avowing and

describing one's state of mind are importantly connected. So even if it were true that we cannot display the phenomenal qualities of our experiences by verbal descriptions in any simple way, that does not mean that we cannot communicate them at all, because much of first-personal experience-talk is not descriptive anyway.

In PI II xi, Wittgenstein illustrates the many facets of such talk:

Are the words 'I am afraid' a description of a state of mind?

I say 'I am afraid'; someone else asks me: 'What was that? A cry of fear; or do you want to tell me how you feel; or is it a reflection on your present state?' -- Could I always give him a clear answer? Could I never give him one?

We can imagine all sorts of things here, for example:

'No, no! I am afraid!'

'I am afraid. I am sorry to have to confess it.'

'I am still a bit afraid, but no longer as much as before.'

'At bottom I am still afraid, though I won't confess it to myself.'

'I torment myself with all sorts of fears.'

'Now, just when I should be fearless, I am afraid!'

To each of these sentences a special tone of voice is appropriate, and a different context. (PI II, ix: 160)

Some of these cases are clearly descriptive ones. There is such a thing as describing one's subjective experience, and such a description can be successful or unsuccessful in communication. But most importantly, Wittgenstein points out that "[w]hat we call '*descriptions*' are instruments for particular uses" (PI §291, emphasis in the original; see also e.g. PI II, xi: 170-171). Therefore, whether a description is successful or unsuccessful is dependent on the purpose for which the description was put forward, and on whether it produces a desired kind of understanding between the speaker and the hearer. It is clear that in different contexts, a good description of a subjective experience will amount to different things. At a doctor's office, a description of one's pains serves its purpose if it makes the necessary distinctions about the location of the pain, its quality, intensity, frequency, duration and so on, enabling the clinician to form a hypothesis about its cause. Such an

account would be a very bad description of a similar pain in a tragic poem, for example; and similarly, what counts as a brilliant poetic simile of excruciating pain might be completely useless for a doctor.

Pains can be described as dull, splitting, burning, sharp, stabbing, and so on. It can be argued that such descriptions are possible only by borrowing our vocabulary from the category of public objects, and that such a derived way of talking can never fully communicate the first-personal, experienced character of pain. I think Wittgenstein has an implicit, if not an explicit, counterargument to the latter claim. I will present it here following Lagerspetz (2012).

Subjective experiences seem ineffable only if we hold a confused view of what counts as a sufficient description. Hume noticed that even the best verbal description of the taste of pineapple, for example, is unable to produce the taste-experience of pineapple in someone who is not already familiar with the taste from his own experience. But descriptions should not even be expected to do such a thing. The way to produce a taste experience of pineapple in someone unacquainted with it is to offer her a suitable *sample* of pineapple. Descriptions of tastes can serve a variety of different purposes, but straightforwardly producing novel taste experiences is not among those purposes. A person can describe the taste of pineapple to someone else so that the other can guess whether she will like pineapple or not; descriptions of different foodstuffs can be used to make a systematic list of them with a number of categories, for practical purposes; or two people tasting pineapple may compare their taste-descriptions to see how their taste-vocabularies differ. In all those contexts what counts as a sufficient description will be relative to the interests of the speaker and hearer. It is not reasonable to expect descriptions to do the same job that samples do, because description is an altogether different instrument. So there is no reason to claim that the inability of descriptions to produce novel subjective experiences in us is a flaw in our ability to describe experiences. (See Lagerspetz 2012: 291-294.)

However, it is still possible to protest that even if it is not the point of descriptions to reproduce the described experience in another person, *no description of a subjective experience can ever be exhaustive*. The felt qualities of experiences like pains and tastes are richer than our vocabulary for describing them, and therefore, the subject of experiences always inevitably knows more about them than she can verbally or otherwise communicate. But this protest might be based on a philosophical prejudice about what counts as an “exhaustive” description. In actual practice, descriptions are said to be inadequate, adequate, vivid or comprehensive in relation to the context of describing and the goals of the persons involved. A description of one’s pain to a doctor is exhaustive insofar as it gives her all the information she needs; an abundant pain-description, explaining all the nuances of the experience, its minute-by-minute development, its exact location and alterations in its location, might from the doctor’s perspective be a *worse* pain-description than a more compact one, because the former will contain lots of redundant information. Descriptions are generally given as answers to actual or potential *questions*, and their adequacy can only be assessed in light of the questions they are meant to answer. (For one particularly good illustration of this, see PI §368.) It does not make any sense to talk about an *absolutely* exhaustive description of one’s experience, abstracted from the communicative context; because it is not clear even for the subject herself what such a description should look like. It is not the case that in principle, it is possible to give a complete description of a subjective experience in the abstract sense, but as a matter of fact we are unable to do that. Rather, there is nothing that would count as such a complete description in the abstract sense, even in principle (Lagerspetz 2012: 290).

Wittgenstein has no reason to see any fatal philosophical problem about our ability to describe our subjective experiences to each other. But this does not mean that he held this kind of communication to be always *unproblematic*. On the contrary, he notes that a “human being can be a complete enigma to another (PI II, xi: 190). Another person may remain such an enigma even when he “does his utmost to make himself understood” (LW II:

28). Wittgenstein sees the breakdowns which sometimes do happen in human communication as breakdowns of *understanding*, that is, failures to relate properly to other people. This can happen when the speaker and hearer do not share a context of common interests, goals and motivations of action, or a background of at least some relevant experiences familiar to both. Understanding, in this case, can be said to consist of the ability to have an exchange of questions and answers which both parties can spontaneously develop and enrich. When this does not happen and the discussion terminates, there will be a feeling that something about the other person remained “hidden”. In normal circumstances, the basic interpersonal attitude which Wittgenstein called our “attitude towards a soul” (PI II, iv: 152) will surely still remain intact. But it will be coupled with uneasiness; uncertainty about what the other is aiming at with her descriptions of her experiences and about how she is using her self-expressive words.

For Wittgenstein, failures to understand what goes on in the minds of others are essentially of this type. They are not the result of some deep ineffability of subjective experiences, but of inability to share contexts of action (or what Wittgenstein called “forms of life”) with others. This is also the proper context for Wittgenstein’s remark: “If a lion could talk, we could not understand him” (PI II xi: 190; LW I §190). If a lion *could* talk, there is no reason why it would not be able to talk about its lion-like subjective experiences as well as humans can talk about theirs. The difficulty would rather be the vast difference between a lion’s life and human life, maybe vast enough to make it impossible for us to find the right questions to ask about lion-like experiences, so that the lion could understand what we want to know of it.

### **5. Inference and behavior**

The third and final form of epistemic asymmetry that I will discuss concerns the relation between our perceptions of human behavior and our beliefs about mental events that are the causes of behavior. Supposedly, our beliefs about the minds of others are formed on the basis of their behavior. The subject’s own beliefs about her mental life, by contrast, are (normally) not grounded in

observations of her own behavior. First-personal experiences are “just felt”; our awareness of them is not grounded in anything further than just the experiences themselves.

When I have a pain, I have direct access to the pain itself, whereas in the case of others I have direct access only to bodily movements like grimaces, gestures, sounds and speech; in short, to pieces of pain-behavior of others. This view has some familiar philosophical consequences. First, it seems that even in ideal circumstances, the best we can do is to become assured that other people *probably* have pains, sensations and other experiences; although this probability will be overwhelmingly high, it will not amount to certainty. Second, it would be strictly wrong to say that pains, sensations and other experiences are ever *perceived* by anyone other than the subject herself; it is only the behavior which suggests the presence of these things that is being perceived. This can be taken to be an existentially flavored philosophical problem in its own right: we cannot meet the minds of others first-hand. Above (in section 3.2) I covered Wittgenstein’s treatment of the first point. I will not here go deeper into Wittgenstein’s notion of certainty in *On Certainty* and elsewhere. Instead I will proceed to the theme of perceiving the minds of others, linking this topic to Wittgenstein’s discussion of aspect-seeing.

### **5.1 Introspection, behavior and evidence**

It would be unintuitive to deny that our evidence of mental events of others is, in a real sense, constituted by their bodily behavior. What Wittgenstein is doing is subtly criticizing the status which we give to that behavioral evidence. The essential point of this criticism is briefly stated in PI §246:

Other people cannot be said to learn of my sensations *only* from my behavior, – for *I* cannot be said to learn of them. I *have* them. (PI §246, all emphases in the original)

The questions of our knowledge of other minds and our knowledge of our own minds are, as ter Hark (1990: 141) puts it, two sides of the same coin for Wittgenstein. For him, the motivation to say that behavioral evidence for the experiences of

others is indirect and insufficient stems from the fact that we compare it with the first-personal case, where we supposedly have another, better kind of evidence. In the first-personal case, we seem to have direct, introspective evidence of our experiences.

Wittgenstein breaks this asymmetry by insisting that it is wrong to construe our relation to (at least some types of) our own mental states in terms of introspective evidence. It is not that I know of my pains and sensations because I introspectively see or feel the sensation or feeling; rather, I simply have the pain or sensation. The relation is even more intimate than the alleged direct introspective access. If I knew of my pains by consulting introspective evidence, then there would be no reason why this introspection could not sometimes go *wrong*, resulting in me being mistaken about my own pains, which Wittgenstein held to be nonsensical (see above, section 3.1). Rather, Wittgenstein writes explicitly:

It is not as if he had only indirect, while I have internal direct evidence for my mental state. Rather, he has evidence for it, (but) I do not. (LW II: 67)

We only construe behavioral evidence for the mental states of others as indirect evidence because we have an idea of some superior type of evidence, compared to which the usual behavioral evidence is a second-best thing. But what we have in the first person is not a good point of comparison, because there the relation between us and our mental states is not evidential (e.g. LW II: 92). Further, it is not clear if any other idea of a superior type of evidence is coherent. In the *Brown Book*, Wittgenstein remarks that “people have often talked of a direct transmission of feeling which would obviate the external medium of communication” (BBB: 185), and proceeds to question whether it makes sense to postulate such a direct medium of communication in contrast of the usual, “indirect” one. Such a medium would be something like telepathy, or what C.D. Broad (1925: 328-330) called “telegnosis”: a cognitive situation where the perceiver would be involved solely with a mental event belonging to another mind. It would be a topic of a whole separate discussion to see whether any such situations can be coherently described. One could conjecture that Wittgenstein

wanted to answer in the negative. Therefore, it is wrong to say that we learn of the sensations of others *only* from their behavior, but right to say that we learn of them (simply) from their behavior. It is just the “only” which is inappropriate.

## **5.2 Behavior and aspect-seeing**

Much of Wittgenstein’s writings on perception and psychology revolve around aspect-seeing (see PI II, xi; LW I §§165-180, 735-785; LW II: 12-18; Z §§208-226). There is a natural link between aspect-seeing and perceiving other minds, which shows itself when Wittgenstein talks about the possibility of seeing other human beings as machines or automata:

But can’t I imagine that the people around me are automata, lack consciousness, even though they behave in the same way as usual? -- If I imagine it now – alone in my room – I see people with fixed looks (as in a trance) going about their business – the idea is perhaps a little uncanny. But just try to keep hold of this idea in the midst of your ordinary intercourse with others, in the street, say! Say to yourself, for example: “The children over there are mere automata; all their liveliness is mere automatism.” And you will either find these words becoming quite meaningless; or you will produce in yourself some kind of uncanny feeling, or something of the sort.

Seeing a living human being as an automaton is analogous to seeing one figure as a limiting case or variant of another, the cross-pieces of a window as a swastika, for example. (PI §420)

This indicates that seeing living human beings as lacking minds is seeing them under a certain aspect. It also indicates that it is a very unusual aspect, one that can be summoned only briefly and only in favorable circumstances. Wittgenstein admits that at least in some sense, people can be imagined to be mindless machines, but points out that such imagining never has *more* than trivial psychological consequences. Our certainty about the minds of others is not threatened in practice; we just cannot see people that way.

In PI II, xi, Wittgenstein explores the range of the concept of “seeing”. He is interested in the conceptual issues around the phenomenon of seeing a picture according to an interpretation, or under a certain aspect, in which case the perceiver in some sense

“sees” things in the picture which are not strictly speaking “in” the picture. One of his examples is the following:

I see that an animal in a picture is transfixed by an arrow. It has struck it in the throat and sticks out at the back of the neck. Let the picture be a silhouette. – Do you *see* the arrow – or do you merely *know* that these two bits are supposed to represent part of an arrow? (PI II, xi: 173, emphases in the original)

Wittgenstein then goes on to say that “it must be possible to give both remarks a conceptual justification” and that the question concerns the sense in which this can be said to be a case of seeing (PI II, xi: 173-174). Eventually, he writes:

‘To me it is an animal pierced by an arrow.’ That is what I treat it as; this is my *attitude* to the figure. This is one meaning in calling it a case of ‘seeing’. (PI II, xi: 175, emphasis in the original)

The term “attitude” immediately reminds us of Wittgenstein’s brief remark in PI II iv, which seems to sum up his view of the nature of human beings’ relations to one another:

My attitude towards him is an attitude towards a soul. I am not of the *opinion* that he has a soul. (PI II, iv: 152, emphasis in the original)

Wittgenstein often discusses seeing feelings and emotions manifested in a human face, and these discussions are entangled with discussions of aspect-seeing, clearly indicating a connection (Luckhardt 1983: 333). Equally clearly he construes the cases of feelings and emotions in others as cases of seeing. In *Zettel* he makes this very explicit (Z §§220-226). When seeing an aspect in a picture, we see both the picture and the aspect, but we do not see them as two separate things. Similarly with seeing an expressive human face:

What do psychologists record? -- What do they observe? Isn’t it the behaviour of human beings, in particular their utterances? But *these* are not about behaviour.

‘I noticed that he was out of humour.’ Is this a report about his behaviour or his state of mind? [...] Both; not side-by-side, however, but about the one *via* the other. (PI II, v: 153, emphases in the original. See also RPP I §§287-292.)

The point is that there is a conceptual distinction to be made between the sense in which we see the physical features of a face and the sense in which we see a feeling manifested in them; but both are cases of seeing. The latter is not inferred from the former (Z §225). Rather it is displayed in the former, leading Wittgenstein to say that “the human body is the best picture of the human soul” (PI II, iv: 152).

Of course, part of the interest of the phenomenon of aspect-seeing is the possibility of aspect-blindness. It can happen, for a variety of reasons, that someone is unable to see a feeling manifested in the behavior of the other person. Here, as much as with the theme of ineffability, it should be noted that Wittgenstein is not claiming interpersonal relations to be always epistemically unproblematic. What he does claim is that the minds of others are not in any peculiar way *inaccessible* to us. I might not always be able to see the feeling manifested in the facial expression of the other person; but when I do see it, I literally see it *in* the face, rather than somewhere behind it.

## **6. Concluding remarks: Wittgenstein and behaviorism**

Wittgenstein’s way of studying mental phenomena quite fundamentally involves the thought that there is no privileged first-person perspective that is helpful in understanding what mental phenomena are (e.g. RPP II §§31-35, 531; PI §314, 413; PI II, xi: 174). To that extent, his approach has a methodologically behaviorist tone. But logical behaviorism is in no way attributable to him; his expressive analysis of first-personal experience-talk rules that out, as Luckhardt (1983), Fogelin (1976: 174-176) and others have observed. In his (1991), ter Hark offers a reading which shows how Wittgenstein’s “attitude towards a soul” and his remarks about the open-ended nature of experience-attributions to others (see 3.2 above) also preclude his account of third-personal experience-talk from being interpreted as logically behavioristic. Anyway, there is still a sense of ambivalence in Wittgenstein’s relation to behaviorism. I think his essential critique of it can be put in terms of inner and outer: Wittgenstein is opposed to behaviorism insofar as it construes the behavior we observe in

others as “outer” events, which forces one to deny the existence of the things we usually call “inner” processes (see PI §308).

What Wittgenstein implies is that the behaviorist is right to insist on the publicity and observability of mental phenomena, but that the notion of “behavior” which underlies or is suggested by behaviorism is misguided. It is an impoverished concept of “mere behavior” (see LPE: 278-279). The bodily movements of others are not mere behavior to us, comparable to the “behavior” of gases or planets, because the former have a uniquely human aspect for us. They are joy-behaviors, pain-behaviors and sorrow-behaviors, and we know what joys, pains and sorrows are by living as members of a community where such things occur.

Introspectionist psychology attempts to study the “inner” events of human consciousness, which are reachable by introspection and then reported through speech or some other medium; but in any case through some “outer” event. Wittgenstein is opposed to this dichotomy. But he is also opposed to behaviorism insofar as it remains trapped in this dichotomy and only denounces one half of it, claiming that those which are usually called “inner” events are in reality nothing but “outer” events (PI §§304-314; LPE: 278-284). The right thing to do is to see the talk of “inner” experiences and their “outer” manifestations as tools for making conceptual distinctions in the continuous fabric of human life. That fabric is whole and open to view as it is, and most of the time it does not invite the quite specialized distinction of inner and outer.

## **References**

- Avramides, A., 2001. *Other Minds*. London and New York: Routledge.
- Bar-On, D. and Long, D.C., 2001. “Avowals and First-person Privilege”. *Philosophy and Phenomenological Research*, vol. 62, no. 2, pp. 311-335.
- Broad, C.D., 1980/1925. *The Mind and Its Place in Nature*. London and Henley: Routledge & Kegan Paul.
- Chalmers, D., 1996. *The Conscious Mind*. Oxford and New York: Oxford University Press.
- Fogelin, R., 1976. *Wittgenstein*. London, Henley and Boston: Routledge & Kegan Paul.

- Hacker, P.M.S., 2005. "Of Knowledge and of Knowing that Someone Is in Pain". In: A. Pichler and S. Säätelä, eds. 2005. *Wittgenstein: the Philosopher and His Works*. Working Papers from the Wittgenstein Archives at the University of Bergen, no. 17, pp. 203-235.
- Hume, D., 2000/1739. *A Treatise of Human Nature*. Ed. D. F. Norton and M. J. Norton. Oxford: Oxford University Press.
- Hyslop, A., 1995. *Other Minds*. Dordrecht: Kluwer Academic Publishers.
- ter Hark, M., 1990. "Wittgenstein's Metaphysics of the Inner and Outer". *Daimon. Revista Internacional de Filosofía*, no. 2, pp. 139-150.
- ter Hark, M., 1991. "The Development of Wittgenstein's Views about the Other Minds Problem". *Synthese*, vol. 87, pp. 227-253.
- Lagerspetz, O., 2012. "Miltä ananas maistuu? Kokemuksen kuvaamisesta" ("What Does Pineapple Taste Like? On Describing Experiences"). In: V. Viljanen, H. Siipi and M. Sintonen, eds. 2012. *Ymmärrys (Understanding)*. University of Turku: Reports from the Department of Philosophy, vol. 25, pp. 288-295.
- Luckhardt, C.G., 1983. "Wittgenstein and Behaviorism". *Synthese*, vol. 56, pp. 319-338.
- Macarthur, D., 2010. "Wittgenstein and Expressivism". In: D. Whiting, ed. 2010. *The Later Wittgenstein on Language*. Basingstoke: Palgrave Macmillan, pp. 81-95.
- Moran, R., 2001. *Authority and Estrangement. An Essay on Self-Knowledge*. Princeton and Oxford: Princeton University Press.
- Revonsuo, A., 2010. *Consciousness: The Science of Subjectivity*. Hove and New York: Psychology Press.
- Robjant, D., 2012. "Learning of Pains; Wittgenstein's Own Cartesian Mistake at Investigations 246". *Wittgenstein-Studien*, vol. 3, issue 1, pp. 261-286.
- Rodríguez, Á., 2012. "How to Be an Expressivist about Avowals Today". *Nordic Wittgenstein Review* 1, pp. 81-101.
- Wittgenstein, L., 1969. *The Blue and Brown Books*. 2nd ed. Malden, Oxford and Carlton: Blackwell. [BBB]
- Wittgenstein, L., 1982. *Last Writings on the Philosophy of Psychology, volume I*. Ed. H. Nyman and G.H. von Wright, trans. M. Aue and C.G. Luckhardt. Oxford: Blackwell. [LW I]
- Wittgenstein, L., 1992. *Last Writings on the Philosophy of Psychology, volume II*. Ed. H. Nyman and G.H. von Wright, trans. M. Aue and C.G. Luckhardt. Oxford: Blackwell. [LW II]
- Wittgenstein, L., 1993. "Notes for Lectures on 'Private Experience' and 'Sense Data'". In: L. Wittgenstein, *Philosophical Occasions 1912-1951*. J. Klagge and A. Nordmann, ed. 1993. Indianapolis & Cambridge: Hackett Publishing Company, pp. 202-288. [LPE]

- Wittgenstein, L., 1980. *Remarks on the Philosophy of Psychology, volume I*. Ed. G.E.M. Anscombe and G.H. von Wright, trans. G.E.M. Anscombe. Oxford: Blackwell. [RPP I]
- Wittgenstein, L., 1980. *Remarks on the Philosophy of Psychology, volume II*. Ed. H. Nyman and G.H. von Wright, trans. M. Aue and C.G. Luckhardt. Oxford: Blackwell. [RPP II]
- Wittgenstein, L., 2001/1953. *Philosophical Investigations*. Ed. G.E.M. Anscombe and R. Rhees, transl. G.E.M. Anscombe. 3rd rev. ed. Malden, Oxford and Charlton: Blackwell. [PI]
- Wittgenstein, L., 1981. *Zettel*. Ed. G.E.M. Anscombe and G.H. von Wright, transl. G. E.M. Anscombe. 2nd ed. Oxford: Blackwell. [Z]

### **Biographical note**

Tero Vaaja is a researcher/doctoral student in philosophy at the University of Jyväskylä, Finland. In his PhD project, conducted at the Department of Social Sciences and Philosophy, he investigates the variations of the problem of other minds, tracking the historical context in which the problem has emerged, its connections with notions like intersubjectivity, consciousness and solipsism, as well as its role in several contemporary discussions. His work has been kindly supported by the Finnish Cultural Foundation and the Faculty of Social Sciences of the University of Jyväskylä.

## II

### REDUNDANCY OF THE ZOMBIE ARGUMENT IN THE CONSCIOUS MIND

by

Heikinheimo, Antti & Vaaja, Tero, 2013

Journal of Consciousness Studies vol 20, no. 5-6, 6-26

### **III**

## **EXPRESSIVISM, SELF-KNOWLEDGE, AND DESCRIBING ONE'S EXPERIENCES**

by

Vaaja, Tero, 2014

Journal of Mind and Behavior vol 35, no. 3, 151-166

Reproduced with kind permission by the publisher.

## Expressivism, Self-Knowledge, and Describing One's Experiences

Tero Vaaja

*University of Jyväskylä*

In this article, I defend an account of self-knowledge that allows us a considerable first-person authority regarding our subjective experiences without invoking privileged access. I examine expressivism about avowals by contrasting it with “detectivist” and “constitutivist” accounts of self-knowledge, following the use of these terms by David Finkelstein. I proceed to present a version of expressivism that preserves some of the valid motivating insights of detectivism and constitutivism as essential parts. Finally, I point out how my account views self-knowledge as a cognitive and conceptual ability that can be cultivated; the account construes self-knowledge as a process.

Keywords: expressivism, first-person authority, avowal

Each of us is normally the best person to ask when it comes to our own feelings and experiences. Speaking about one's own mental states is generally held to carry a special epistemic authority. Moreover, this authority belongs exclusively to the first person; others are not admitted to have a similar claim to know someone's experiences even if they are extremely well-informed and familiar with them. I take these to be facts on first-person authority as they appear in the practice of human life quite universally.

Such authority has a central place in social life; denying it can easily (and legitimately?) be taken as an offence. However, it might be that philosophers have historically been overconfident about the special security of our knowledge of our own minds. Carruthers (2011) argues that self-knowledge is interpretive and prone to confabulation. Schwitzgebel (2011; Hurlburt and Schwitzgebel, 2007) claims that we might be regularly wrong about even quite fundamental features of our conscious experience. Therefore it is important to be clear

---

Correspondence considering this article should be addressed to Tero Vaaja, Yhteiskuntatieteiden ja filosofian laitos, PL35 (Ylistönmäentie 33), 40014 Jyväskylän yliopisto, Finland. Email: [tero.t.vaaja@jyu.fi](mailto:tero.t.vaaja@jyu.fi)

about the nature of first-person authority, and the conditions in which it may be legitimately challenged.

In this article, I seek to give a modest account of self-knowledge that still respects the special status of the subject as a knower of her own mental states. I treat commonsensical first-person authority as an explanandum, setting aside accounts that seek to dethrone the notion altogether. I start by presenting two contrasting views about the nature of self-knowledge and the basis of first-person authority. I point out how each of these views, “detectivism” and “constitutivism,” is unsatisfactory and how expressivism about avowals, an idea inherited from Wittgenstein (1953), can be seen as preferable to them. I owe the terms detectivism and constitutivism, as well as the main drift of the argument in the first half of this paper, to Finkelstein (2003). Another way to refer to these two contrasting views would be to call them (species of) empiricism and rationalism about self-knowledge, as is done in Gertler (2011). I proceed to present a version of expressivism that incorporates some of the good insights made by detectivism and constitutivism. As explained in the conclusion, I hope my view to be meritorious in respecting commonsensical first-person authority without invoking privileged access, i.e., an idea of a special epistemic channel that makes self-knowledge unproblematic to come by. I also seek to do justice to the meaning of “self-knowledge” as a process that has to do with the personal development of one’s conception of oneself.

### Detectivism

What is it that makes psychological self-ascriptions, or avowals, especially secure?<sup>1</sup> One way of answering is to appeal to introspection, combined with some form of privileged access. The idea is simple: people come to know what their own mental states are like because they are the ones who directly feel or perceive those states. We are assumed to have an “inner sense,” or some naturally evolved capacity that enables us to inwardly monitor our mental states. These are forms of what Finkelstein (2003) calls detectivism: the view that the source of self-knowledge is a perceptual or quasi-perceptual act of detecting that allows us to find out our own mental states.

So, one possible explanation for first-person authority is a combination of two ideas: first, there is a special way of detecting one’s own mental states; and second, that way of detecting is remarkably reliable. Maybe subjects are not completely infallible about everything that goes on in their conscious experience,

---

<sup>1</sup>I will use “avowal” as an umbrella term to refer to any sincere utterance whereby the subject speaks about her mental condition. This liberal use is not a standard one. According to more restricted uses of the term, what I will later refer to as primitive avowals and intellectual self-ascriptions would not necessarily qualify as avowals.

but they have such a propensity of being right about those things that it cannot be paralleled by any other person.

It is hard to deny that in an obvious sense, the subject of a painful sensation is in a better position to observe that particular pain than anyone else. But it is still far from obvious that this is what grounds the typical way in which first-person authority is granted to subjects, or if this is a good account of what self-knowledge is. Next, I attempt to illustrate the issue by an example; my chosen example in this paper will be a case of describing a sensation of pain.

### *Example 1*

I have an abdominal pain that I need to describe to a physician. I am able to point out its location and give an evaluation of its intensity on a scale of 1 to 10. I will also describe its qualitative character by a few adjectives. After careful consideration and some effort to find the right words, I say (at time  $t_1$ ) that my pain is located about ten centimeters up from my waistline, on the left side of my middle abdomen, its intensity is 6, and it is stinging, sharp, distressing, and penetrating.

When I have finished giving my description, I overhear the word “rip,” or someone suggests it to me. I say (at time  $t_2$ ): “Ripping! Yes. That’s what my pain is like. That’s right; I could not come up with it myself.”

When I eventually say that my pain is ripping, I presumably say it with first-person authority. The fact that I needed help in finding the word might give reason for an interlocutor to not take it completely at face value; a question like “Are you sure that is the right word?” might be justifiable. But if I say sincerely and after careful consideration that “ripping” describes my pain perfectly, it is unclear what could ground the claim of someone who insists that I must nevertheless be wrong. In this kind of a situation, any doubt that another person might harbor about the appropriateness of my pain-description will more naturally target my adeptness in the use of the word, rather than the accuracy of my introspective act.

According to detectivism, my statements about my pain are based on perceptual or quasi-perceptual observing. In this case, I am supposedly monitoring my sensation of pain and detecting a ripping quality in it. But detectivism makes it hard to see why my eventual description of my pain as ripping should carry any special authority. It was, after all, based on the same introspective observation that I had already done at  $t_1$ , without at that time judging my pain to be ripping. We can make the example clearer by stressing that my sensation of pain stays the same from  $t_1$  to  $t_2$ : I am not judging my pain to be ripping at  $t_2$  because it started as non-ripping and then suddenly turned into ripping. Someone could suggest that at  $t_1$  I did not attend to the pain as completely as I did at  $t_2$ ; the suggestion could be that upon hearing the word “rip,” I introspectively probed the pain again to see if the new word fits it, and found a novel ripping quality in it. But it is possible that I would sincerely deny that too, and

testify that my pain features in my experience exactly in the way as it did at  $t_1$ . The quality that made me describe it as ripping was in my awareness from the start; I merely came up with a better description of it.

I think it is fairly plausible that in this situation, where I explicitly admit that I do not derive my eventual pain-description from any distinct introspective act, few people would feel that the authority of my avowal diminishes from  $t_1$  to  $t_2$ . This suggests that detectivism is not adequate to explain the basis of first-person authority.

Maybe we should waive the detectivist idea and state that inward perceptions are not the source of the authority of my avowals. Instead, it could be suggested that first-person authority is only a matter of mastering a language. Adult persons who are competent language-users have learned a stock of everyday phenomenological vocabulary, and they are considered to be beyond criticism in their psychological self-ascriptions just by virtue of the fact that they generally use that vocabulary in a coherent and consistent manner, without regularly coming into conflicts with other competent language-users. Upholding the first-person authority might be seen as a mere pragmatic or social convention.

If we think this way, how unassailable a subject's descriptions of her conscious experiences are will be a function of her adeptness in using experience-vocabulary. The descriptions of a fully competent adult will be authoritative, the descriptions of a young child or a non-native speaker less so. However, what should we do in situations where two people, both perfectly competent in introspecting and describing conscious experiences and who we have independent reasons to believe to be undergoing a similar experience, nevertheless describe that experience in mutually inconsistent ways? Do we then have to assume that at least one of them makes an introspective error? Are we then entitled to waive the first-person authority of one or both of them? For Schwitzgebel (2011), cases like that form the basis of one group of arguments to the effect that people are not in general reliable judges of their own conscious experiences.

### Constitutivism

If Schwitzgebel is right, much of the first-person authority that we normally grant to competent adult people is based on false prejudice. However, there is an alternative view of self-knowledge that denies that describing our experience is essentially a matter of having an accurate perception of one's inner episodes, which is then translated into words. This view, called "constitutivism" by Finkelstein (2003), is also friendlier to first-person authority than detectivism ends up being. Its central idea is that our judgments concerning our inner episodes play a constitutive role in determining what those inner episodes are.

Constitutivism seems insightful especially concerning propositional attitudes like beliefs. When we self-ascribe a belief, it seems that we most typically do

that by *rationaly committing* ourselves to a belief, via judging that something is the case. Self-ascribing a belief seems to be the act of forming a belief or settling on a belief, rather than finding one via introspection. As the so-called transparency theories of self-knowledge have emphasized, self-ascriptions of attitudes need not involve any judgment turned inwards, so to speak; they are rather part and parcel with the judgments we make of the outside world.

So at least in some cases, my non-introspective judgments may *constitute* my mental states. Also in the case of descriptions of sensations, my authority may be thought to be “not like the authority of an eyewitness [. . . , but rather like] that of an Army colonel when he *declares* an area off limits” (Finkelstein 2003, p. 28; emphasis in the original). A slightly adapted example will illustrate the point:

### *Example 2*

Two people have an abdominal pain that they describe to a physician. It has been established that their pains are caused by a similar medical condition; they are of the same age, gender and build, the patterns of activation in their nervous systems are highly similar, and their pain-descriptions agree for the most part. In short, we have good independent grounds for believing that they are describing qualitatively similar experiences.

One person describes her pain as sharp and ripping. The other person disagrees, saying: “I don’t think it is ripping at all, not really sharp either. It’s more like crushing and suffocating.”

Maybe we always have to leave some room for the possibility that, despite all clues to the contrary, the subjective experiences of the two people are, after all, different. But even if we assume that the experiences are similar and the subjects are just giving mutually incompatible descriptions of the same pain, we can interpret this as a case of faultless disagreement.

We can suggest that what the subjects are doing is not that they observe by introspection features of their inner experiences accurately or inaccurately. Instead, they are making spontaneous applications of concepts, and in doing this they engage in *defining* what their experiences are like. They are flagging a certain description as the correct thing to say about their experience. First-person authority, according to this view, is a matter of being in the unique position of choosing how experience-vocabulary is to be applied to one’s subjective experience. What ultimately makes it the case that a subject’s pain is ripping is the fact that the subject *judges* it to be ripping. Even if there is another, incompatible description of a qualitatively identical experience — even if the description of the first subject is highly anomalous — there is no need to ascribe error to any party. The deviant description can be treated just as a different application of experience-vocabulary, an application that is within the subject’s rationality to make, and which may be psychologically interesting in itself. It does not force us to waive the first-person authority of any speaker involved. First-person

authority is the acknowledgement that subjects' statements about their experiences are (treated as) true in their own conversational context.

Constitutivism, in the case of describing my pain, would be friendly to first-person authority by holding that my sincere testimony is the primary court of appeal which determines what my pains are like. The fact that I judge my pain to be ripping plays a constitutive role in making it the case that my pain is (rightly characterized as) ripping. First-person authority exists, according to this view, because the primary way of establishing the character of someone's experiences is to refer to that person's sincere avowals about those experiences. For that reason, my judgment to the effect that I have a ripping pain is essential in making it the case that my pain is indeed ripping, as opposed to crushing or suffocating. Of course, there will be constraints on how I can describe the pain; I cannot normally characterize my pain as "dark green" or "prestigious," for example. But it can be argued that this would not be because those descriptions are erroneous in light of some independent standard, but because they violate some conversational maxims; I would normally know that those words are probably uninformative to others as pain-descriptions. Insofar as I want to communicate, I should not use unhelpful concepts, but otherwise I am free to describe my pain in whatever way seems to me most suitable. In determining what is true to say about my experiences, those avowals of mine will be the primary point of reference. First-person authority just reflects this state of affairs.

Is constitutivism preferable to detectivism? Two points of criticism are important. First, it cannot really be praised as an account of self-knowledge. Instead, it makes it hard to characterize my pain-descriptions and other avowals as instances of (self-)knowledge at all. Knowledge conceptually requires some kind of systematic avoidance of error. Roughly speaking, if something counts as an instance of knowledge, it should involve a judgment that succeeds in representing some state of affairs *correctly*, in virtue of some laudable systematic method. If constitutivism generally holds, and truths about persons' inner states are primarily determined by referring to their avowals, then there will be no such thing as the cognitive achievement of getting a psychological self-description *right*. It will be no more of a cognitive achievement than launching an arrow into a wall and drawing a bulls-eye around its head is an archery achievement.<sup>2</sup>

Second, constitutivism seems to make us responsible for mental facts about ourselves in a way that is not plausible across the board. Here it becomes evident why constitutivism fits better together with accounts of beliefs and other similar attitudes. When we consider the latter, constitutivism seems advantageous, because we generally want to be personally responsible for the contents of our

---

<sup>2</sup>I believe that something like this thought is behind those remarks of Wittgenstein that suggest a "non-cognitive thesis of avowals," as Hacker (1975) calls it.

beliefs and desires. But sensations are different in this regard. According to constitutivism, what ultimately makes it right to say that my pain is ripping instead of crushing is the fact that I judge it to be ripping instead of crushing. But in many cases, I will be unable to accept this account from my own viewpoint. It will at least usually, if not always, strike me as false to say that my pain is ripping because I judged it to be ripping. In a typical situation, I say my pain is ripping because my pain calls for exactly that word, and I will be inclined to insist that I really have no rational control over that matter. If I complain of a sharp pain, no one can seriously suggest to me: “Learn to *judge* it to be dull instead, and then it will not be sharp anymore!” Not all conscious experiences, as they appear to me in first person, leave room to intellectually decide the most appropriate verbal characterization for them. Some experiences do not let me rationally judge what I want to say of them; they will rather take control of me, and demand an expression. This uneasiness from the first-person viewpoint should justify looking for a better account to surpass both detectivism and constitutivism.

### Expressivism

Finkelstein (2003), Bar-On and Long (2001), Bar-On (2004), and Rodríguez (2012) have examined expressivism as a superior alternative for making sense of our relation to our own inner sphere. This view develops a point inherited from Wittgenstein (1953), saying that much of psychological talk in the first person is not descriptive in nature; it does not stem from an observation of an inner object. Instead of merely rejecting detectivism, however, Wittgenstein insisted on continuity or at least a possible connection between verbal avowals and primitive, “natural” expressions:

How do words refer to sensations? [. . .] The question is the same as: how does a human being learn the meaning of the names of sensations? — of the word “pain” for example. Here is one possibility: words are connected with the primitive, the natural, expressions of the sensation and used in their place. A child has hurt himself and he cries; and then adults talk to him and teach him exclamations and, later, sentences. They teach the child new pain-behavior. “So you are saying that the word ‘pain’ really means crying?” — On the contrary: the verbal expression of pain replaces crying and does not describe it. (Wittgenstein 1953, §244)

According to the possibility Wittgenstein points out, the avowals that we use to talk about our experiences work in the same way as pre-verbal grunts and cries. The point of the avowals is not to be parts of fact-stating discourse, but to give voice to wants and needs in social interaction. The avowals can also be drawn out of me against my will, like primitive expressions. This is a point in favor of expressivism against constitutivism, as the latter threatened to over-intellectualize the subjective sphere.

For the question of why my descriptions of my own experience carry a special authority, expressivism offers a deflationary answer. According to it, avowing is not a matter of describing one's pains or feelings at all. Avowals only superficially look like descriptions. Actually they are sophisticated and cultured expressive behavior: utterances that are in the business of reacting to my surroundings, and thereby doing other things, such as eliciting pity or asking for help. This was Gilbert Ryle's view in his *Concept of Mind*:

[M]any unstudied utterances embody explicit interest phrases, or what I have elsewhere been calling "avowals," like "I want," "I hope," "I intend," "I dislike," "I am depressed," "I wonder," "I guess," and "I feel hungry"; and their grammar makes it tempting to misconstrue all the sentences in which they occur as self-descriptions. But in its primary employment "I want..." is not used to convey information, but to make a request or demand. [. . .] Nor, in their primary employment, are "I hate..." or "I intend..." used for the purpose of telling the hearer facts about the speaker; or else we should not be surprised to hear them uttered in the cool, informative tones of voice in which we say "he hates . . ." and "they intend . . ." We expect them, on the contrary, to be spoken in a revolted and a resolute tone of voice respectively. (Ryle 1949, pp. 183–184)

However, even if Ryle's view of the primary employment of avowals is correct, he realizes that he cannot boldly generalize this point. The existence of a "primary" employment implies that there are one or more secondary employments. Surprising or not, sometimes "I hate . . ." and "I intend . . ." are uttered in a cool and measured manner, in order to give a self-description. The view that avowals are simply expressive and lack truth-values is rightly met with suspicion (Hacker 1975; see also Malcolm, 1954). Obviously, if this simple view is what expressivism amounts to, it will explain (apparent) first-person authority, but it will not be an account of self-knowledge. According to it, my verbal avowals are no more instances of self-knowledge than distinctive grunts and gestures are. On the other hand, those avowals cannot be meaningfully corrected by another person, but this is for the trivial reason that they have no factual content to disagree on.

Wittgenstein (1953, II, ix) plausibly acknowledged that avowals can play the role of both expressions and descriptions, or something in between. A non-naive version of expressivism holds that my speech about my own mental states is fundamentally continuous with my natural bodily expressions, but such speech still linguistically expresses or "manifests" facts about my thoughts and feelings so that it is capable of stating truths or falsehoods about me. Bar-On (2004) has developed such a version and labeled it "neo-expressivism." Sophistication is clearly necessary, because it is hard to deny that avowals are in some sense also in the business of stating facts about their speaker. Avowals have contents that can feature in logical inferences, they can be contradicted by other statements, and so on. It seems that expressivism has to face an objection that is parallel to the Frege–Geach problem for metaethical non-cognitivism (for a summary, see Sinclair, 2009): How can this way of talking be fundamentally expressive, when it evidently in many contexts functions like descriptive, fact-stating talk?

In what follows, I will present a development of expressivism to shed light on the nature of avowals, the first-person authority associated with them, and the limitations of that authority. I attempt to combine a number of what I take to be valid insights. First, I will endorse a view that I attribute to Wittgenstein: avowals can function as expressive utterances but also as descriptions, and there is no categorical line separating the two cases. Second, I agree with Rodríguez (2012) in holding that Bar-On's (2004) influential expressivist account has the undesirable feature of taking apart avowals as expressive *acts* and avowals as the linguistic (truth-evaluable) *products* of those acts. I suggest that the putting forward of a linguistic description of one's experience is a single expressive act, whose expressive quality and truth-value are assessed in an interdependent fashion. Third, I seek to integrate detectivism and constitutivism in the picture, by highlighting the kinds of cases where each works best.

### Primitive Avowals, Intellectual Self-Ascriptions, and Deliberations

For heuristic purposes, I will distinguish between three different types of psychological self-ascription. These are not meant as rigid categories. Instead, they represent the end and middle points of a scale on which avowals, and interpretations of avowals, can move. One extreme is a purely expressive, spontaneous avowal; another extreme is a detached, cool self-ascription done as if from a third-person perspective. Between these, there is a vast range of avowals that express the speaker's state of mind *by* asserting something about it. A good label for these latter cases is hard to come by; I will call them *deliberations*, owing the word, and some of my inspiration, to Moran (2001).

Other advocates of the expressivist view have made the point that (some) avowals have a special epistemic authority because of their peculiar expressiveness. They are taken to be immediate, non-judgment-involving airings of the subject's mental states. My aim is to qualify, and clarify, this point by suggesting that some avowals (deliberations) have a special epistemic authority when they are expressive in a certain spontaneous and unstudied way while *also* being honest attempts of a revisable self-description.

#### *Primitive Avowals*

First, I endorse Wittgenstein's point about verbal expressions of feelings being able to take over and extend the function of primitive, non-verbal expressions. Assuming that more articulate and considered expressions can build on simple primitive expressions, I will propose a way of seeing these as a procession on a single, continuous scale. Primitive, natural expressions like cries and smiles are devoid of cognitive content. They are not attempts to convey factual information. They may be expressions of attitudes, means of drawing attention, devices of eliciting reactions from others or otherwise communicative, but they are not

statements or descriptions of the subject's mentality. They can be called purely expressive acts. The simplest form of verbal avowals can be equated with them. Cases where "It hurts!" is used spontaneously and passionately to serve the same function as would be served by a scream, or a case where a spontaneous "I feel so good!" takes the same communicative role that could be taken by an exhilarated smile, can be called purely expressive avowals. These have a character of naturalness and spontaneity; they are drawn out of a person, rather than formulated and put forward by the subject in a controlled fashion. This is one end of my proposed spectrum.

### *Intellectual Self-ascriptions*

On the other end of the spectrum, there are self-ascriptions of mental states that are purely descriptive. Whereas purely expressive avowals are not descriptive to any extent, the self-ascriptions of mental states at the other end of the spectrum are not expressive to any extent. The latter are instances where the subject takes a detached, third-person perspective toward her own mentality, and produces a studied verdict from that perspective. She may or may not like the contents of that verdict; she may even want to disown it. I will call these intellectual self-ascriptions. They will include a case where I reluctantly admit, after a lengthy work to sort out my thoughts, that I am angry with my father because of his strictness as a disciplinarian, while at the same time admitting that I should not and do not want to be angry with him. In another case, I notice my slowing pace of work and carelessness and conclude that I must be tired and frustrated, although I do not feel like saying that I am either of those things; but my physical and behavioral condition force me to make that conclusion anyway. I know, after all, that lethargy and carelessness are objective criteria for a person's being tired.

At this latter end of the spectrum, it can be legitimately said that I come to know my own mental states by *detecting* them in myself, although that detection is not necessarily carried out by inward glances of introspection. In any case, in these instances I attribute a mental state to myself as a result of self-observation of some kind, and this observation has no special claim of authority over anyone else's word. My self-observation can be mistaken for the same mundane reasons as any observation can be mistaken; it will make perfect sense to ask me to do my self-observation more carefully or more attentively, in order to avoid error. It is possible that I mistake the symptoms of a medical condition for symptoms of tiredness, or that I misidentify as repressed anger something that further reflective work reveals to be some other complex feeling. In short, this is a class of cases where I am sufficiently alienated from my own mental state to treat that mental state as an external object of scrutiny. The account of detectivism, while not easily generalizable, fits well here. This kind of self-scrutiny was what

Ryle (1949), who rejected privileged introspective access as the basis of self-knowledge, eventually treated as the paradigm case of real self-knowledge.

### *Deliberations*

There is a purely expressive case of avowal; these I have called primitive avowals. There is also a purely descriptive case of avowal (according to my liberal use of the term); these I have called intellectual self-ascriptions. Now I will distinguish a third case, which is *the speech act whereby the subject puts forward an expressive linguistic utterance to serve as a self-description*. I believe that many, maybe most, avowals in typical human communication can be seen as instances of this type. They are characterized by a desire of the speaker to strike a balance between saying something that can be taken to be an objectively accurate description of her, and voicing her own impulses and wants, all in a single speech act. They are expressive utterances of the subject, but these expressive utterances acknowledge that they are attempts at manifesting a mental event that is an object of scrutiny also from the subject's own perspective; an event for which the giving of an adequate description is a cognitive challenge. They are instances where the subject assesses two things at once: first, what she wants to say about her experience; and second, how objectively plausible her statement is as a self-description. I call these avowals *deliberations*. One more modified example will serve to illustrate the point.

### *Example 3*

I have an abdominal pain that I need to describe to a physician. I am asked to assess my pain's intensity on a scale from 1 to 10. I have used the pain-scale before, and I consider the guidelines I associate with different numeric degrees of pain. I judge that my pain is of the level of 7. Then I am asked to think carefully:

"You describe your pain otherwise in the same way as in those earlier instances when you have judged it to be 5 or 6. You also don't show signs of greater distress over it. Are you sure that 7 is not too much?"

I answer: "Yes, I understand that, but I just feel that this week it is harder to bear. I'm not sure if it is the pain itself that intensifies or if I am just depressed, but 6 would be too small a number now. I'm saying 7."

Here, I am doing several things at once. First, I am giving a description of my pain. My utterance of "7" occurs in a context of giving a description; it is meant to inform the other about a certain feature of my conscious experience, to go down in my medical record as a true proposition about my condition. Second, I am using words (or rather, numbers) expressively: the point of my saying "7" is to let the other know how I feel about my pain, to voice my sentiment. Third, however, in this particular example I acknowledge that I have

some reservations about whether my avowal accurately describes a change in the pain itself or in the overall quality of my mental condition (“I’m not sure if it is the pain itself that intensifies or if I am just depressed”). Here I admit that my decision to say 7 instead of 6 might be borne out of my growing concern over my pain, my overall feeling bad physically and mentally, or something like that. In a way, I give the hearer some freedom to evaluate what conclusions to draw from my utterance.

Now, it seems to me plausible to agree with constitutivism to an extent. My description of my pain has a unique claim to being true. This is because my avowal has a special status in determining what is deemed right to say about my pain’s intensity. My honest avowal of my pain as 7 is a central criterion for it being the case that my pain indeed has the intensity of 7. I am the only one who can apply pain-vocabulary to myself in the first person, so my judgments about my pain are crucial in determining how pain-vocabulary is to be applied to me in particular cases. However, my avowals are not the only criterion for determining what my pain is like; there are bodily and behavioral criteria for different kinds of pains too (as the interlocutor in Example 3 notices).

In light of this, I suggest that my avowal is a complex communicative act: it is, in effect, *a request for others to accept my pain-avowal as a valid description of my pain*. It has a double nature. It is put forward as a description of my state of consciousness, but it is also a kind of an act of pleading: an expression of my want to make others treat my pain as a pain of the level 7. Most of the time, my avowal will be accepted as a valid and authoritative description without a scruple, insofar as people generally accept the first-person authority of subjects over their own mental states. But sometimes there will be room for scrutiny, as in Example 3.

In Example 3, I am saying that my pain is 7 in circumstances where, as far as any onlooker can see, I could as well say 5 or 6. So why am I saying 7? If this unusual question would be put to me, I could approach it in a number of ways.

(a) First, I could try to ground my judgment in some objectively available behavioral evidence. “Look, I may not show signs of greater distress over my pain just now, but there are some signs anyway: it distracts me more than before, it is harder to concentrate on anything else, I am constantly more stressed about it than before . . . . It must have intensified from 6 to 7.” Here, I am taking a more detached position toward my pain by allowing that it is a matter of evidence to decide whether my pain is 6 or 7. This is to move my avowal more in the direction of what I have called an intellectual self-ascription. I loosen my claim to first-person authority somewhat, by allowing that my judgment about my pain might be wrong according to some standards that can override my own statement.

(b) Second, I could (in principle) decide to be a hard-headed constitutivist. “I just feel like saying 7. It seems to me to be the correct application of the

pain-scale to what I am feeling right now. And I am automatically right in this, because it is me who gets to decide how pain-vocabulary is applied to my inner experiences. End of story.” I think it is evident that by these words, the subject would make her avowal sound in a certain way suspicious. It seems that her attitude toward her pain is not an attitude of a person who wants to communicate something about her pain to others. It is rather the attitude of a person who is merely interested, for one reason or another, to ensure that the hearers withhold further inquiry and accept her statement. Concerns about the honesty of the avowal would be raised, and there would be some hesitation about whether, or to what extent, her utterance can be taken seriously as an avowal. It would be sensible to protest that the subject does not get to *decide* whether his pain is 6 or 7 just like that.

(c) Finally, there is the option that seems to be natural and plausible: “I just feel that I have to say 7. I cannot help it. It just feels worse today.” What I acknowledge here is that my avowal shares the nature of a primitive expression: the number 7 is drawn out from me, somewhat in the way spontaneous grunts or smiles are drawn out from me, rather than rationally decided to be my chosen number for the pain.

Now, I suggest that the first-person authority of my self-ascription is at its strongest when it has a nature like that described in (c). When it is in this way akin to a spontaneous, primitive expression, then the subject has a special force behind her request that her pain-avowal is treated as a valid description of her pain. Her avowal will then represent her genuinely best effort to give a linguistic expression to an event of her consciousness that does not allow for just any arbitrary expression.

In other words, I am suggesting that the authority of an avowal as a self-description is dependent on whether the avowal is taken to share in the nature of a primitive avowal. But I am also arguing that it is necessary for an avowal to be plausible as a description from a detached perspective, if it is to work in its role as an avowal. Once more, I will illustrate by an example. Let us imagine that, in Example 3, I am struck with a sudden fear and anguish over my pain, and I start to feel my constant, familiar pain as so unbearable that I want help with it immediately at any cost. Then, when asked about the intensity of my pain, I will respond “10.” Now, it seems that another person would have a good reason to say to me: “Look, you cannot really say that. I know you feel bad, but 10 is the highest point of the scale, it is meant to represent a pain that is so unmanageable that you have never experienced anything worse than it. A person with a pain that has the intensity of 10 would be incapacitated, which you clearly are not.” In a way, my “10” would be a failed avowal; it could not be taken seriously as an avowal.

In the previous case, I am uttering “10” as a kind of a purely expressive call for help that does not even purport to be a measured attempt of self-description.

This kind of an avowal will be appropriate in some conversational contexts, but faulty in many others. In particular, it will be unhelpful for the physician, or at least it will put the physician in a position where she has to contemplate how to interpret my utterance. It will not be a fully functional avowal in its context.

I take these considerations to show the following. Insofar as my (deliberative) avowals are descriptions of my mental state, they are also requests for others to accept my description as valid. But the acceptability of my avowal as a valid self-description is largely dependent on whether my avowal is taken to be expressive in the right way (i.e., in the way of a primitive, unstudied expression). And my avowal, however honestly expressive, will not be fully taken seriously as an avowal unless it at least attempts to be a descriptive act (i.e., is constrained by my aspiration to inform others about what my pain is like, and not only by what I want to say about it). The descriptive and expressive aspects of an avowal are interdependent.

### **Conclusion: Avowals, Self-Knowledge, and the Nature of First-Person Authority**

I will now conclude by spelling out some consequences for the issues of self-knowledge and first-person authority that can be drawn from my discussion. First, it seems to me that a crucial part of what is commonly called “self-knowledge” is manifested in a person’s ability to reflect on her use of the different modes of avowals, and to some extent choose between them. Avowals are called for in many different communicative situations. Sometimes, when another person asks me “How do you feel?,” what is expected from me is just a spontaneous manifestation of my feeling of pain, affection, or anxiety. Then, it is an exercise of self-knowledge to be able to recognize and let out my spontaneous and unstudied reaction, suppressing any need to take a detached perspective and survey my state of mind as a part of my objective personal psychology. At other times, it will be necessary for me to study my psychology as if from a third-person perspective, in order to uncover biases or unconscious motivations, acknowledging that my own assessment of my psychology is nothing but an assessment by a fallible human being. Then, it will be necessary to contain my spontaneous and unstudied reactions, and to keep in mind the possibility that my first thoughts about my pains, affections or anxieties might not be the (whole) truth about them. (“I feel like saying that this pain is 7; but don’t I usually have a low pain threshold? Maybe most other people would call it 6, or even 5? And I admit that I am feeling depressed; maybe that is affecting all my judgments more than I realize.”) Understanding that my unstudied expressions and correct descriptions of my psychology (according to some standards that I myself can accept when speaking in third person) can come apart, and finding out how they can be expected to come apart in diverse situations, is a vital part of my

self-knowledge. In deliberations, I talk expressively, and in so doing I manifest my wants and needs to characterize my mental life in certain ways, but at the same time I am subjecting my avowals to interpersonal assessment by presenting them as *descriptions* of myself. Seeing how those expressively grounded descriptions manage with and against those descriptions of me that are given from the perspective of another person helps me to cultivate an important kind of self-knowledge. I am learning how my conception of myself plays together with other people's conception of me.

This characterizes self-knowledge in a sense in which it is a process. It is a sense of self-knowledge that is easily overlooked if the crucial expressive function of first-person psychological talk goes unnoticed. In deliberations, how I can plausibly describe myself constrains how I should feel appropriate to express myself, and at the same time how I need to express myself constrains how I describe myself. Competent use of deliberative avowals might be characterized as communication that is at the same time both self-studying and self-defining — a remarkable feat of human thought.

Second, pointing out the combination of expressiveness and descriptiveness in avowals produces a modest and commonsensical view of first-person authority and its limitations. There is little motivation to assume that individuals have magically accurate introspective powers, so that they would be uniquely authoritative judges of their own mental states in a detectivist manner. But what people do have is a subject's perspective to those mental states, and a desire to define and characterize those states from that perspective. Conscious attempts of persons to work out what their subjective experiences are like — what I have called deliberative avowals — have a special epistemic status insofar as they are properly expressive honest utterances while also being attempts of self-description. A description that I give of my own experience is authoritative when, and insofar, it is based on an expressive act that is ungrounded and natural in the same way as a primitive bodily reaction is. The subject is the only one who is in a position to give a description with this peculiar basis; therefore, naturally, an avowal of this kind carries special weight. When moving away from deliberations toward primitive avowals, or toward intellectual self-ascriptions, motivation to demand a special authority for the avowals wanes: in the case of primitive avowals, because they are not issued or interpreted as statements with factual content, and in the case of intellectual self-ascriptions, because they are not made from the special perspective of the subject-position.

First-person authority is, first and foremost, recognition that each person has a unique status as a generator of knowledge about her own mental reality. Properly expressive deliberative avowals have a special epistemic job to do. They are not infallible, not always even highly reliable, but they are acts of giving voice to a personal experience: they are the subject's applications of concepts to her personal experiences in a certain situation and at a certain time, and as

such they have a constitutive role. They serve as the starting point of inquiry into her experience, and enjoy a certain amount of resistance to corrections. The role of such avowals as (partly) self-defining acts also means that a subject can, in principle, decide to stick to her self-description even when it is anomalous from the perspective of an outside observer. If a subject is truly brought to see her self-description as erroneous, this must happen by eventually bringing her to revise her avowal in such a way that she can, after the revision, own it as her honest self-expression, not only as a third-person description of her forcibly given from outside. This seems an essential characteristic of an autonomous, self-standing subject. Consequently, first-person authority has an ethical dimension in addition to an epistemic one. Respecting it is to grant to other people an authoritative voice in telling what their experiences are like. Disregarding it is to say that it is in principle possible to overrule a subject's self-expressing voice by a third-person, more authoritative account of what her experiences are *really* like. It is doubtful whether those who are subjected to the latter treatment have a chance of seeing themselves as subjects in the full sense.

### References

- Bar-On, D. (2004). *Speaking my mind: Expression and self-knowledge*. Oxford: Oxford University Press.
- Bar-On, D., and Long, D.C. (2001). Avowals and first-person privilege. *Philosophy and Phenomenological Research*, 62, 311–335.
- Carruthers, P. (2011). *The opacity of mind: An integrative theory of self-knowledge*. Oxford: Oxford University Press.
- Finkelstein, D.H. (2003). *Expression and the inner*. Cambridge, Massachusetts: Harvard University Press.
- Gertler, B. (2011). *Self-knowledge*. London: Routledge.
- Hacker, P.M.S. (1975). *Insight and illusion: Wittgenstein on philosophy and the metaphysics of experience*. London: Oxford University Press.
- Hurlburt, R.T., and Schwitzgebel, E. (2007). *Describing inner experience? Proponent meets skeptic*. Cambridge, Massachusetts: MIT Press.
- Malcolm, N. (1954). Wittgenstein's Philosophical Investigations. *The Philosophical Review*, 63, 530–559.
- Moran, R. (2001). *Authority and estrangement: An essay on self-knowledge*. Princeton, New Jersey: Princeton University Press.
- Rodríguez, Á.G. (2012). How to be an expressivist about avowals today. *Nordic Wittgenstein Review*, 1, 81–101.
- Ryle, G. (1949). *The concept of mind*. New York: Barnes & Noble.
- Schwitzgebel, E. (2011). *Perplexities of consciousness*. Cambridge, Massachusetts: MIT Press.
- Sinclair, N. (2009). Recent work in expressivism. *Analysis*, 69, 136–147.
- Wittgenstein, L. (1953). *Philosophical investigations* (third edition). [G.E.M. Anscombe, Trans.]. Oxford: Blackwell.

**IV**

**WITTGENSTEINIAN CRITERIA AND PERCEIVING OTHER  
MINDS**

by

Vaaja, Tero, 2015

To be submitted for publication.

# Wittgensteinian Criteria and Perceiving Other Minds

Tero Vaaja

*Abstract.* In his later work, Wittgenstein employs a notion of “criteria” in the course of discussions of knowledge and application of concepts. This subsequently much discussed notion has been widely taken to offer an original solution of the problem of other minds, by suggesting that some types of behavior are, as a matter of linguistic convention, criteria for mental states. I set out to examine the contribution the notion of criteria makes to the problematics of other minds in a Wittgensteinian context, in particular by investigating how it plays together with his overall critique of the private “inner” and the behavioral “outer”. I argue that in order to be consistent with this overall picture, the criteria should be taken as able to be constitutive of the mental states they are criteria of. The crucial challenge will be to show how such a view can avoid falling into implausible behaviorism. By help of some ideas from John McDowell, and elaboration of the concept of expression, I argue that instances of perceptible bodily behavior can be, in favorable circumstances, criteria for mental states in a constitutive sense. This is when they are expressions in a sense in which there is there is no gap between the expression and the expressed.

## 1. Background

In his later work, Wittgenstein employs a notion of “criteria” in the course of discussions of knowledge and application of concepts. It is commonly stated that the notion offers an original solution, or a “dissolution”, of the problem of other minds (see e.g. Hyslop 2014, Baggini & Fosl 2010, Thornton 2014, and the survey articles of Lycan 1971 and Addis 1995). The purpose of this paper is to clarify the contribution Wittgenstein’s remarks on criteria make on the issue of other minds. This is done by relating them with his critique of the “inner and outer”, the theme of perceptibility of other minds, and the concept of expression.

Wittgenstein initially introduced the term "criterion" in his *Blue Book* (Wittgenstein 1958), where he contrasted criteria with symptoms. These passages are preceded by a discussion about the ways in which we know about an "inner" experience of another person (toothache, to be precise).

"Let us introduce two antithetical terms in order to avoid certain elementary confusions: To the question 'How do you know that so-and-so is the case?', we sometimes answer by giving 'criteria' and sometimes by giving 'symptoms'. If medical science calls angina inflammation caused by a particular bacillus, and we ask in a particular case 'why do you say this man has got angina?' then the answer 'I have found the bacillus so-and-so in his blood' gives us the criterion, or what we may call the defining criterion of angina. If on the other hand the answer was, 'His throat is inflamed', this might give us a symptom of angina. I call 'symptom' a phenomenon of which experience has taught us that it coincided, in some way or other, with the phenomenon which is our defining criterion. Then to say 'A man has angina if this bacillus is found in him' is a tautology or it is a loose way of stating the definition of 'angina'. But to say, 'A man has angina whenever he has an inflamed throat' is to make a hypothesis.

In practice, if you were asked which phenomenon is the defining criterion and which is a symptom, you would in most cases be unable to answer this question except by making an arbitrary decision *ad hoc*. It may be practical to define a word by taking one phenomenon as the defining criterion, but we shall easily be persuaded to define the word by means of what, according to our first use, was a symptom. Doctors will use names of diseases without ever deciding which phenomena are to be taken as criteria and which as symptoms; and this need not be a deplorable lack of clarity. For remember that in general we don't use language according to strict rules – it hasn't been taught us by means of strict rules, either." (Wittgenstein 1958, 24-25)

Symptoms of the fact or condition X are signs of the presence of X; they are non-conclusive pieces of evidence for the obtaining of X. Criteria, by contrast, are aspects of the situation that make the situation count as an instance of X; they spell out what it is that is correctly called an X. But, as Wittgenstein goes on to point out, the difference between the two can be vague.

In another much briefer but equally famous remark, Wittgenstein says that "an 'inner process' stands in need of outward criteria" (Wittgenstein 1953/2001, §580). Thus, Wittgenstein seems to be claiming that there are outward bodily events that are not merely signs of mentality in others, but are connected with mental states by the meanings of mental state terms – by a convention of language, by "grammar". This makes those outward bodily events criteria for their respective mental states. It is commonly assumed that a novel and important way of dealing with skepticism about other minds is provided by this view. A textbook example is given by Baggini & Fosl (2010, 94):

"[The idea of criteria], Wittgenstein believed, provided a way out of some old philosophical difficulties: How can we know that other people have minds? And how can I avoid solipsism – the idea that only I exist? These

problems dissolve (rather than are solved) because the criteria for the correct use of words like 'pain' and 'minds' are behavioural and social – even though that does not mean that pain and minds *are* only behaviours.”

The critique of a dichotomy between a private, subjective "inner" and a behavioristic "outer" is a major theme of Wittgenstein's later writings about mind and psychology. His main insight is that there is something deeply wrong with the common (roughly "Cartesian") idea of the human mind as reified, or construed as a metaphysically private "place". That place is supposedly accessible to the subject herself, but ultimately inaccessible to anyone else, except via sounds emitted and movements made by human bodies, those sounds and movements allowing us to infer things, more or less confidently, about the hidden private realm of mind. Thinking about the mind like this is a prime example of what Wittgenstein condemned as a confusing philosophical "false picture" (Wittgenstein 1953/2001, §115). In contrast with the idea of a human being as the combination of private inner experience and outer physical behavior, Wittgenstein famously remarked that "the human body is the best picture of the human soul" (Wittgenstein 1953/2001, II, iv, 152).

## **2. What Are Criteria?**

In what follows, it is necessary to bypass many interpretative issues that have been discussed extensively in commentary literature since the publication of *Philosophical Investigations*. (For an overview of the issues, see Addis 1995.) One question is whether a systematic theory of criteria can be acquired from Wittgenstein's remarks. Assuming that Wittgenstein was not in general aiming at formulating theories, it seems sensible to also assume that he employs "criterion" (or "Kriterium", in German) as a common term with an ordinary established meaning, not as a specifically coined technical term with a special meaning. He thought that the workings of language could be elucidated by that concept, apparently in particular in the context of other minds. My purposes in this paper do not require me to take a stand whether or not a detailed theory of criteria based on Wittgenstein's writings can be formulated. I will focus on a limited number of points: in particular, the relation between criteria and evidence, and an interpretation of criteria for X as constitutive of X.

Wittgenstein made remarks of the form "Y is a criterion for X" in a variety of contexts. This raises another question, namely, whether the entities standing in criterial relations should be thought to be states of affairs, propositions, or something else. Here too, my account does not depend on giving a definite answer. I employ a broad construal of criteria as ways of telling whether something counts

as an X, following Glock (1996, 94) in assuming that “[t]he basic point [of Wittgenstein’s talk about criteria] is that certain phenomena or facts license the application of certain words”.

Calling criteria “ways of telling” is a common turn of phrase in the literature (see e.g. Glock 1996, de Gaynesford 2002, Glendinning 1998). Signs are ways of telling too; observing someone’s inflamed throat is a way of telling that the person has angina. But if the presence of a certain bacillus in the bloodstream is the criterion of angina, then observing that bacillus in the person’s bloodstream is in a distinct sense a (or *the*) way of telling that the person has angina. The criterion is connected to angina by virtue of what “angina” means. Malcolm (1954) used strong terms as he interpreted Wittgenstein’s criteria in general as decisive grounds for giving a verdict. Criteria “settle the question” of whether certain words apply: “The satisfaction of the criterion of *y* establishes the existence of *y* beyond question [...]. [I]f the criterion of [a man’s] being in pain is satisfied then he *must* be in pain” (Malcolm 1954, 543-544; emphasis in the original).

The phrase “ways of telling” also highlights that criteria should not be assimilated to truth-conditions, although in Wittgenstein’s initial example it could be said that the criteria for someone’s having angina are the same as truth-conditions for someone’s having angina. A salient difference is that criteria imply pragmatic usability. Truth-conditions can sometimes be transcendent to human capacities of knowledge, but criteria cannot. If something is a criterion, then it must be an actually usable way of giving a verdict (cf. de Gaynesford 2002).

Albritton (1959) gave an influential early interpretation of Wittgenstein’s use of the notion of criteria. Albritton interpreted criteria for X as “logically good evidence for” X; a type of evidence whose status as good evidence for X is a matter of linguistic convention. This sets criteria apart from signs or symptoms for X, whose status as evidence for X is a matter of a contingent empirical correlation. According to this interpretation, both signs or symptoms and criteria are evidence, but criteria are a special kind of evidence, their quality as evidence guaranteed a priori by “grammar”. Taking criteria to be cases of evidence involves another important interpretative issue. In cases like Wittgenstein’s initial example in the *Blue Book*, it is odd to construe criteria as “evidence”. There Wittgenstein talks about the bacillus in the patient’s bloodstream as “the defining criterion” of angina; the defining criterion spells out what is meant by ascribing angina to the patient. Observing a symptom, like the patient’s inflamed throat, is to gain a piece of evidence for the claim that the patient has angina, but observing the bacillus in her bloodstream is not equally obviously to gain a piece of evidence that she has angina. Rather, it is to perceive *that she has angina*. Because in later writings Wittgenstein did not generally talk about single defining criteria any more, and because in

the context of other minds it would apparently be crude behaviorism to take some bodily behaviors to be defining criteria of pains or other mental states, it has been commonly assumed that Wittgenstein's use of the term shifts significantly. Bodily criteria for the mental states of others should be thought as a special, non-inductively grounded type of evidence for them, the thought goes, rather than spelling out what mental states are; and the initial passage from the *Blue Book* should not be taken to be representative of his use of the term as a whole. In this paper, I will motivate a different reading, discouraging the construal of the criterial relation as evidential.<sup>1</sup>

Below, I attempt to dissect the concept of criteria more comprehensively, and sum up the central characteristics of criteria. This is meant to be a sum-up of both how Wittgenstein used the term in his remarks, and how the concept has been interpreted in subsequent commentaries. I distinguish four central characteristics in a way which I hope to be relatively uncontroversial. A range of other and more fine-grained analyses are available in the literature (again, see the survey articles of Lycan 1971 and Addis 1995 for an overview), but the present one is adequate for my purposes: I will point out an apparent inherent tension in the notion of criteria, and set the issue against the background of Wittgenstein's criticism of the inner and outer.

Let us start once more with the *Blue Book* passage about the criterion of angina. The first salient thing is the distinction between criteria and accidental signs or symptoms: the former are connected with their respective phenomena in virtue of a conceptual link. So, first,

a) *Criteria stand in an intrinsic ("grammatical") relation to what they are criteria of.*

This attempts to paraphrase in a neutral fashion the essential feature of criteria which makes them special and distinguishes them from symptoms. The grammatical relation can be put in a variety of ways. The status of the criteria for X as criteria is a matter of the meaning of the term X; criteria are non-inductive, a priori good evidence for X (as for Albritton); criteria for X play a role in learning or forming the concept of X.

---

<sup>1</sup> I have written a conference paper (Vaaja 2012) which contains the motivating thoughts of this paper in an earlier form. In that paper, I operate under the idea that criteria are grammatically grounded evidence. Considering that we see criterial behavior as expression of mentality, I earlier contended that we are able to see mental states of others in a certain sense "directly" by seeing them under the expressive aspect, but still there is a distinction between the expression and the expressed, leaving a sort of a gap between us and others. In contrast, I now think that an interpretation of criteria for X as constitutive of X, instead of special evidence for X, is worth pursuing, and that "expression" can be conceived in a way that does not include a necessary distinction between the expression and the expressed.

Considering that the most interesting use of Wittgensteinian criteria is their assumed ability to enable us to know about the "inner" experiences, such as pains, of others, it can be seen why this feature may seem problematic. Criteria for toothache in another person – ways of telling about the other's toothache that we hold as conclusive – will presumably be things like holding one's cheek, moaning, carefully probing one's tooth, saying "My tooth hurts", and so on. The problem is that while "angina" can very well be defined as having a certain bacillus in one's bloodstream, it seems implausible to say that "toothache" includes as a part of its meaning the behavioral pattern of holding one's cheek, moaning, probing one's tooth, and saying "My tooth hurts", or any extension of that pattern with similar behavioral items, let alone is definable in terms of them. That route seems to lead to behaviorism, from which Wittgenstein was careful to distance himself from (Wittgenstein 1953/2001, §§304-308; see also e.g. Luckhardt 1983).

The *Blue Book* passage makes another equally important point about criteria as well; namely,

*b) Criteria are **context-dependent**; what is a criterion for X in some circumstances may not be a criterion for X in other circumstances; and what is not a criterion for X in some circumstances may be a criterion for X in other circumstances.*

Wittgenstein remarks that there is "fluctuation in grammar between criteria and symptoms" (and additionally that this fluctuation "makes it look as if there were nothing at all but symptoms") (Wittgenstein 1953/2001, §354). In the case of judgments about other minds, what counts as criterial pain-behavior in typical circumstances will not count as such in all circumstances, such as when the person displaying the behavior is a known pretender, or when the behavior is part of a stage act.

Furthermore, and going beyond the *Blue Book* passage, it is clear that

*c) Criteria can be **multiple**.*

The *Blue Book* passage is about a situation where there is a single defining criterion of angina. But again in the case of other minds, it is quite evident that there is an arbitrarily large, open-ended set of criteria that provide us with ways of telling that another person has toothache; different kinds of variations of pain-behavior for different persons, different circumstances and different kinds of toothaches. It is also evident that no part of the open-ended, dynamic set can be taken apart in a non-arbitrary fashion and declared as *the* criteria for toothache. Wittgenstein, however, implies that

the variety and dynamicity of ways of telling about pains should not trick us into thinking that in any given situation, all of them were merely *symptoms* of pain, and none were criteria (cf. Wittgenstein 1953/2001, §354).

Finally, commentators of Wittgenstein have assumed that

*d) Criteria can be (and, in the case of other minds, are) defeasible.*

This disclaimer seems necessary to avoid implausibility when talking about bodily criteria for the "inner" experiences of others. Even if all the imaginable criteria for another person's being in pain are satisfied – groaning, wincing, grimacing following a bodily damage, verbal testimony of pain, or whatever have you – it remains in principle possible that the person is not in pain. The felt pain is after all, presumably, something distinct from the bodily events that constitute its criteria. Now, (a) and (d) seem to be in tension. According to (a), there is an intrinsic connection between pain and pain-behavior, and this is the crucial feature that sets criteria apart from symptoms. But still, according to (d), pain-behavior by itself does not make it the case that the other is in pain. In any case, it may seem that the only alternative to (d) would be unacceptable behaviorism, claiming that "inner" states just are to be defined in terms of patterns of observable behavior.

Albritton tries to resolve the tension by interpreting the intrinsic connection in the following way. The behavioral criteria of toothache are those types of behavior that are, as a matter of necessary truth, automatically justifying evidence that the other has toothache:

"That a man behaves in a certain manner, under certain circumstances, cannot entail that he has a toothache. But it can entail something else, which there is no short way of stating exactly, so far as I can find. *Roughly*, then: it can entail that anyone who is aware that the man is behaving in this manner, under these circumstances, is *justified in saying* that the man has a toothache, in the absence of any special reason to say something more guarded [...]. Even more roughly: That a man behaves in a certain manner, under certain circumstances, can entail that he *almost certainly* has a toothache." (Albritton 1959, 856; all emphases in the original.)

But this means that we have at our disposal some way of making sense of the distinction between the criterial toothache-behavior, and the thing that such a behavior "almost certainly" indicates, that is, what such a behavior is near-infallible evidence *for*: the toothache as such. If such a distinction is intelligible, then there is a way to conceive toothache as a bare, private sensation, divorced from its bodily criteria. And admitting this threatens to compromise (a) enough to make it insignificant.

What is the conceptual connection between toothache and its bodily criteria, if toothaches can be conceived, and their presence or absence conjectured, in abstraction of those bodily criteria?

A possible modest answer is that behavioral criteria of pain introduce the term “pain” into language by enabling us to recognize those situations that “almost certainly” involve pain; they specify what an appearance of pain looks like. What the behavioral criteria of toothache do is not to make it the case that a person displaying those criteria has toothache, but to make it the case that what he displays is the appearance of a *toothache*. This is the point Stanley Cavell forcefully makes, with the crucial emphases, when he writes that criteria do not tell us of a thing’s existence,

“...but of something like its identity, not of its *being* so, but of its being *so*. Criteria do not determine the certainty of statements, but the application of the concepts employed in statements.” (Cavell 1979, 45)

Surely, it is part of the explanation of the concept of “pain” to specify how we distinguish between situations that involve at least appearances of pain (situations that involve either real pain or at least apparent, play-acted or feigned pains) and situations that do not involve appearances of pain. But the account presented in previous paragraphs implies that we are in touch with pains of others only in terms of the bodily evidence for them, even if some of that evidence is “criterial”. The perceptible bodily criteria of pain only constitute an appearance of pain, veridical or otherwise; whether the pain itself is present is a further question.

I suggest that there is one major reason to be uncomfortable with this account of Wittgensteinian criteria in the context of other minds. The reason is that it does not fit well together with Wittgenstein’s rejection of the metaphysical privacy of the subjective sphere, and his critique of the distinction of the private “inner” and the perceptible “outer” as a dubious philosophical picture. He thought that, in at least some cases, our cognitive relation to others is such that we do not see on their bodies the outward evidence or indication of “inner” mental phenomena, but we see (in some acceptable sense of the word “see”; cf. Wittgenstein 1982, §§735-785) the mental phenomena themselves:

“Consciousness in another’s face. Look into someone else’s face, and see the consciousness in it, and a particular *shade* of consciousness. You see on it, in it, joy, indifference, interest, excitement, torpor and so on. The light in other people’s faces.

Do you look into *yourself* in order to recognize the fury in his face? It is there as clearly as in your own breast.” (Wittgenstein 1981, §220)

“We do not see the human eye as a receiver, it appears not to let anything in, but to send something out. The ear receives; the eye looks. (It casts glances, it flashes, radiates, gleams.) One can terrify with one’s eyes, not with one’s ear or nose. When you see the eye you see something going out from it. You see the look in the eye.” [...] ‘If you only shake free from your physiological prejudices, you will find nothing queer about the fact that the glance of the eye can be seen too.’ For I also say that I see the look that you cast at someone else. And if someone wanted to correct me and say that I don’t really *see* it, I should take that for pure stupidity.” (Wittgenstein 1981, §§222-223)

“‘We *see* emotion.’ -- As opposed to what? -- We do not see facial contortions and make inferences from them (like a doctor framing a diagnosis) to joy, grief, boredom. We describe a face immediately as sad, radiant, bored, even when we are unable to give any other description of the features. -- Grief, one would like to say, is personified in the face.” (Wittgenstein 1981, §225)

It is often pointed out (See Bohl & Gangopadhyay 2013, Gangopadhyay & Miyahara 2014) that the phenomenology of social cognition is straightforward: upon encountering an angry face, we immediately see the face as angry, without making a quick theoretical assumption about the “inner state” of anger that more or less probably causes these facial contortions. Obviously, frowns, flaring nostrils and a reddening face are not the same thing as anger; frowns, flaring nostrils and reddening faces can occur in contexts other than anger, and there can be concealed anger with no observable bodily effects. But frowns, flaring nostrils and a reddening face are the kind of things that could be cited as examples of Wittgensteinian criteria for anger – ways of telling that the person is angry that are more than inductive evidence. So, it seems right to state Wittgenstein's position like this: In at least some cases, although the criteria for, say, anger that are visible on another person’s face at a moment are not a definition of what anger is, we are nevertheless able to *perceive the other person’s anger by perceiving the criteria*. In those cases, there is no conceivable inner-outer gulf to cross: what we are prone to describe as an “inner” state actually lies open to perception.

This should motivate one to look for a plausible way of interpreting Wittgenstein’s notion of criteria more austerely, and in a way that does not assume a sharp difference between Wittgenstein’s initial use of the term in the *Blue Book*, where the relation between angina and the defining criterion of angina seems clearly constitutive (Albritton (1959, 855) judges that this “distorts” Wittgenstein’s subsequent use of the term), and his overall conception of criteria. The suggestion will be that bodily criteria of the mental states of others can be, in some sense, constitutive of the mental states they are criteria of. The crucial challenge will be how to make sense of such an option without lapsing into (and saddling Wittgenstein with) unacceptable behaviorism.

### 3. Criteria as Constitutive of What They Are Criteria Of

The view of the criterial relation as constitutive rather than evidential finds support from ordinary uses of the term. The following are conceptual remarks that at least purport to be points about ordinary language. They are not self-evident (such remarks never are, because “ordinary language” is not monolithic), but they seem to me plausible enough to note. When some phenomena are cited as criteria for deeming a case a case of X, the function of those criteria is to spell out what X is taken to amount to, at least in this particular instance of using the concept X. They are a clarification, typically for some practical purpose of making a decision, of what X is treated as, so to speak.

”Does this case satisfy the criteria of X?” and ”Is this an X?” are not merely two related questions, but for pragmatic purposes they are *one* question. The answer to the latter question is given by settling the former question. And reversely: When we encounter a case where our criteria for X are met, but we are still reluctant to judge the case at hand as a case of X, it seems insufficient to simply admit that the criteria were defeasible, in the same sense that we can admit that evidence for X was misleading, still justifiably holding that it *was* in any case *evidence* for X. Instead, these kinds of situations require also calling into question the status of the employed criteria as criteria.

Here is an example to illustrate the matter:

Let us say that we need to know whether a certain patient has mental capacity for making decisions. There is a set of criteria for determining mental capacity: the patient should be able to understand the decision to be made and information about the decision and its consequences; the patient should be able to retain that information long enough to make the decision; the patient should be able to use the information in weighing the pros and cons of the decision; and the patient should be able to communicate the decision. Then it is assessed to what extent the patient meets these criteria. Let us suppose that the patient is judged to meet the criteria to a satisfactory degree. The patient is thereby judged to have mental capacity. No inferential move takes place between the judgment about the fulfillment of the criteria and the judgment of mental capacity. The judgment about the fulfillment of the criteria *is* the judgment about mental capacity.

Now, suppose that, for whatever independent reasons, it is discovered that the patient did not have mental capacity after all, although the patient continues to meet the usual criteria for mental capacity. Now, there are two possible construals of the situation:

- i) The patient meets the criteria for mental capacity, but does not in fact have mental capacity

ii) It seemed as if the patient met the criteria for mental capacity, but in fact he or she does not

If the employed criteria are well established (socially, legally, and so on), the construal (i) is surely possible, in which case it means "The patient meets the *typically used* criteria for mental capacity, but does not in fact have mental capacity". But without that clarification, (i) seems internally incoherent – because judging someone to meet those criteria normally is to *thereby* judge that he or she has mental capacity. The construal (ii) seems preferable, because it involves the following thought: In a "deceptive" case like this, it needs to be concluded that *the criteria applied were not good for their purpose* – they were not good for settling the case whether or not the patient had mental capacity. What was found out was that *they could not be used as criteria for mental capacity in this particular case, in these circumstances*.

In varying circumstances, criteria can gain or lose their status as criteria. Noticing this is to notice the second of the two main points about criteria that Wittgenstein makes in his initial *Blue Book* passage: the **(b) context-dependence** of criteria, the statement that there is "fluctuation" between criteria and symptoms. Whatever features of the situation motivate us to judge that the patient should not be counted as having mental capacity, they will also motivate us to judge that those features should be included in the set of criteria of mental capacity in this particular kind of case (and some other maybe dropped, although in the example above I assumed this is not the case, to avoid unnecessary complication).

Criteria are standards that need to be operative, implicitly or explicitly, in order to assess the relevance of any evidence. This seems to be the essential purpose of introducing them, or making them explicit, in a discourse. Criteria are instructions on what is relevant to settle a matter, instructions on what has to be the case in order for X to be the case. Thereby they clarify what kind of evidence is evidence for X. Evidence for the patient's mental capacity would be things that speak for, e.g., that he can plausibly weigh the pros and cons of a proposal (his results in a psychological test, maybe), but the subsequent statement that he can weigh the pros and cons of a proposal is not itself giving a piece of evidence for his mental capacity. It is giving a part of the criteria for mental capacity, which means that it spells out, partly, what evidence for mental capacity is evidence *for*. The abilities given in the criteria are not factors that are found to be reliably associated with some independent property of the patient called "mental capacity". They are the phenomenon itself. If these remarks about the established use of the term "criterion" are plausible, and given the assumption that Wittgenstein did not intend to use the term in a way radically different from its

established use<sup>2</sup>, then there is motivation to interpret criteria as constitutive of what they are criteria of.

This constitutive interpretation of criteria, or at least something closely related to it, can be found in McDowell (1983). Arguing against the interpretation of Wittgensteinian criteria as defeasible evidence, McDowell makes the point that Wittgenstein has been illegitimately assumed to mean that if Y is a criterion for X in certain circumstances, then all Y's, as a type, are criteria for X in all circumstances. If, as in the case of other minds, there is always the possibility of Y occurring without X occurring, then this is taken to mean that Y's, as a type, are at best defeasible evidence for X. A preferable reading holds that it is the status of Y as a criterion that varies across different situations. Frowning and cursing are criteria of anger in typical circumstances, and even though we know that there are other circumstances (pretense being the simplest example) where a non-angry person frowns and curses, this does not mean that these are cases where criteria of anger are satisfied while anger is not present. This thought, as McDowell points out, depends on the assumption that what is a criterion in some circumstances is a criterion *universally*, across circumstances. But properly understood, our knowledge that there are circumstances where a non-angry person frowns and curses only means that whether frowning and cursing *are* criteria of anger or not, on a given occasion, depends on the circumstances. (McDowell 1983, 377)

According to McDowell, the satisfaction of the criteria for X should not be thought as something that "falls short of the fact" (McDowell 1983, 386-387) that X is the case, in a way defeasible criteria apparently necessarily do. Deceptive cases, such as a case where someone displays all the typical bodily behavior that suggests her being in a certain mental state, but in fact is not in that mental state, should not be construed as cases where criteria for X are satisfied, but X still fails to obtain. Rather, they should be construed as cases where it only *appears as if* the criteria for X were satisfied. This is an application of McDowell's general "disjunctive" conception of experience: an instance of perception can be *either* a case of encountering a misleading appearance *or* a case where worldly facts make themselves directly manifest to perception, without the mediation of "appearances" (McDowell 1983; Witherspoon 2012). It is an unfounded prejudice caused by a concern with skepticism, says McDowell, to assume that we have cognitive access to a domain, such as the mental states of other, only in terms of some "neutral" information that is common to a good case (that is, a case where I take another person to be in a certain mental state, and she actually

---

<sup>2</sup> It is evident that commentators have interpreted Wittgenstein's "criterion" as a technical term, and e.g. Baggini & Fosl (2010) take it for granted that the Wittgensteinian meaning of "criteria" is different from its meaning in everyday English. But even if the term has been taken as a technical term in the literature, this does not show that it was so for Wittgenstein.

is in that mental state), and a bad case (where the appearance is deceptive). Such neutral information could be something like the bare spatio-temporal information about, say, another person's facial contortions that impinges on our senses, and requires us to make a hypothesis or a conjecture about an underlying mental state.

This offers a new way of understanding the **(d) defeasibility** of criteria, stemming from their **(b) context-dependence** and **(c) multiplicity**. Concepts governed by context-dependent and multiple criteria are dynamic, and there is always the possibility of encountering a deceptive case, where the usual criteria for applying the concept are deemed inadequate. This means that any general kind of criterion can fail in specific circumstances. As a simplified example: Frowns, in general, often serve as criteria for anger. However, there are situations in which a frowning person is not angry. This commonsensical admission spells out the way in which criteria can be said to be "defeasible": There is no guarantee that anything that is a criterion for a mental state in usual circumstances will continue to be so in all circumstances. But criteria are *not* defeasible in another sense, namely, in the sense in which someone could be taken to satisfy the criteria of, say, anger, and at the same time be taken to not be angry. A judgment that *frowning is a criterion for A's anger in this situation* contradicts the judgment that *A is frowning but A is still not really angry*.

The point can also be made in the following way. We can talk about "frowns" in two senses. In the first sense, a "frown" is a facial contortion of a certain kind; a certain kind of look in the face and eyes. In the second sense, something is a frown only insofar as it is a manifestation of anger; something's being a frown is a matter of whether the person displaying it is angry. We employ the first sense when we, say, describe the appearance of gorillas by saying that they have a permanent frown on their face. We employ the second sense when we say that my friend, who put on an angry face but did not mean it seriously, was not really frowning at me. Now, the question we implicitly ask ourselves when encountering an angry appearance in another person is: Do the circumstances allow us to conclude that this person who frowns at us in the first sense also frowns at us in the second sense? If the answer is yes, then frowning is, at this situation and in these circumstances, a criterion of anger for us. If the answer is no, then it is not. Obviously, we often make mistakes. We often take something to be a criterion of anger in circumstances where it was not one. However, in *good* cases we can perceive the other person's anger, and not only external evidence of it, by perceiving the criteria of anger. In the good case the obtaining of the criteria for X is something that, to use McDowell's wording, does not fall short of the fact of X.

The sense in which McDowell means that encountering criteria for X “does not fall short of the fact” that X is that the satisfaction of the criteria is not compatible with X not obtaining (McDowell 1983, 387). Encountering a frown in the second sense is not compatible with there being no emotional fact about the person manifested in the frown; something is a frown in this sense only insofar as the situation involves an occurrence of an emotion. This does not make it unambiguous what the specific relation between the bodily criteria and the mental state is, though. McDowell suggests in passing that in some cases, most plausibly in the case of facial expressions of emotional states, what is accessible in experience (the criteria) may be “the fact [of the emotional state of the other] itself [...] directly presented to view” (McDowell 1983, 387; see 387n34). But in other cases, it may be appropriate to apply the idea less straightforwardly:

“[I]n at least some cases of knowledge that someone else is in an ‘inner’ state [...] we might think of what is directly available to experience in some such terms as ‘his giving expression to his being in that “inner” state’; this is something that, while not itself actually being the ‘inner’ state of affairs in question, nevertheless does not fall short of it in the sense I explained.” (McDowell 1983, 387)

This distinction between the “inner” state and the expression of the “inner” state will concern me in the next section, where I will attempt to push further the possibility of construing the bodily criteria of mental phenomena as constitutive of the mental phenomena themselves, in the interest of doing justice to Wittgenstein’s critique of the inner and the outer.

#### **4. Criteria and Expression**

What I am now examining is the following reading of Wittgensteinian criteria: Criteria are context-dependent and typically multiple ways of telling that something is the case, defeasible (only) in the special sense that their status as criteria is dependent on the circumstances; and in good cases, they are constitutive of the phenomena they are criteria of. The intricacy of McDowell’s discussion should testify that the last part of this characterization is not easy to make appealing in the context of other minds. The salient question is whether any bodily events *can* be criteria for mental states in this sense. Can it in any way be held that bodily events are constitutive of mental states, without embracing some blunt form of behaviorism? Can it be plausible in the cases of feelings or emotions, let alone in a striking case like pain?

If trembling, grinding teeth, grimacing, and the like are criteria of pain in the sense I proposed, this implies that in a good case – a case where the person before us is actually in pain – the trembling,

grinding teeth and grimacing are constitutive to her being in pain. They are what "being in pain" in this case, in these circumstances *is* – and seeing the trembling, grinding teeth and grimaces is to see the pain of the other person.

Now, it might seem that this claim amounts to too much to be plausible. It is easy to protest that it is a category mistake to talk about "seeing pains". Surely, we ordinarily and legitimately talk about, for example, "seeing the pain in another person's eyes", when we want to emphasize the immediateness of our judgment about her experience. But clearly that is just figurative speaking? Properly speaking, pains can be *felt* but not seen (at least apart from extraordinary synesthetic experiences in the first person). Pain as a subjective experience must at any rate be distinguished from the bodily effects of pain, and it is the latter we can see, not the former. Is this protest justified?

In contemporary discussions of social cognition, "direct social perception" accounts (see e.g. Gallagher & Zahavi 2008, Gangopadhyay & Miyahara 2014) propose that in many instances of social cognition, it is a simple perceptual process that allows us to understand the mental state of a fellow human being. This stands in contrast to the established duo of "theory-theory" and "simulation theory", who assume that such social understanding is achieved by an elaborate cognitive process that uses perceptual input of the bodies of others as data. However, even the accounts that claim that another person's mental state can in favorable conditions be disclosed to us in direct perception (as the locution "seeing the pain in another person's eyes" suggests) arguably have to maintain that we do not perceive mental states in quite the same way as we perceive physical objects; claiming otherwise would be to "violate certain fundamental intuitions about other minds" (Gangopadhyay & Miyahara 2014, 15). Our experience of others is characterized in many cases by immediate givenness, but it is also characterized by the idea of *transcendence* of other minds: the mentality of others is something that does not fully disclose itself to the perception of others, even when being perceptible in some manner (see e.g. Overgaard 2007, Gangopadhyay & Miyahara 2014).

Surely there is something distinctive in our perceptions of fellow human beings (and to some extent our perceptions of living things in general), compared to our perceptions of inanimate objects. According to a common way of putting the matter, we see the bodily events of other human beings as *expressive*. According to Gallagher (2008, 542), "one important element of our understanding of others depends on a direct perception of the other person's actions and expressive movements".

Thanks to such direct perception, typically it is not needed to infer or theorize in order to gain an understanding of what is going on in the mind of another person:

”[I]n most intersubjective situations we have a direct understanding of another person's intentions because their intentions are explicitly expressed in their embodied actions [...]. For the most part this understanding does not require the postulation of some belief or desire that is hidden away in the other person's mind, since what we might reflectively or abstractly call their belief or desire is expressed directly in their behavior” (Gallagher 2005, 224).

The common idea here is that our mode of seeing the mentality of other human beings in, or on, their bodies is to see the bodily events as expressions of mental states. Apparently human babies (normally) develop the capacity to perceptually distinguish between expressive movement of living things and non-expressive movement of non-living things very early in their development, and according to the advocates of ”direct social perception”, this capacity remains the main tool of social cognition in adult life too. Now, expressions are said to be expressions *of* something. The term seems to involve a distinction between the expression and what gets expressed in the expression. That is why saying that we come to know other minds by perceiving the expressive behavior of others can account for the directness of social cognition without flouting fundamental intuitions about other minds, or without falling into behaviorism. Insofar as we (most of us) have a natural ability to recognize expressive behavior and take it in as already diffused with mental meaning, we can say that we perceive the mental states of others ”directly” *by* perceiving the expressions of those mental states (”directly” in the sense that there is no inference or theoretical interpretation involved). But expression of a mental state is not the same thing as the mental state; and we ”see” the mental state in a way distinct from the way in which we literally see its expression, when the latter is conceived simply as bodily movement.

So, considering screams or frowns (and the like) as Wittgensteinian criteria for pains or anger respectively, it seems advisable to think that they are criteria precisely insofar as they are expressions of their mental states. Pain and anger are the mental states that these expressions, among countless others, are expressions *of*, and this is an inherent part of our concepts of pain and anger. But this is to admit that criteria in the context of other minds, after all, are not constitutive of what they are criteria of, because expressions are not constitutive to the mental states they are expressions of. They are acts of displaying the mental states outward, but distinct from the mental states themselves. Recalling McDowell’s remarks quoted above, we can conceive of “direct” and “indirect” cases. In the first type of case, what we perceive may be “the fact itself [...] directly presented to view”, while in the second type of case what we perceive is “his giving expression to

his being in that ‘inner’ state”, which is “not itself [...] the ‘inner’ state of affairs” (McDowell 1983, 387). In both cases, what we perceive on the bodies of others does not, in a sense, “fall short of the fact” of the mental state of the other, but in the latter case what we – directly or primarily – perceive is an expression of a mental state, not a mental state. Is this distinction between the two cases unavoidable?

I suggest that it is not. The concept of “expression” allows a different construal too. Krueger & Overgaard (2012) make the point that “expression” is ambiguous, and that it is not always clear what exactly is meant by saying that we perceive the minds of others in their expressive behavior. From the variety of senses of “expression”, we can distinguish as relevant the following:

First, there is a minimal and simple sense of “expression”, in which it means simply the physical configuration of a face, that is, “facial expression”. This sense of “expression” does not imply that expression *of* anything is involved. Even a wax figure in Madame Tussauds can have an “expression” in this objectified sense.



*Fig. 1. Expression, sense 1.*

Second, there is a sense in which “expression” means the act or vehicle of giving a visual or a verbal representation of something “inner”; a thought, experience, or feeling. It is giving a visual or auditory form for what goes on in the subject’s mind. It is rendering an inner event public; coming up with means of communication that convey to the audience an idea of the contents of the subject’s mind. Thanks to this sense, single words or phrases can be called “expressions”. In this sense, we can say of someone who struggles to verbalize an elusive feeling that she is trying to find “a good expression for her feeling”, and in this sense we can praise a poet’s “skill of expression”. In this sense, expression is construed as consisting in what is expressed, on the one hand, and what does

the expressing, on the other hand. The expressing results from the expressed; it is a transformation or an interpretation of the expressed.

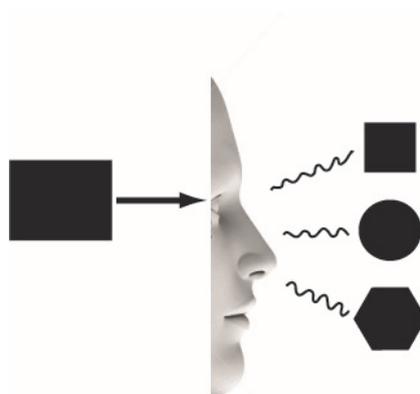


Fig. 2. Expression, sense 2.

Third, there is a more literal meaning of "expression", faithful to the etymology of the word as the combination of *ex + pressare, ex + premere*, "pushing something out". In some cases of what we call human "expressive" behavior, it seems plausible to say that this sense plays a role. It is when we talk about spontaneous, naturally occurring, and to some extent non-voluntary reactions like weeping, sighing, smiling, frowning, drooping, or startling. When we say, for example, "Those tears are an expression of grief" or "Her face expresses fear", the idea of something being literally pressed out of the person is not far-fetched. Here is a feeling surfacing and becoming visible, in the sense in which it is also possible to talk literally about, for example, "the expression of gas from the tank".<sup>3</sup>

---

<sup>3</sup> However, surely both the second and the third sense are together in play when we talk about human expressive behavior in general. If I'm not mistaken, it is somewhat atypical to refer to sweating as an "expression of anxiety" or trembling as an "expression of fear", insofar as these are completely involuntary, and thus lack the "purposeful conveying" aspect of the second sense.

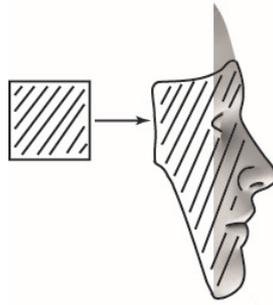


Fig. 3. Expression, sense 3.

If "expression" is construed in the third sense, then the physical, observable object of perception (which can also be referred as the person's "expression" in the first sense) is actually construed as being constitutive of what is expressed. It is, if not the whole, at least a part or an aspect of what is expressed. "Expression" in this sense is an act of pressing out, and what is expressed is the same thing as that which is visible in the expression in the first sense. Krueger & Overgaard (2012) recommend understanding bodily expression in a constitutive sense like this, in order to make proper sense of the claim that the mental states of another person can be objects of direct perception. The account is not obviously vulnerable to accusations of behaviorism, because it does not need to claim that the bodily expressions constitute all there is to the mental state in question; it only claims that they constitute a proper part or an aspect of it.

In light of this way of thinking, it becomes possible to say that bodily events like screaming, grimacing, trembling, and clutching a body part are criteria for pain insofar as they are expressions of pain in the third sense, and that in this capacity they are also constitutive of the other person's pain. Here is a proposal of how to put the picture together:

*There is a diverse and open-ended set of patterns of bodily behavior that serve as our ways of telling that another creature has pains. The patterns of this set have an intrinsic connection to the concept of pain. They belong inseparably to the concept, in the sense that "were we to try to divorce our concept of pain from the behaviors that typically express it [...], that would alter our concept of pain beyond recognition" (Witherspoon 2012, 10). In perceiving other people, we encounter these patterns in the form of events on their faces and bodies, which we may sometimes describe in an objectifying mode, such as when we speak of someone's facial configuration as an "expression" in the first sense. Some cases of*

*encountering such patterns are vague or downright deceptive. Contortions of a stage actor or those of dismembered frog legs do not involve pain. But as soon as it is clear that the case is vague or deceptive, we are no longer entitled to claim that criteria for pain were satisfied in the first place. Rather, we conclude that in these circumstances, the usual criteria for pain are not obviously valid.*

*On the other hand, in other cases (let us call them good cases) the patterns of bodily behavior are actually expressions (in the third sense) of pains. In the good cases, the criterial bodily behavior is constitutive of pain by being an expression (in the third sense) of it. Here we construe pain as a single phenomenon that can remain internal and unexpressed, but when it is expressed, it assumes a perceptible form in facial contortions, screams, and the like.*

It is admittedly possible to insist that what pain *is*, essentially, is just the subjective episode of hurting, and to insist that counting facial and bodily events as literally parts of the mental phenomenon of the other person's pain is just to overlook a self-evident distinction. But even if the view I examine is not our most typical construal of what belongs to the phenomenon of pain, this does not show that it is not a possible and instructive one. And if the view is too much to accept in the case of pain, maybe its plausibility is clearer at least in the case of emotion (cf. McDowell 1983, 387n34). It is fruitful to read some of Wittgenstein's remarks against this background:

“We say ‘The expression in his voice was *genuine*’. If it was spurious we think as it were of another one behind it.—*This* is the face he shews the world, inwardly he has another one.—But this does not mean that when his expression is *genuine* he has two the same.” (Wittgenstein 1953/2001, §606; all emphases in the original)

It is possible to understand the remark that that “the human body is the best picture of the human soul” (Wittgenstein 1953/2001, II, iv, 152) in light of the previous passage. It is not that the outer form of the human body just happens to be a reliable picture of what is going on “on the inside”. It is that our very conception of the “inside”, at least in many contexts, is modeled on the concrete, perceptible outer. Inner, unexpressed sadness is not necessarily, or even primarily, conceived as an abstract, singular phenomenal feeling of sadness. It is more tangibly conceived in terms of having, so to speak, a sad face inside, which has retreated from the surface and left behind a mismatching “outer” face to cover it, thus hiding itself from view. But its hiddenness is no more metaphysically grounded than the hiddenness of the face of a masquerader who hides her face behind an artificial

mask. And just like the masquerader's face, the "inner" sad face – the *same* face, not a copy of it – can become perceptible for others again, in the form of a genuine expression. Consistent with the way McDowell interprets the ultimate aim of Wittgenstein's discussions of bodies, minds and criteria, this line of thought should allow us to "restore the concept of a human being to its proper place, not as something laboriously reconstituted [...], but as a seamless whole of whose unity we ought not to have allowed ourselves to lose sight in the first place" (McDowell 1983, 384).

## 5. Concluding Remarks

I have argued that the Wittgensteinian bodily criteria for mental states should be interpreted as constitutive of those mental states, rather than external evidence for them. This interpretation is consistent with Wittgenstein's overall critique of the inner and outer, and with his view that (at least some) mental states of others can be perceived by perceiving their bodies. It is an important challenge to this view to show that it is not implausible behaviorism, and that ascribing the view to Wittgenstein does not saddle him with a behaviorist position that he would have rejected.

Things that can in various situations serve as criteria for mental states – tears of sadness, cries of pain, smiles of joy, and so on – are the kind of things of which we customarily say that they are expressions of mental states. The mere fact that we are able to see such things as expressions shows that we do not conceive of sadness, pain, or joy in a behavioristic fashion. That is, we do not take these things to be only dispositions to behave in certain ways, but subjective experiences that are shown outwards, expressed. However, this may make us think that even in the best possible case we can only perceive in others the expressions of their mental states, as opposed to the "inner" states that those expressions are expressions of, the latter being some completely distinct and ultimately private things. If so, we will still be in the grips of what Wittgenstein criticizes as a false picture of the inner and outer. We can escape the picture by acknowledging that there is not necessarily an absolute distinction between the expression and the expressed. In many cases, human expressions can be understood as literal expressions: instances of coming into view of something that was previously hidden.

This is not to say that there is no difference between, say, feeling sadness oneself, and seeing sadness expressed in the face and body of another person. That difference is obviously very real. But the deeper point is that the concepts of sadness, pain, joy and the like are concepts that find their application in the shared human life. Both the "inner" manifestations of sadness, pain or joy as subjective experiences, and their "outer" manifestations as expressions, are united by those concepts

as single phenomena of human life – phenomena that each of us encounters in two ways: as experienced in first person, and as perceived in the expressions of others.

## References

Addis, Mark (1995): “Criteria: The State of the Debate”. *Journal of Philosophical Research*, vol. 20, pp. 139-174.

Albritton, Rogers (1959): “On Wittgenstein’s Use of the Term ‘Criterion’”. *The Journal of Philosophy*, vol. 56, no. 22, pp. 845-857.

Baggini, Julian & Fosl, Peter S. (2010): *The Philosopher’s Toolkit: A Compendium of Philosophical Concepts and Methods*. Second edition. Oxford: Wiley-Blackwell.

Bohl, Vivian & Gangopadhyay, Nivedita (2013): “Theory of Mind and the Unobservability of Other Minds”. *Philosophical Explorations*, vol. 17, no. 2, pp. 203-222.

Cavell, Stanley (1979): *The Claim of Reason. Wittgenstein, Skepticism, Morality, and Tragedy*. Oxford: Oxford University Press.

Gallagher, Shaun (2005): *How The Body Shapes the Mind*. Oxford: Clarendon Press.

Gallagher, Shaun (2008): “Direct Perception in the Intersubjective Context”. *Consciousness and Cognition*, vol. 17, no. 2, pp. 535-543.

Gallagher, Shaun & Zahavi, Dan (2008): *The Phenomenological Mind. An Introduction to Philosophy of Mind and Cognitive Science*. London and New York: Routledge.

Gangopadhyay, Nivedita & Miyahara, Katsunori (2014): “Perception and the Problem of Access to Other Minds”. *Philosophical Psychology*, DOI: 10.1080/09515089.2014.895935

- de Gaynesford, Maximilian (2002): "Blue Book Ways of Telling: Criteria, Openness and Other Minds". *Philosophical Investigations*, vol. 25, no. 4, pp. 319-330.
- Glendinning, Simon (1998): *On Being With Others. Heidegger – Derrida – Wittgenstein*. London and New York: Routledge.
- Glock, Hans-Johann (1996): *A Wittgenstein Dictionary*. The Blackwell Philosopher Dictionaries. Oxford: Blackwell.
- Hyslop, Alec (2014): "Other Minds". *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition), Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/spr2014/entries/other-minds/>>.
- Krueger, Joel & Overgaard, Søren (2012): "Seeing Subjectivity: Defending a Perceptual Account of Other Minds". In *Consciousness and Subjectivity*, ed. by Miguens & Preyer, Heusenstamm: Ontos Verlag, pp. 273-296.
- Luckhardt, C.G. (1983): "Wittgenstein and Behaviorism". *Synthese*, vol, 56, no. 3, pp. 319-338.
- Lycan, Gregory W. (1971): "Noninductive Evidence: Recent Work on Wittgenstein's 'Criteria'". *American Philosophical Quarterly*, vol. 8, no. 2, pp. 109-125.
- Malcolm, Norman (1954): "Wittgenstein's Philosophical Investigations". *The Philosophical Review*, vol. 63, no. 4, pp. 530-559.
- McDowell, John (1983): "Criteria, Defeasibility, and Knowledge". In McDowell: *Meaning, Knowledge, and Reality*, Cambridge (Mass.): Harvard University Press, 1998, pp. 369-394.
- Overgaard, Søren (2007): *Wittgenstein and Other Minds*. New York and London: Routledge.
- Thornton, Stephen P. (2014): "Solipsism and the Problem of Other Minds". *The Internet Encyclopedia of Philosophy*, ISSN 2161-0002, <http://www.iep.utm.edu> [27.7.2014]
- Vaaja, Tero (2012): "Meeting Other Minds". In *From Minds to Persons: Proceedings from the First MLAG Graduate Conference*. Ed. by Vieira da Cunha, Morando, & Miguens. FLUP: Porto.

Witherspoon, Edward (2012): "Wittgenstein on Criteria and The Problem of Other Minds". *Oxford Handbooks Online*. Oxford University Press.

<http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199287505.001.0001/oxfordhb-9780199287505-e-22> [16.2.2015]

Wittgenstein, Ludwig (1953/2001): *Philosophical Investigations*. Translated by G.E.M. Anscombe. The German text, with a revised English translation, 3<sup>rd</sup> edition. Oxford: Blackwell.

Wittgenstein, Ludwig (1958): *The Blue and Brown Books. Preliminary Studies for the "Philosophical Investigations"*. Oxford: Blackwell.

Wittgenstein, Ludwig (1980): *Remarks on the Philosophy of Psychology: volume II*. Edited by von Wright & Nyman. Translated by Luckhardt & Aue. Oxford: Blackwell.

Wittgenstein, Ludwig (1981): *Zettel*. 2<sup>nd</sup> edition. Edited by Anscombe & von Wright. Translated by Anscombe. Oxford: Blackwell.

Wittgenstein, Ludwig (1982): *Last Writings on the Philosophy of Psychology: volume I*. Edited by von Wright & Nyman. Translated by Luckhardt & Aue. Oxford: Blackwell.