

# Subjectivity in psychoanalysis, phenomenology and cognitive neuroscience: A conceptual study

Jasper Feyaerts

Supervisor: Prof. Dr. Stijn Vanheule

Co-supervisor: Prof. Dr. Gertrudis Van de Vijver

A dissertation submitted to Ghent University in partial  
fulfilment of the requirements for the degree of  
Doctor of Psychology



“[...] “cogito ergo sum” ubi cogito, ibi sum [...]  
Of course, this limits me to being there in my being  
only insofar as I think that I am in my thought;  
to what extent I really think this concerns me alone  
and, if I say it, interests no one”

-Lacan, *Instance of the Letter*-



# Table of Contents

---

Acknowledgements	viii
<b>Introduction</b>	
Subjectivity: problems and context	1
Research questions	4
Methodology	6
Overview of the thesis	6
References	9
<b>Chapter 1: Madness, subjectivity and the mirror stage</b>	
<b>Lacan and Merleau-Ponty</b>	11
Introduction	11
The metaphysics of madness	12
Naturalistic attitude and lived experience	14
Lacan reads Wallon	15
Merleau-Ponty reads Lacan reads Wallon	19
Conclusion	23
References	25
<b>Chapter 2: How to return to subjectivity?</b>	
<b>Natorp, Husserl and Lacan on the limits of reflection</b>	27
Introduction	27
Natorp's critical psychology: subjectivation vs objectivation	30
Husserl's response: "the principle of all principles"	33
Natorp and Husserl reconsidered	40
The Lacanian 'split subject'	43
Conclusion	48
References	52
<b>Chapter 3: Cartesian meditations</b>	
<b>Lacan and the question of enunciation</b>	56
Introduction	56
Enunciation vs enunciated	57
Unconscious, truth, subjectivity	57
References	63
<b>Chapter 4: He or It (the Thing) thinks, not I</b>	
<b>Self-monitoring and verbal hallucinations</b>	65

Introduction	65
Thinking as movement	67
Comparator-supposed-to-know	72
Ghostly doubles	73
I think ... I think?	76
Who (He or It (the Thing)) thinks what?	79
Conclusion	81
References	84
<b>Chapter 5: The logic of appearance</b>	
<b>Dennett, phenomenology and psychoanalysis</b>	87
Introduction	87
Dennett and phenomenology: the seeming of seeming	88
The principle of appearance qua appearance	89
Dennett's heterophenomenology: Cartesianism without a theater	95
Return of the Cogito: serious, all too serious	98
Dennett and psychoanalysis: the curious case of the 'objectively subjective'	101
Freud and Lacan between phenomenology and cognitive science	102
Psychoanalysis' 'abominable mess'	105
Decentering appearance: belief in belief	106
Belief in the uncanny or uncanny belief?	109
Conclusion	112
References	114
<b>Chapter 6: Expression and the unconscious</b>	118
Introduction	118
Explaining first-person authority: from detection to expression	122
Detectivism	123
Wittgenstein's expressivism	127
Expression and the 'as-if' of the unconscious	132
Unconscious: adverbial vs substantial	132
The unconscious structured as a cogito	135
The cogito structured as an unconscious	139
Conclusion: expression and the unconscious	142
References	148
<b>Discussion and Conclusion</b>	
Discussion of our conclusions	152
General considerations: the question of subjectivity	157
The concept of 'subject' and its critique	158
Psychological verbs and the question of self-reference in schizophrenia	162
References	166
<b>Nederlandstalige samenvatting:</b>	
<b>Subjectiviteit in de psychoanalyse, fenomenologie en cognitieve neurowetenschap</b>	168

Probleemstelling	168
Onderzoeksvragen	170
Methode	171
Opzet	172
Resultaten per hoofdstuk	174
Discussie	176
Referenties	179



## Acknowledgments

First of all, I would like to thank the University of Ghent and the Department of Psychoanalysis & Clinical Consulting for offering me the opportunity to freely conduct my research and to devote my time to the abstract labor of writing this dissertation. Given the current *Zeitgeist* of pragmatism and control, I take this to be a most precious commodity.

The members of my doctoral guidance committee have been a continual support throughout this period. First and foremost, I want to thank Stijn Vanheule for listening as a true analyst to the jumble of thoughts figuring in my Cartesian theater - however chaotic, rash and untimely they may have been – and for his open-minded and elegant abstinence throughout these years of free association (who dares to say now that psychoanalysis is without effect?). Next, Gertrudis Van de Vijver for her careful reading of my papers and the biting, but as always, intelligent and instructive criticisms. Paul Verhaeghe deserves a special mention for enduring his persistent headache in guiding my thesis while ‘knowing-quite-well’ that every new paper would make matters worse – acted like a true heroic fetishist! Last but not least, Lieven Jonckheere for reminding me about the fact that psychoanalysis is not merely a lofty, theoretical enterprise, but also a clinical practice (at least, sometimes).

Next, I am grateful to the wonderful Center for Subjectivity Research in Copenhagen for hosting me as a visiting researcher – special thanks to Dan Zahavi, Josef Parnas, Mads Henriksen and Merit Lynnerup for their warm hospitality. Likewise, the members of my department deserve credit for putting up with me throughout these years: Gregory Bistoën, David Schrans, Joachim Cauwe, Dries Dulsster, Eline Trenson, Junior Ingouf, Nathalie Laceur, Melissa De Smet, Rosa De Geest, Mattias Desmet, Reitske Meganck, Els Vandebussche and all the others. My international friends and interlocutors are thanked for the many conversations and their valuable suggestions to this work: Dominiek Hoens, Jan De Vos, Marc de Kesel, Aaron Schuster, Patrice Maniglier, Russell Grigg, Pierre-Henri Castel, Robert Pfaller, Vincent Descombes, ...

Yet, most importantly, I want to thank my parents for their blind and unremitting confidence without which this thesis would not have been possible (I’ll skip over the details). To my father, I hope this counts as a ‘return on investment’. To my mother, I hope this takes away at least one of your worries. Finally, I’m grateful for the love and support of my partner-in-crime, Dorinda Bielen (say ‘Bon voyage!’).

J.F.





# Introduction

---

## Subjectivity: problems and context

In his 'Philosophy and the Scientific Image of Man' from 1963, the American philosopher Wilfrid Sellars proposed a concise, yet compelling diagnosis of our philosophical, scientific and social predicament. Sellars famously talked about a discord fueled by the conflict between two apparently opposing and competing "images" (1963, p. 5) of man in the world: one the one hand, a *manifest* image which represents the tentative and provisional collection of ideas and assumptions we spontaneously adopt to characterize ourselves and the world we live in. Chief among these are ideas closely related to our immediate self-understanding as being 'subjects' endowed with (self-)consciousness, a certain extent of freedom of choice, gifted with (un)reason, desires, affects, and so on. On the other hand, a relatively recent but continually expanding *scientific* image of man as a "complex physical system" (1963, p. 25) – that is, an image conspicuously *unlike* its manifest double, but one which can be progressively distilled from various burgeoning scientific disciplines, including physics, neurophysiology, evolutionary biology, and, more recently, cognitive science. In Sellars' view, the aforementioned discord results from the observation that modern science, although it has immeasurably enriched and challenged our understanding of phenomena by way of techniques and resources quite foreign to common sense, has now begun to deploy those resources closer to home in the course of the investigation into the nature of mind. In that way, it has at the same time begin to encroach on a realm of phenomena hitherto thought (or maybe even hoped) to lay *beyond* the purview of science, specifically, those phenomena which are variably grouped under the headings of 'subjectivity', 'meaning', 'experience' and 'the first-person perspective', and which for many philosophers – past and present – harbor the very key of what makes that image 'human' in the first place. One of the crucial tasks facing philosophers today is then, according to Sellars, to arbitrate the apparent conflict between these two increasingly divergent images. In what way should the competing claims of the manifest and scientific images be adjudicated and how should their relative rights and privileges be understood? Is science, for example, and as some philosophers insist, constitutively incapable of providing a satisfactory account of what we mean by 'meaning' or 'first-person experience' (e.g. Nagel, 1986)? Is any attempt to study first-person, subjective experience from the third-person, objectifying viewpoint bound to lead the investigator into a "performative contradiction" (Habermas, 2008)? Or is it rather, to voice the other end of the spectrum, the authority of our pre-scientific intuitions which should be denounced as a typical philosopher's fallacy – i.e. a failure of imagination paraded as an insight into (transcendental) necessity

(Dennett, 1978) – and hence, our spontaneous ‘folk-psychological’ evidences that finally need to be called into question (Churchland, 1989)?

In more recent years, Sellars’ antagonistic prognosis has found a specific application in the contemporary discussions surrounding the presumed – to use Levine’s felicitous expression – ‘explanatory gap’ between cognitive science and philosophical accounts of subjectivity. Yet, although complaints about alleged ‘gaps’ separating the breathtaking achievements of the natural sciences and the relative paucity of valuable results in the sciences of mind have indeed become something of a persistent *leitmotiv* in the philosophical and psychological literature ever since Hume or Leibniz, it seems that the current lament has taken up a somewhat different flavor. That is, although cognitive science is generally accepted as the first truly scientific theory of mind and considered to have achieved for the mind what Galileo achieved for movement – i.e. putting the investigation on the right track, the track leading through the well-known game of conjecture and refutation, to an increasingly better understanding of the cognitive dimension of the mind -, many researchers nowadays nevertheless seem to express a certain *Unbehagen* with regard its current status. That nagging awareness is never more explicit than in recent discussions about the relation between, on the one hand, cognitive descriptions of the mind’s functioning, as they are typically conceived in terms of some kind of ‘information processing’ operating within our black box, and on the other hand, subjectivity, loosely defined within the literature in terms of Thomas Nagel’s (1974) famous idiom of ‘what it is like’ to be a cognizing mind, or the way things ‘seem’, ‘appear’ or are ‘experienced’ from a subjective or first-person perspective. More to the point, the core of the argument which is raised against cognitive science is that, although it is granted cognitive science succeeds in charting the cognitive, inner *reality* of the mind, although it has indeed provided us with ever-increasing sophisticated descriptions of the cognitive dimension of the mind ‘as it is in itself’, it nevertheless (or at least, up until now) *fails* to account for its phenomenal *appearance*, for how this mind appears ‘to itself’ or how it is experienced subjectively. In the concise wordings of Roy et al. (1999): “explaining what is happening *in* the black box is not explaining what is happening *for* the black box” (p. 12, italics added). According to the advocates of this explanatory gap argument then, it is one thing to try to account for what is going on in our brain – at whatever level of explanation, whether neurobiologically or functionally defined – when we engage in various kinds of cognitive behavior (e.g. memory, perception, attention, ...), it is something fundamentally *different* to try to account for what appears or seems to be going on when we engage in those types of behavior from a first-person point of view.

So it is this surprising Kantian twist in the recent assessment of cognitive science’s putative results which has led many researchers to argue that a mere functional analysis of the inner workings of our mind does not provide an exhaustive nor adequate explanation as long as the first-person dimension of subjective experience is not taken into consideration as an explicit area of investigation. Furthermore, in light of this explanatory gap and the related phenomenological critique, several authors in different research domains - ranging from cognitive (neuro)science, analytic philosophy of mind to

psychopathology and the so-called consciousness studies -, have argued for a *return* to classical phenomenological approaches of subjectivity in the hope to enhance our understanding of the relation between cognitive processes and their phenomenal manifestations. Still according to these authors, phenomenology, as originated in the work of Husserl and subsequently developed by e.g. Merleau-Ponty, Heidegger and Sartre, could provide us with a refined theory and a descriptive approach of subjectivity as the necessary requirement prior to any naturalistic effort of ‘bridging the gap’. On strong versions of that account, phenomenological descriptions of our most immediate ‘lived’ experience become the *quod erat explicatum* for any reductionist undertaking, it explicates the manifest image as manifest image and unilaterally dictates which gap cognitive science still is required to bridge. As Chalmers put it: “Experience is the most central and manifest aspect of our mental lives, and indeed is perhaps the key *explanandum* in the science of mind. Because of this status as an *explanandum*, experience cannot be discarded like the vital spirit when a new theory comes along” (1995, p. 206).

Yet, not all authors seem to share the same confidence with respect to what has been called a veritable “phenomenological renaissance” (Parnas, Sass & Zahavi, 2008, p. 578) and its ability to offer a genuine contribution to the perennial project of establishing a naturalistic explanation of mind. This critique has been advocated along two main lines of argument. The first line maintains that although approaching subjectivity from a naturalistic perspective is possible, a phenomenological approach in terms of privileged first-person access and phenomenal experience is not. Most notably, Daniel Dennett has disputed the idea that subjectivity and consciousness would entail any ‘data’ which could only be accessed from a first-personal point of view, and in fact, more strongly, has argued that there is “no such thing as actual phenomenology” (1991, p. 365) or the possibility of a “first-person science” (Ibid.). The second line maintains that although a science of subjectivity is possible, a *naturalized* one is not. This critique is firmly rooted in the *transcendental* nature of Husserlian phenomenology. Moreover, it could be argued that the whole point of Husserl’s enterprise is to claim that there is a dimension of the mental that necessarily escapes, while being simultaneously presupposed by, the natural sciences and requires a specific kind of investigation. And this dimension is all the more crucial because Husserl considered it as ontologically anterior to natural being and ‘matters of facts’, both psychological and physical. Indeed, Husserl goes as far as writing “We are fighting against the naturalization of consciousness” (1910-1911/2002, p. 254).

Surprisingly, one notorious intellectual tradition which is almost entirely *absent* from these discussions is the Freudian and Lacanian discipline which goes by the name of psychoanalysis. This observation is all the more surprising given that the Sellarsian set-up, and more in particular, its most recent manifestation in terms of the uneasy tension between the first- and third-person perspectives, arguably touches the very core of what is at stake in psychoanalytic theory. Given the peculiar distribution of the distinction between neurocognitive reality and manifest appearance in the contemporary debate, are there indeed not valuable lessons still to be gained from Freud’s original insight that, even in the field of ‘psychical appearances’, “reality is not necessarily what it appears to

be” (Freud, 1915, p. 171)? Or what about Lacan’s distinctive claim that the Cartesian cogito is not so much a mere fiction ready to be dissolved through scientific progress as was the case for other fictive entities such as the ‘phlogiston’ or ‘élan vital’ (Dennett, 1991, p. 79), but itself the most suitable *persona* from which the meaning of ‘the unconscious’ could be deduced (Lacan, 1981, p. 126)? Yet, neither should this overt silence be conveniently put away as an insignificant occurrence: as noted by Ricoeur (1997) and Laplanche (1999), from its inception, psychoanalysis was precisely founded, for better or worse, on a systematic *refusal* of the disjunction between the sub-personal and personal levels of explanation. Regardless of whether this constitutes its “invaluable contribution” (Ricoeur, 1997, p. 66) or rather an “abominable mess” (Wittgenstein, 1978, p. 313), this Janus-faced character at least partly explains psychoanalysis’ remarkable timidity in the face of the philosophical challenges that arise from the contemporary debate. On the one hand, what exactly is left of the once so seductive and sweeping Freudian image of the three successive humiliations of man now that we can safely add a fourth one – e.g. Dennett? Indeed, it seems that Freudian psychoanalysis, steeped as it is in the intentional language of desires, negations and disavowals, now finds itself amongst those traditional ‘manifest’ philosophies threatened by the more recent neurocognitive disenchantment. And is it, furthermore, also not quite imaginable that Dennett’s response to Lacan’s fierce anti-humanistic formula in *Science and Truth* according to which “science’s man does not exist, only its subject does” (1966/2006, p. 730), would be the sobering ‘nor does its subject, although it might *seem* to’? On the other hand, however, given Freud’s inaugurative departure from Brentano’s and Husserl’s views on intentionality and self-consciousness (see Seron, 2015; Bernet, 2002) or Lacan’s unreserved critique of both Sartre’s existential psychoanalysis (1949/2006, p. 80) and Merleau-Ponty’s bodily unconscious (Lacan, 1982-1983), it is equally clear that the intellectual relation between phenomenology and psychoanalysis is complicated enough to avoid any easy wedding of psychoanalysis with the philosophies of manifest experience.

### **Research questions**

Taken together, the general discussion we just briefly outlined on the vicissitudes of subjectivity within Sellars’ distinction between the manifest and scientific images, its most recent expression in terms of the contemporary debate on the ‘explanatory gap’ between cognitive science and phenomenology, and finally, the ambiguous position of psychoanalytic theory with regard to that philosophical conjecture, reveal the need

- (1) to develop a philosophical account of how and on what grounds the competing claims from neurocognitive and phenomenological theories of subjectivity should be adjudicated;
- (2) to explicate which philosophical assumptions guide the different traditions in their respective approaches of subjectivity and the question of its naturalistic explanation or elimination;

- (3) to elucidate the precarious nature of psychoanalytic claims on the nature of subjectivity through a juxtaposition of psychoanalytic theory with contemporary phenomenology and neurocognitive science.

These general problems lead in turn to the following more specific research questions:

- (1) In what way is the relation between the first- and third-person perspectives conceptualized in neurocognitive accounts, phenomenology and psychoanalytic theory? Is this relation a matter of epistemic asymmetry and if so, how is the issue of first-personal knowledge conceived with respect to the question of epistemic justification?
- (2) What exactly is meant in each tradition with such popular terms as ‘subjectivity’, ‘manifest experience’, ‘seeming/appearance’ and ‘the first-person’? For example, does talk about ‘subjectivity’ necessitate a ‘subject’? If so, what are the differences/similarities amongst the respective disciplines in terms of how this ‘subject’ is understood and how does it differ from the ordinary concept of ‘the empirical person’?
- (3) If subjectivity is acknowledged as a legitimate and necessary field of inquiry, what kind of methods are deemed appropriate to investigate and make this subjective dimension accessible? How are the proposed methods itself related to the way subjectivity is conceived in each tradition?
- (4) Is the difference between first-personal descriptions of experience and third-personal explanations in neurocognitive accounts, reducible to the difference between ‘consciousness’ and ‘unconsciousness’? Specifically for psychoanalytic theory: how does its account of ‘the unconscious’ relate to sub-personal neurocognitive accounts? Does the psychoanalytic conception of ‘unconsciousness’ leave the authority of the first-person untouched or does it constitute a challenge to that authority?
- (5) What is the relevance of these discussions with regard to concrete examples from the research literature on psychopathology? More specifically, in what way do philosophical assumptions on subjectivity influence the way in which schizophrenic pathology is approached?

## **Methodology**

Since these questions are all generally philosophical and conceptual in scope, they will be variously addressed and elucidated by means of a series of conceptual research articles. These articles are all papers that are published in scientific journals or as chapters in an academic book, accepted for publication, or recently submitted. Throughout these articles, careful attention will be paid to the conceptual study of both classic and more recent philosophical authors which have developed influential views on how to approach subjectivity in their respective traditions. For the phenomenological tradition, these authors include Husserl, Merleau-Ponty and Sartre; for neurocognitive science and analytic philosophy of mind, special attention is given to Dennett, Frith and Wittgenstein; for psychoanalytic theory, Freud and Lacan will be our focus. In each article, sustained effort is put into developing each perspective in its own right, while the different claims which emerge from these readings are only subsequently juxtaposed in order to assess their relative strengths and weaknesses during the concluding discussion. Where possible and appropriate, the conceptual analyses will be discussed and illustrated by means of empirical examples from the research literature on psychopathology and its implications for future research made clear. While our general research questions are the main framework in each consecutive article, enough space will be provided to focus on the specific details and questions which, while not explicitly set forth in the introduction of our thesis, may nevertheless emerge from our discussion of these different authors and topics. In that case, the relevance and implications of these specific issues for our general research questions will be pointed out and discussed.

## **Overview of the thesis**

In this section we provide the reader with a broad overview of this thesis. Since this dissertation consists of a collection of articles, the reader might have the impression that link between chapters is not always clear or that arguments and findings unnecessarily repeated. This overview aims at addressing these concerns.

The *first* chapter of our thesis – ‘Madness, subjectivity and the mirror-stage: Merleau-Ponty and Lacan’ – discusses the intimate relation between subjectivity and madness as treated in Lacan’s early writings and in Merleau-Ponty’s seminal work *Phenomenology of Perception* (2002). Central within this paper is the difference between Lacan’s and Merleau-Ponty’s respective readings of Henry Wallon’s psychogenetic model of the ‘mirror stage’. Both authors took special interest in this model because of its potential to clarify their own theoretical positions: whereas for Merleau-Ponty the mirror image constitutes an ‘objectifying’ illusion which both can and needs to be reduced to a pre-objective stratum of lived experience, Lacan considers the specular image as being non-reducible and formative for subjectivity as such. We then further discuss how the difference separating Lacan’s and Merleau-Ponty’s

viewpoints on the relation between madness and subjectivity is related to the function they ascribe to misrecognition in the formation of subjectivity, as well as to the subsequent possibility (Merleau-Ponty) or impossibility (Lacan) of overcoming this misrecognition.

Both the concepts of misrecognition and objectification in relation to subjectivity are further discussed in the *second* chapter of our thesis – ‘How to return to subjectivity? Natorp, Husserl and Lacan on the limits of reflection’-, in which the contemporary discussion on the ‘explanatory gap’ is used to clarify these concepts. In the first part of that chapter, a more detailed discussion is offered of the current phenomenological arguments as to the supposed limitations of neurocognitive explanations vis-à-vis subjectivity. In the next parts, the methodological question of the possibility of studying subjectivity through phenomenological reflection is considered through a reading of the Neo-Kantian philosopher Paul Natorp, his discussion and disagreement with the founding figure of transcendental phenomenology Husserl, and concluded with a Hegelian synthesis of these positions through Lacan’s conceptual couple of enunciation/enunciated. In the final part, the implications and relevance of our discussion for the ‘explanatory gap’-argument is pointed out and the critical position of psychoanalysis with regard to that argument discussed.

Lacan’s conceptual distinction between ‘the subject of enunciation’ and ‘the subject of the statement’ is taken further in our *third* chapter – ‘Cartesian mediations: Lacan and the truth of enunciation’- in which the distinction is used to offer a reading of Freud’s ambiguous therapeutic precept according to which “in analysis, the neurotic has to tell more than he knows”. From this small remark, a specific psychoanalytic version of the Cartesian and Husserlian procedures of *epoché* is proposed in order to clarify the nature of ‘truth’ in psychoanalysis. Differences and similarities with respect to phenomenological notions of truth and the Sartrean figure of ‘conscious negativity’ are pointed out.

In the *fourth* chapter – ‘He or it (the thing) thinks, not I: self-monitoring and verbal hallucinations’, the results of our philosophical discussions on subjectivity thus far are used to offer a critical reading of recent, neurocognitive explanations of verbal hallucinations in schizophrenia. More specifically, Frith’s predictive, self-monitoring model is questioned in terms of both, on the one hand, the general idea of subjectivity set forth by this model, premised on the distinction between ‘ownership’ and ‘agency’ of self-conscious thought, and on the other hand, the specific description it offers of the phenomenological nature of hallucinatory experience. On the basis of our critical discussion, in the last part, we delineate a number of requirements which should be met by neurocognitive explanations of verbal hallucinations.

In the *fifth* chapter – ‘The logic of appearance: Dennett, phenomenology and psychoanalysis’ – we further pursue our discussion from chapter two on the competing claims of neurocognitive and phenomenological accounts with regard to subjectivity, this time however in terms of Sellars’ distinction



between the manifest and scientific images of man. Starting from this distinction, we aim to develop and contrast three different possible ways of responding to Sellars' diagnosis: Dennett's philosophical reconstruction of neurocognitive science, contemporary phenomenology and psychoanalysis. We will suggest that these respective traditions and the substantial differences between them can be understood in terms of a 'logic of appearance'. Related to this are differing ideas about the rights and limits of the first-person perspective, the relation between conscious experience and belief, and the issue of naturalization. In the final part of that chapter, we try to specify, on the basis of a detailed reading of the disagreement between Dennett and phenomenology, in what way psychoanalytic theory could respond to these different issues.

The conclusion from our fifth chapter as to the psychoanalytic critique of the idea of first-personal 'immunity to error' with respect to the phenomenological shift to 'seeming' and 'appearing' is then further elaborated in our ultimate, *sixth* chapter – 'Expression and the unconscious'. The central focus of this chapter is twofold: on the one hand, to develop Wittgenstein's account of the phenomena of 'first-person authority' and the 'asymmetry between the first- and third-persons' with respect to the use of psychological expressions in the present indicative. This account is critically juxtaposed with phenomenological and neurocognitive, materialist proposals to explain or either to reject 'first-person authority' in terms of (the deficiency of) self-reference and epistemic asymmetry. On the other hand, in the second part, we employ Wittgenstein's concept of 'expression' to offer a non-epistemic account of the meaning of 'unconscious'. Both Freud's and Lacan's respective ways to justify the use of unconscious are critically discussed and an alternative solution for the logical meaning of unconscious is proposed.

## References

- Bernet, R. (2002). Unconsciousness consciousness in Husserl and Freud. *Phenomenology and the Cognitive Sciences* 1(3), 327-351.
- Chalmers, D. (1995). Facing up to the Problem of Consciousness. *The Journal of Consciousness Studies* 2 (3), 200-219.
- Churchland, P. (1989). Folk Psychology and the Explanation of Human Behavior. *Philosophical Perspectives* 3, 225-241.
- Dennett, D. (1978). *Brainstorms: philosophical essays on mind and psychology*. Harmondsworth: Penguin.
- Dennett, D. (1991). *Consciousness Explained*. New York: Back Bay Books/Little, Brown and Company.
- Freud, S. (1915). The unconscious. *Standard Edition* 14, 166-215.
- Habermas, J. (2008). The language game of responsible agency and the problem of free will: How can epistemic dualism be reconciled with ontological monism? *Philosophical Explorations* 10 (1), 13-50.
- Husserl, E. (2002 [1910-1911]). Philosophy as Rigorous Science. *New Yearbook for Phenomenology and Phenomenological Philosophy* 2, 249-295.
- Lacan, J. (1981). *The Seminar 1964, Book XI, The Four Fundamental Concepts of Psychoanalysis*. New York & London: W.W. Norton & Company.
- Lacan, J. (1982-1983). Merleau-Ponty: In Memoriam. *Review of Existential Psychology and Psychiatry* 18(1/2/3), 73-81.
- Lacan, J. (2006 [1949]). "The Mirror Stage as Formative of the *I* Function", in J. Lacan & J-A. Miller (Eds.), *Ecrits*. New York and London: W.W. Norton, 75-81.
- Lacan, J. (2006 [1966]). "Science and Truth", in J. Lacan & J-A. Miller (Eds.), *Ecrits*. New York and London: W.W. Norton, 726-745.
- Laplanche, J. (1999). *Essays on Otherness*. London: Routledge.
- Merleau-Ponty, M. (2002). *Phenomenology of Perception*. London and New York: Routledge.
- Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review* 83 (4), 435-450.
- Nagel, T. (1986). *The View From Nowhere*. Oxford: Clarendon Press.
- Parnas, J., Sass, L.A. & Zahavi, D. (2013). Rediscovering psychopathology: the epistemology and phenomenology of the psychiatric object. *Schizophrenia Bulletin* 39(2), 270-277.
- Ricoeur, P. (1977). *Freud and Philosophy: An Essay on Interpretation*. New Haven: Yale University Press.
- Roy, J.M., Petitot, J., Pachaud, B. & Varela, F.J. (1999). "Beyond the Gap: An Introduction to Naturalizing Phenomenology, in J. Petitot, F.J. Varela, B. Pachaud & J.M. Roy (Eds.), *Naturalizing Phenomenology*. Stanford (California): Stanford University Press, 1-82.
- Sellars, W. (1991). "Philosophy and the Scientific Image of Man," in *Science, Perception and Reality*, Atascadero: Ridgeview Publishing Company, 1-40.
- Seron, D. (2015). "L'équivalence entre "mental" et "conscient" chez Brentano," in M. Gyemant & D. Popa (Eds.), *Approches phénoménologiques de l'inconscient*. Hildesheim : Georg Olms Verlag, 15-30.



# 1

## Madness, Subjectivity and the Mirror Stage Lacan and Merleau-Ponty<sup>1</sup>

---

### 1. Introduction

In his 1949 essay on *The Mirror Stage as Formative of the I Function*, while discussing the function of the Ego within psychoanalytic experience, Lacan holds that “the subject’s capture by his situation gives us the most general formulation of madness – the kind found within the asylum walls as well as the kind that deafens the world with its sound and fury” (p.80). In a similar way, in his *Phenomenology of Perception* from 1945, Merleau-Ponty considers pathological subjectivity as “a loss of plasticity” (2002, p.151) in the subject’s intention to renew its perceptual field. Furthermore, both Lacan and Merleau-Ponty contend that the ‘capture’ or ‘loss of plasticity’, which are distinctive for madness, reveal something on the nature human subjectivity as such.

In this chapter, we discuss the intimate relation between subjectivity and madness as treated in Lacan’s early writings<sup>2</sup> and in Merleau-Ponty’s seminal work on the *Phenomenology of Perception*. Most clarifying for our discussion is the difference between Lacan’s and Merleau-Ponty’s readings of Henry Wallon’s psychogenetic model of the mirror stage. Both took special interest in this theory because of its potential to clarify their own theoretical positions: whereas for Merleau-Ponty the mirror image is an illusion that needs to be reduced to lived experience, Lacan considers the specular image as non-reducible, and as formative for the experience of subjectivity. Next, we reframe this difference in terms of the above stated convergence between both authors on the status of madness vis-à-vis subjectivity. This brings us to the conclusion that the difference separating Lacan’s and Merleau-Ponty’s viewpoints on the relation between madness and subjectivity is related to the function they ascribe to *misrecognition* in the formation of subjectivity, as well as to the subsequent possibility (Merleau-Ponty) or impossibility (Lacan) of overcoming this misrecognition.

---

<sup>1</sup> This chapter is published as a book chapter – Feyaerts, J. & Vanheule, S. (2015). “Madness, subjectivity and the mirror stage: Lacan and Merleau-Ponty”, in P. Gherovici & M. Steinkoler (Eds.) *Lacan on madness: madness, yes you can’t*. NY: Routledge, 159-172.

<sup>2</sup> Lacan’s discussion of madness and psychosis changed substantially throughout time (see Vanheule, 2011). In this chapter, we only focus on Lacan’s discussion of psychosis and madness prior to the start of his public seminar.

## 2. The metaphysics of madness

Before addressing Lacan's and Merleau-Ponty's respective appropriations of Wallon's theory of the mirror stage, it is helpful to turn to Lacan's paper *Presentation on Psychological Causality*, dating from 1947. This text presents a remarkable proximity to the thread followed by Merleau-Ponty in his *Phenomenology of Perception*. In the same vein as Merleau-Ponty's phenomenological critique of science's 'naturalistic attitude'<sup>3</sup>, Lacan takes issue with the "organicist theory of madness" of the French psychiatrist Henry Ey, who, according to Lacan, "cannot relate the genesis of mental problems [...] to anything but the play of systems constituted in the material substance (l'étendue) located within the body's integument" (1947, p.124). Such play "always rests in the final analysis on molecular interaction of the *partes-extra-partes*, material-substance type that classical physics is based on" (Ibid.). In giving this critique, Lacan follows the phenomenological refutation of the naturalistic attitude, which assumes that madness, qua natural kind, is a determined object among other objects in the world. He proposes that instead of yet another mechanistic biological model, a theory on psychic causality is needed, one that actually transcends the particular issue of madness:

The problem that madness thus kindles in us owing to its pathos provides a first answer to the question I raised about the human value of the phenomenon of madness. And its metaphysical import is revealed in the fact that it is inseparable from the problem of signification for being in general – that is the problem of language for man (Ibid., p.135).

So far from reducing the psychological causality involved to some sort of fortuitous curiosity, Lacan stresses its *metaphysical* dimension in the sense that madness sheds light on the nature of subjectivity as such. Moreover, Lacan argues that the metaphysical value of madness is bound up with the problem signification and language poses for man. Language enables the human being to fictionalize reality and to live it through as a problem of truth, in a continuous interweaving between misrecognition and recognition. "The madman believes he is different than he is" (Ibid., p.139), but, as Lacan contends, the same holds for "the king who really thinks he is a king" (Ibid.). This split in subjectivity between '*being*' and '*believing*', takes us to the heart of subjectivity itself. So rather than viewing madness as an expression of biological frailty, or as a deplorable adversity that goes against the true nature of human subjectivity, Lacan holds that madness is the "permanent virtuality of a gap opened in his essence" (Ibid., p.144). In sum, according to Lacan, madness is a permanent possibility for human subjectivity, and this is what makes up its metaphysical significance. Instead of being a contingent aberration, subjectivity

---

<sup>3</sup> The 'naturalistic attitude' is the stance taken by he who aims "to grasp the nature of reality" and to describe such "reality" in terms of some objective description which will accurately characterize the "thing-in-itself" apart from one's experience of it" (Toombs, 1993, p.40).

cannot genuinely be comprehended, and would not be what it is, without madness as the limit of its freedom. We return to this idea later in the text.

In 1947 Lacan relates the idea of psychical causality to the concept of the *imago* which plays a constitutional role in the formation of subjectivity. The *imago* reveals the historical dimension of subjectivity in that its history will take shape as a “more or less typical series of ideal identifications” (Ibid., p.145). These successive imaginary identifications, taken as an ever-developing ensemble, a *whole-à-venir*, constitute the Ego. More precisely, the *imago* is the specular image by which the Ego of the subject is constituted through identification, without thereby conflating the actual being of the subject with this image. It is the image by which man believes himself a man, and by which the madman believes himself other than he is. The *imago* is the image by which man recognizes himself (as an image) while at the same time misrecognizing his being<sup>4</sup>. In other words, at the basis of both madness and human subjectivity the same kind of identification can be found. This is why madness partakes in the metaphysical condition of the subject and should be comprehended in terms of the problematic relation between being and believing as such.

Likewise, Merleau-Ponty (2002) fulminates against reducing madness to the mere outcome of numerous causal factors that would determine its make-up. Again, rather than placing madness *jenseits* human existence, he believes that madness reflects a metaphysical condition that should be comprehended in its own right:

There can be no question of simply transferring to the normal person what the deficient one lacks and is trying to recover. Illness, (...), is a complete form of existence and the procedures which it employs to replace normal functions which have been destroyed are equally pathological phenomena. It is impossible to deduce the normal from the pathological, deficiencies from the substitutive functions, by a mere change of the sign. We must take substitutions as substitutions, as allusions to some fundamental function that they are striving to make good, and the direct image of which they fail to furnish. The genuine inductive method is not a ‘differential method’; it consists in correctly reading phenomena, in grasping their meaning, that is, in treating them as modalities and variations of *the subject’s total being* (Ibid., p.124-125).

Consequently, for Merleau-Ponty, madness should not be reduced to nothing but a bit of the world, enclosed within the realm of science as a mere object of biological, psychological or sociological

---

<sup>4</sup> As such, Lacan’s subject as subjected to the image/signifier is closer to a certain ‘ontological dualism’ where “thought is in disharmony with the soul” (Lacan, 1990, p.6). Moreover, this brings Lacan in closer proximity to Descartes, in that Lacan tends to place the imaginary/symbolic orders that engender subjectivity and the organic body in separate domains. However, this does not imply that Lacan is uncritical of Descartes. Lacan’s alleged proclamation for ‘a return to Descartes’ should not be taken as the uncritical return to the *res cogitans* of the Cogito-like subject. As Adrian Johnston observes (2008, pp. 53-54), Lacan’s formulations on the relation between subjectivity and the body nevertheless allow for the rejection of a false dilemma tacitly governing many contemporary discussions. That is, the choice between on the one hand, a strict mind-body dualism (à la Descartes) and, on the other hand, a reduction of subjectivity to its corporeal substance (whether this substance be the organic body of the natural sciences (e.g. “the brain”) or the lived body of modern phenomenology). We will return to this issue in our conclusion.

investigation. Contrary to empirical and intellectualistic explanations which respectively treat madness as, on the one hand, a contingent epiphenomenal effect of pathogenic distortions, and on the other, as a “perversion of the will” (Ibid., p. 144), like Lacan, Merleau-Ponty considers madness as a ‘modality or variation’ that concerns ‘the subject’s vital area’, as an expression of ‘the subject’s total being’, which moreover should be revealed through phenomenological analysis.

Interestingly Lacan explicitly addresses this phenomenological analysis in his 1947 paper. More specifically, he revisits Merleau-Ponty’s point that *lived experience* needs to be considered prior to any objectification and even prior to any reflexive analysis that interweaves objectification and experience. As Lacan contends:

For Merleau-Ponty’s work decisively demonstrates that any healthy phenomenology, that of perception, for instance, requires us to consider lived experience prior to any objectification (...). Let me explain what I mean: the slightest visual illusion proves to force itself upon us experientially before detailed, piecemeal observation of the figure corrects it; it is the latter that allows us to objectify the so-called real form. Reflection makes us recognize in this form the a priori category of extension [l’étendue], the property of which is precisely to present itself "partes extra partes," but it is still the illusion in itself that gives us the gestalt action that is psychology's true object here (1947, p.146).

However, Lacan’s recourse to Merleau-Ponty’s phenomenological maxim is highly ambiguous. After all, Lacan only focuses on the lived experience of the *mirror image*, as the fundamental but necessary illusion lying at the heart of subjectivity, instead of the phenomenological focus on the lived experience of perception in its pre-objective dimension. Before turning to our discussion of Lacan’s and Merleau-Ponty’s respective readings of this mirror stage, let us first try to clarify this phenomenological maxim by shortly presenting Merleau-Ponty’s analysis of lived experience and his critique of the naturalistic attitude.

### **3. Naturalistic attitude and lived experience**

As suggested by Lacan, one of the central motifs guiding Merleau-Ponty’s phenomenological account of perception consists indeed in the uncovering of the unreflective ground of lived experience, a dimension which is simultaneously presupposed and forgotten within the naturalistic attitude. For Merleau-Ponty, the first task of a genuine philosophy is therefore the phenomenological critique of this attitude which, both in its empirical and intellectualistic guises, tends to forget the subjective ground on which it nevertheless operates. That is, the naturalistic attitude obfuscates the constituting role subjectivity plays in perception and sees the world as an already constituted entity that hence appears as a whole of ready-made objects. Merleau-Ponty thoughtfully summarizes the logic of this naturalistic repression as follows:

Obsessed with being, and forgetful of the perspectivism of my experience, I henceforth treat it as an object and deduce it from a relationship between objects. I regard my body, which is my point of view upon the world, as one of the objects of that world. My recent awareness of my gaze as a means of knowledge I now repress, and treat my eyes as a bit of matter. They then take their place in the same objective space in which I am trying to situate the external object and I believe that I am producing the perceived perspective by the objects on my retina. In the same way I treat my own perceptual history as a result of my relationships with the objective world; my present, which is my point of view on time, becomes one moment of time among all others, my duration a reflection or abstract aspect of universal time, as my body is a mode of objective space (2002, pp.81-82).

First of all, as this passage indicates, Merleau-Ponty recourse to *the body* as the primary medium of our involvement with the world marks an important difference between his ‘bodily phenomenology’ and Husserl’s incessant focus on ‘consciousness’ as the central frame of departure. If, for Husserl, the phenomenological return to the world is first of all a return to the world of transcendental consciousness, Merleau-Ponty’s own version is “a return to the world which precedes knowledge, of which knowledge always speaks, and in relation to which every scientific schematization is an abstract and derivative sign-language - as is geography in relation to the country-side in which we have learnt beforehand what a forest, a prairie or river is” (Ibid., p.x). And for Merleau-Ponty, this pre-predicative knowledge of lived experience is effectuated through the body: instead of a transcendental subject as a bodiless spectator of the world, Merleau-Ponty’s subject is an embodied subject with hands and feet immersed in the world. Secondly, the object of the naturalistic repression is indeed this embodied subjectivity which constitutes our world.<sup>5</sup> As a consequence of this repression, the body is usually treated as a desubjectivized object among other objects, owing nothing to the experience of the world. Thirdly, Merleau-Ponty’s phenomenological analysis aims at lifting this naturalistic repression, which should enable the pre-objective, lived experience to come to the fore. Pathological subjectivity is an important issue within this phenomenological effort to go beyond fixed representation precisely because it points towards the ambiguity of existence, like a real-life thought experiment set up to reveal the dimensions of constitution/constituted, pre-objective/objective and becoming/being, operative in normal subjectivity.

#### **4. Lacan reads Wallon**

---

<sup>5</sup> However, the question remains as to why the body is repressed in the first place? In our opinion, the concept of the body is at one and the same time a step in the right direction in comparison with Husserl’s and Heidegger’s bodiless analytics, however, to the extent that Merleau-Ponty’s recourse to the body results in an unproblematic *embodied subjectivity*, it also represents the unthought remainder which threatens to haunt the temporal logic of naturalistic repression and concomitantly proves to be a serious drawback to his phenomenology. It is precisely at this point that Lacan’s own conceptualizations on the relation between Cartesian subjectivity and the *Real of the body* can provide a welcome alternative.



In order to refine our analysis of this phenomenological rapport that conjoins madness and subjectivity, we now turn to the different readings Lacan and Merleau-Ponty propose of Wallon's psychogenetic theory of the mirror stage as formulated in *Les origines du caractère chez l'enfant* (1949, pp. 218-237) and how these bear upon their respective theories of subjectivity.

In the chapter of *Les Origines* entitled "The Body proper and Its Exteroceptive Image" (Ibid., p. 219 *et passim.*), Wallon introduces a whole zoo of creatures to demonstrate the decisive disparity that separates animal and human ways of relating to the mirror image. For Wallon, the distinctive feature separating the human infant from, for example, the drake, is the ability to grasp the reciprocal relation between the self and its reflection. An animal, by contrast, is a creature that is unable to identify with its image, as is illustrated by the example of a drake that acquired the strange habit, since the death of its partner, of staring at a reflecting windowpane. Wallon writes: "Without doubt his own reflection could more or less fill in the void left by the absence of his companion" (1949, p.219, our translation). The ability of the drake to find consolation in the image is concomitant with its inability to identify with the mirror image. In a similar way, Lacan contrasts the behavior of the human infant and the chimp when confronted with their mirror image: the infant engages in triumphant jubilation and playful self-discovery, while the chimps ends up with sheer indifference. Lacan relates this fundamental difference qua lived experience of the mirror image to the remarkable contrast between the early instrumental self-sufficiency of the animal, and man's prematurity at birth, which he tentatively associates with the perceptual tendency for recognizing the human Gestalt early in life.

A second important idea from Wallon's work that resonates in Lacan's study of the mirror stage is the consideration of the jubilant mirror experience as the mythical beginning of self-differentiation. The mirror stage enables the transition from a passive state of raw immediacy to a situation of imaginary mastery and, then, to symbolic representation: "The development of the infant demonstrates by what degrees immediate experience, the undifferentiated, dispersed, and transitory impressions of brute sensibility must become dissociated, fixed by images initially concrete and seemingly coextensive with their object, and then give way to symbolic transmutations of pure and stable representation" (Ibid., p.183). For Wallon, the mirror experience is thus also the "prelude to symbolic activity", enabling a transition from partial, sensorial perceptions to what he calls the "symbolic function" (Ibid., pp. 230-231). Lacan's early work is often read in the same way: the mirror experience forms the imaginary ground for the symbolic *I* to come to the fore.

Wallon's detailed observations clearly function as a conceptual paradigm for Lacan's understanding of the mirror image. However, the Lacanian appropriation of the mirror stage is more than the simple juxtaposition of *Wallon avec Lacan* presented thus far. In several respects, Lacan's ongoing revision and articulation of the mirror mechanism throughout his teachings represents a complex synthesis of several strands of thought in psychoanalysis, philosophy and experimental psychology. Let us address this issue by means of the following questions:

- (i.) *Why does the yet to be constituted subject look in the mirror?*
- (ii.) *What does the yet to be constituted subject see in the mirror?*

The first question is obviously a question of ‘beginnings’, more specifically a question concerning the anterior ontogenetic conditions that drive the infant towards the seductive lure of the mirror image. In the opening chapter of *Civilization and Its Discontents*, Freud (1929) already poses the same question a propos the genesis of his psychological topography. In a discussion of the ‘oceanic feeling’ – the lived experience surrounding the ego – Freud argues:

The idea of men’s receiving an intimation of their connection with the world around them through an immediate feeling which is from the outset directed to that purpose sounds so strange and fits in so badly with the fabric of our psychology that one is justified in attempting to discover a psychoanalytic – that is, a genetic – explanation of such a feeling (1929, pp.65-66).

Henceforth, Freud concludes, “a unity comparable to the ego cannot exist in the individual from the start” (1914, p.77). Lacan endorses this Freudian inclination towards a ‘genetic explanation’ of the Gestalt-like illusion of the ego:

In effect, if one starts from the notion of original narcissism, perfect as regards libidinal investment, if one conceives of the primordial object as primordially included by the subject in the narcissistic sphere, as a primitive monad of enjoyment [jouissance], to which is identified the infant nursling [nourisson], one has difficulty seeing what would be able to lead a subjective way out [sortie subjective] (2001, p.410).

Lacan’s concerns about the notion of original narcissism are relatively straightforward here: original narcissism, as the blissful state of a self-sufficient and unified wholeness, is, considered from an ontogenetic perspective, untenable insofar as it curtails the possibility to conceive the development of any subjectivity whatsoever. The question as to ‘why’ the Lacanian subject-à-venir looks in the mirror should thus be comprehended starting from an ontogenetic perspective which stipulates the non-all of human nature prior to the acquisition of ego-like subjectivity.

Contra Wallon, Lacan consequently stresses that man is, by definition, a *disadapted animal* burdened with a disordered Gestalt-like imagination. The mythical tale of the mirror stage is not so much the *nec plus ultra* of human teleological development that Wallon described, as it is the expression of a ‘primordial discord’ which characterizes man’s relation to nature.<sup>6</sup> More specifically, both Freud

---

<sup>6</sup> Although Wallon similarly stresses the child’s prematurity at birth, he considers it as a positive condition for the subsequent sociability of the child. The difference with Lacan’s reading lies in the fact that Wallon thinks of prematurity as an anterior condition that is replaced by a more adaptive sociability. Lacan, by contrast, emphasizes the fundamental impossibility to abolish this primordial discord and disadaptation to the natural environment.

and Lacan relate this non-all of human nature to the idea that all human beings are born *prematurely*. As a consequence, human infants are utterly depended on others for carrying out all basic vital tasks. Adopting Freud's terminology, Lacan defines this prolonged 'primordial discord' as a state of helplessness (*Hilflosigkeit*). It is against this ontogenetic background that the peculiarly (de)formative function of the specular image has to be understood: the illusory completeness of the subject's body image as reflected in the mirror provides him with the promise of a unified wholeness that compensates for human helplessness. Yet, at the same time, the attraction exercised by the seductive lure of the Gestalt-illusion answers to a completely different logic than the one observed in other animals: animals respond instinctively to (Gestalts of) other animals, but they do not alienate themselves in the mirror image. This is why Lacan holds that "the mirror stage is a drama whose internal thrust is precipitated from insufficiency to anticipation" (1949, p.78): the non-all of human nature is supplemented and warded off by the ideal imaginary unity of the mirror image.

This brings us to the second question: what does the yet to be constituted subject see in the mirror? Or what does the image in the mirror reflect? For Wallon, the reflected mirror image seen is in a certain way indifferent: the psychological developmental task the infant has to accomplish is first of all the active mental integration of real model and mirror image. For this integration to occur the reciprocal relation between the child's own body and the mirror image is not needed; other bodies can do the same developmental job. In Wallon's analysis, the difficulty lies in a *spatial realism* dominating the child's epistemological relation to its *Umwelt* that prevents it from linking the actual figure with the virtual one in the mirror. That is, the pre-mirror stage child does not yet understand that the two bodies located at two points in space – the introceptively felt body and the virtual image in the mirror – constitute only one body. Therefore, the essential factor that develops or is achieved during the mirror phase is the recognition of spatial values or, more precisely, the coordination of what was perceived as two bodies in two distinct places. The child's behavior suddenly demonstrates a jubilant comprehension of the reciprocity between model and image. The Aha-Erlebnis of the mirror stage thus concerns an epistemological break-through in which spatial realism is replaced by a more accurate conception of the relation between model and image.

For Lacan, by contrast, the account of the mirror image is far less heroic. First of all, Lacan stresses the fact that the image seen in the mirror is not indifferent at all. Indeed, for Lacan, the mirror image concerns the child's *own body image*, which henceforth helps to transcend any "human or artificial support" (1949, p.76). Lacan's recurrent emphasis on the reflection of one's own body – and not that of the mother or any other caretaker – expresses the idea that by identifying with the mirror image, the subject attains a *causa sui*-status in transcending the obstructions of his support. The triumphant jubilation accompanying the subject's identification demonstrates the sudden experience of mastery over the premature distress, the possibility of being '*maître/m'êtré à moi-même*' via the reflected totality of the salutary imago. This identification with the imaginary totality of one's own body makes up the ground for the formation of the ego and later for successive secondary identifications.

Secondly, however, in Lacan's reading of the mirror stage, the epistemological status of the subject's identification with the mirror image is far removed from Wallon's *adequatio rei et intellectus* of the correspondence between real model and virtual image since it also sets the stage for structural misrecognition: experiences that don't fit the seductive wholeness of the mirror image are subsequently warded off. To put it in the words of Mladen Dolar: "when I recognize myself in the mirror, it is already too late" (1996, p.138). There is a fundamental price to be paid for recognizing oneself in the mirror: the seductive illusion of the mirror image makes me lose my 'self-being' prior to the identification, that is, the immediate coincidence with myself in being. This is why Lacan equates identification with alienation: to "Know thyself" by means of the mirror image is to "Alienate thyself" in an image that is other, to recognize oneself before the mirror is to misrecognize oneself in the alienating Gestalt of an illusion of totality. Important, however, is that despite this dramatic depiction of the mirror stage in terms of the annihilating effects of the image vis-à-vis authentic being beyond representation, the presumed choice between the illusion of the Gestalt and the mythical state of self-being is in fact a Hegelian one. To reformulate Lacan's favorite example of the forced choice between 'Your money or your life' in terms of our discussion: either one chooses the illusion of the mirror image ... or one loses both. For Lacan, subjectivity itself is engendered by a loss of being, it is, in its essence, the loss of the immediacy of 'self-same being'. Consequently, the forced choice between 'subjectivity' and 'being' is no choice at all: one can only 'choose' the illusory nature of mediated subjectivity – holding on to presumed more original state would entail losing the 'self' of being and thus both.

To conclude: Lacan proposes a reading of Wallon's mirror experiments that differs from the latter's orthopedic teleological version in different respects. For Lacan, the specular image by which the subject comes to (non-)being is without any doubt an illusion. Nonetheless, it is an illusion that has the structure of a truth. As Lacan puts it in his 1947 essay: there is a "law of our becoming" commanding one to "become such as you are" (p. 145). This illusion is indeed the human illusion par excellence by which we obey to this "law of our becoming", by which one pays the price of becoming by separating oneself from one's own being. In this respect, madness is the permanent possibility for human subjectivity because it is a faithful expression of "this gap opened up in his essence" (Ibid., p. 144). The liberation from the slavish adhesion to immediacy of 'being', enables the subject to experience his freedom as detached from 'what there is'. Madness is therefore the *nec plus ultra* of this freedom. In madness, alienation to the illusory image that makes up subjectivity is most extreme. So rather than constituting a blame for human subjectivity, madness provides the most faithful expression of the mechanism that commonly makes us think that we are who we are.

## **5. Merleau-Ponty reads Lacan and Wallon**

Merleau-Ponty's phenomenological reading of Wallon's mirror stage differs from Lacan's and was provided during a series of lectures in 1950-1951, entitled *Les relations avec autrui chez l'enfant*

(1997). As noted in our introduction, like Lacan, Merleau-Ponty takes special interest in Wallon's theory because of its potential to enhance his own phenomenological project of 'returning to the things themselves'. Moreover, Wallon's description of the mirror experience explicitly points towards the constitutive relation between visual perception (the Gestalt discerned in the mirror) and any subsequent symbolic activity. As such, it opens towards the ambiguity of perception, so cherished by Merleau-Ponty since it sets the stage for a critique of the naturalistic attitude. Let us recall Merleau-Ponty's description of naturalistic repression, this objectivistic attitude that always-already forecloses embodied subjectivity which nonetheless functions as its constitutive ground:

I detach myself from my experience and pass to the *idea*. Like the object, the idea purports to be the same for everybody, valid in all time and places, and individuation of an object in an objective point of time and space finally appears as the expression of a universal positing power. I am no longer concerned with my body, nor with time, nor with the world, as I experience them in antepredicative knowledge, in the inner communication that I have with them. I now refer to my body only as an idea, to the universe as an idea, to the idea of space and the idea of time. Thus 'objective' thought (in Kierkegaard's sense) is formed – being that of common sense and of science – which finally causes us to lose contact with perceptual experience, of which it is nevertheless the outcome and the natural sequel (2002, p.82).

For Merleau-Ponty, it is phenomenology's first philosophical task to return to this 'antepredicative knowledge', to return to this original perceptual experience that forms the natural ground for any ulterior objectification. In this respect, it should be clear why Merleau-Ponty rejoins Wallon's description of the mirror stage as the gradual subordination of the virtual body, as seen in the mirror, over the proprioceptive real body, as it is originally experienced by the child. Recall that Wallon considers the mirror stage as the phase during which the child overcomes its 'spatial realism', its premature epistemological attitude towards the independence of the interoceptive 'felt body' and the 'virtual body' in the mirror. In order to overcome this solipsistic stance towards imaginary representation, the child has to relate the *pre-objective experience* of his own body to the objectified body in the mirror image. As Merleau-Ponty describes this transition:

There is, says Wallon, a space that adheres to the image. Every image tends to occur in this space, the mirror image too. This inherent spatiality will, according to Wallon, be reduced through intellectual development. We gradually learn to fold the specular image to the interoceptive body, and conversely, to treat the quasi-locality, the pre-spatiality of the specular image as an appearance that stands beneath the unique space of real things. (...) In this way, the space adhering to images is substituted for an ideal space (1997, p. 194, our translation).

This transition from 'the interoceptive body' to 'the ideal body' of the specular image hence designates, to borrow the language of the *Phenomenology of Perception*, the transition from the 'the pre-objective

perception' of the body, the quasi-sacral 'lived body' of Merleau-Pontian philosophy, to the objective body of the naturalistic attitude which represses the former in favor of a specular illusion. The objectified body in the mirror, departs from the original body that nonetheless "is always near me, always there for me, (...) that is never really in front of me, that I cannot array (...) before my eyes, that remains marginal to all my perceptions, that is *with me*" (Ibid., p.104, our translation).

In the same vein, Merleau-Ponty addresses Lacan's interpretation of the mirror stage as a drama of alienation through which the original experience of the body is substituted for a frustrating idealization of body as 'body-ideal':

The self-image, at the same time that it enables self-knowledge, enables a kind of alienation: I am no longer what I immediately felt I was, I am this picture of me offered by the mirror. There occurs, in the words of Dr. Lacan, a "captivation" of me by my spatial image. Suddenly, I leave the reality of the lived experience of myself for a constant reference to this ideal self, fictitious or imaginary, of which the mirror image forms the first draft. In this way, I am torn from myself and the mirror image prepares me for another, even more serious alienation, which will be the alienation by the other (1997, p.203, our translation).

Merleau-Ponty thus seems to read Lacan in the following (phenomenological) way: the mirror stage represents the dramatic transition from 'a lived experience of the self' to 'the fictional experience of an ideal imago'. It is this second experience which is the presupposition of the empirical and intellectualistic renditions that erroneously treat the subject as an object in the world, and which therefore ought to be reduced in order to liberate the phenomenological subject of original self-experience. Just like Lacan, Merleau-Ponty stresses that the subject is captivated by the illusory ideal image of its reflection. However, contra Lacan, the formative illusion of the mirror image is henceforth treated as a hindrance towards real phenomenological subjectivity. For Merleau-Ponty, the illusion seen in the mirror deceives the subject about its own true origin. The ambiguity of perception, i.e. subjectivity as both constituting-constituted, is once again eliminated in favor of a solidified world of ready-made (body-) objects. Instead of treating the body as the way *I am* in and for the world, the subject captured by the mirror image sees the body as a contingent feature that *I have*. The body is thus reduced to an image that is nothing more "than an appearance in a visible world which has nothing to do with me" (Ibid., p.197). Consequently, imaginary recognition is for Merleau-Ponty above all a misrecognition vis-à-vis the original bodily experience of the phenomenological subject. Conforming to the temporal logic of naturalistic repression, the image by which the subject becomes estranged from itself, now appears as the ground of its own reflection. The body-as-object appears as the formal ground of my self-experience, congealing the whole of bodily experience "as a crystal placed in a solution suddenly crystallizes it" (2002, p. 82).

Three conclusions can be drawn from Merleau-Ponty's interpretation of the mirror stage. First, Merleau-Ponty seems to adhere to what we might call a 'subjective dualism' in that he neatly

distinguishes between two forms of 'lived experience' or 'subjectivity'. On the one hand, true phenomenological subjectivity which is intrinsically linked to the pre-objective experience of the body and remains faithful to the constitutive pole that surrounds the ambiguity of human existence. On the other, ideal subjectivity which is concomitant with the constituted pole of human existence, cut off from its primordial bodily dimension by the alienation in an image that is 'other'.

Second, apart from being a neutral phenomenological analysis, his phenomenological rendition of the mirror stage seems to engender a *normative* analysis of subjectivity. That is, what Merleau-Ponty considers to be the philosophical maxim for any adequate account of perception, equally applies to his analysis of the mirror stage. As he explains:

The task of a radical reflection, the kind that aims at self-comprehension, consists, paradoxically enough, in recovering the unreflective experience of the world, and subsequently reassigning to it the verificatory attitude and reflective operations, and displaying reflection as one possibility of my being. (...) Hence reflection does not itself grasp its full significance unless it refers to the unreflective fund of experience which it presupposes, upon which it draws, and which constitutes for it a kind of original past, a past which has never been present (2002, pp. 280-281).

Although Merleau-Ponty seems to hesitate with regard to the question of the precise ontological status of the unreflective fund underlying any original experience of the self and the world, he nevertheless sees it as a *conditio sine qua non* for any genuine account of subjectivity, something phenomenology time and again 'has to return to'.

Third, this ontological/normative scheme permeating Merleau-Ponty's reading of the mirror stage and the greater part of his *Phenomenology of Perception* obviously bears upon his phenomenological analysis of pathological subjectivity. Indeed, more than anyone else, it is the madman who, for Merleau-Ponty, represents the subject as completely immersed in the constituted pole of a blocked subjectivity, cut off from the permanent possibility to renew the sense of its own existence, imprisoned in the static world of former habits and lost prospects:

For these patients the world exists only as one readymade or congealed, whereas for the normal person his projects polarize the world, bringing magically to view a host of signs which guide action, as notices in a museum guide the visitor" (Ibid., p.129).

The phenomenological rapport between the illusory nature of mirrored subjectivity and madness as presented by Merleau-Ponty thus reveals itself in the following way: the captivation in the image is indeed once again an illusion, leading to a statical and objective view of ourselves. The illusion is however a deceiving one: "In order to be determined by an external factor, it is necessary that I should

be a thing. However, “neither my freedom nor my universality can admit of any eclipse” (Ibid., p. 505). The eclipsing nature of the mirror image exists, when we consider the genuine subject of phenomenological corporeality, only as a second order derivative vis-à-vis the field of possibility that characterizes our bodily being-in-the-world, our original and universal commitment to the world. The madman, in contrast, is completely absorbed in the constituted pole of its own existence, and to the extent that he is indeed ‘determined by an external factor’, be it that of an image or the incessant prospect of an inevitable delusion, fundamentally bereaved of the ground of every deliberate *Sinngebung*. In this way, the phenomenological analysis of madness, although it is spared from the objectifying tendency of naturalistic explanations that negate the madman’s being-in-the-world, ultimately amounts to viewing the madman as the incarnated version of the naturalistic fallacy itself.

## 6. Conclusion

Whereas both Lacan and Merleau-Ponty take madness seriously, and assume that madness can learn us something of the nature of human subjectivity, they actually arrive at contrasting conclusions. From a Lacanian perspective, three fundamental remarks can be made about Merleau-Ponty’s analysis of the mirror stage as spelled out above.

First, whereas Merleau-Ponty distinguishes between phenomenal and ideal subjectivity as two concomitant subjective states, Lacan juxtaposes ‘subjectivity’ and ‘being’ as mutually exclusive categories. From a Lacanian point of view, it is not only ‘too late’ when I recognize myself in the mirror, there is also no way to return. Thus considered, the phenomenological project of ‘returning to the things themselves’ is nothing but a reverie that takes shape as an *après-coup* effect of a fundamental loss of being. For Lacan, there is no ‘true’ subject to return to beyond the alienation: phenomenological reduction as the royal road to a more original self-experience ultimately is a dead end. For Lacan, lived experience is precisely an effect of the alienation in the image. The latter is – to put it Kantian terms – the condition of possibility of any subjectivity whatsoever.

Second, and closely connected to what we called Merleau-Ponty’s *normative analysis*, the very notion of a pre-reflective *embodied subjectivity* is, from a Lacanian point of view, a contradiction in terms. Despite many similarities, Lacanian psychoanalysis and Merleau-Ponty’s phenomenology are fundamentally at odds with regard to the (non-) rapport between mediated subjectivity and its corporeal conditions. Merleau-Ponty tends to replace the cartesian dichotomy between *res cogitans* (thinking substance) and *res extensa* (extended substance) with a more Aristotelian fusion between mind and body as it is expressed in the concept of the ‘lived body’. Like Merleau-Ponty, Lacan objects to considering the cogito as a point of certainty or a metaphysical ‘first’ principle. However, unlike him, Lacan does not reject the Cartesian cogito, ‘ce pauvre moi’ of the specular image, because it is a non-existent fallacy. He rather considers it as an effect, an outcome of symbolic-imaginary (mis-)recognition, instead of a first-level principle of indubitable departure. To explain why Lacan continues to refer to the



Cartesian cogito, Miller productively resumes Lacan's distinction between 'being' and 'having' in its relation to the body:

"Having a body" is significant in its difference from "being a body". We justify the identification of being and body in the animal, but not in man, because no matter how corporeal he is, corporified, he is also made a subject by the signifier, that is to say, he is made lack-in-being. This lack in being as effect of the signifier divides being and body, reducing the body to the status of "having it" (2001, p.21).

For Lacan, the subject is only a subject because of its subjection to symbolic-imaginary mediation and therefore it cannot identify itself with its body. This is also the reason why Lacan continues to refer to a cogito-like subjectivity and why his theory is simultaneously a critique of Merleau-Ponty's phenomenology: for Lacan, the only outcome of the normative call to return to the phenomenological 'lived body' is the ultimate effacement of the subject.

Third and finally, given these Lacanian reservations vis-à-vis Merleau-Ponty's phenomenology, it becomes clear how Lacan's reflections on the mirror stage fundamentally differ from the phenomenological version and how they consequently lead to a different view on the relation between subjectivity and madness. As Lacan contends in his paper on the mirror stage:

This misrecognition [the one that characterizes the ego] can be seen in the revolt through which the madman seeks to impose the law of his heart onto what seems to him to be the havoc of the world. This is an "insane" enterprise - but not because it suggests a failure to adapt to life, which is the kind of thing people often say in our circles, whereas the slightest reflection on our experience proves the dishonorable inanity of such a viewpoint. It is an insane enterprise, rather, in that the subject does not recognize in this havoc the very manifestation of his actual being, or that what he experiences as the law of his heart is but the inverted and virtual image of that same being. He thus doubly misrecognizes it, precisely so as to split its actuality from its virtuality (1947, p.140).

Once more, Lacan defends 'the insane enterprise of madness' as a fundamental expression of human subjectivity. Rather than being a failure to adapt to life as it actually is, madness is the proper heir of the subject's lack of being. Madness is the price we all pay for the loss of the immediacy of being, hence, of the so-cherished freedom that is considered to be the distinctive feature separating man from the animal. Although the madman is indeed blocked in a constituted subjectivity, and although he is indeed 'the limit of our freedom', the madman is – instead of being the perfect opposite of a more original phenomenological subjectivity – our most faithful companion. In short: both madness and freedom are possibilities of a human being due to "the permanent virtuality of a gap opened up in his essence" (Ibid., p.144). To deny this gap in a nostalgic movement towards phenomenological subjectivity is to deny subjectivity itself.

## References

- Dolar, M. (1996). "At First Sight", in R. Salecl & S. Zizek (Eds.), *Gaze and Voice as Love Objects*. Durham, N.C.: Duke University Press, 129-153.
- Freud, S. (1914). "On Narcissism", in *Standard Edition* 14. London: Hogarth Press, 67-102.
- Freud, S. (1929). "Civilization and Its Discontents" in *Standard Edition* 21. London: Hogarth Press, 57-146.
- Lacan, J. (2006 [1947]). "Presentation on Psychological Causality", in J. Lacan & J.-A. Miller (Eds.) *Ecrits*. New York and London: W.W. Norton, 123-160.
- Lacan, J. (2006 [1949]). "The Mirror Stage as Formative of the I Function", in J. Lacan & J.-A. Miller (Eds.), *Ecrits*. New York and London: W.W. Norton, 75-81.
- Lacan, J. (2001 [1960-61]). *Le Séminaire 1960-1961, Livre VIII: Le transfert*. Paris: Seuil.
- Lacan, J. (1990). "Television", in J. Copjec (Ed.), *Television/ A Challenge to the Psychoanalytic Establishment*. New York and London: W.W. Norton, 1-46.
- Johston, A. (2008). *Zizek's Ontology: A Transcendental Materialist Theory of Subjectivity*. Northwest University Press.
- Merleau-Ponty, M. (2002). *Phenomenology of Perception*. London and New York: Routledge.
- Merleau-Ponty, M. (1997). *Parcours*. Editions Verdier.
- Miller, J.-A. (2001). The Symptom and the Body Event, *Lacanian Ink* 19 (Fall), 4-47.
- Vanheule, S. (2011). *The Subject of Psychosis: A Lacanian Perspective*. Palgrave Macmillan.
- Wallon, H. (1949). *Les Origines du Caractère chez l'Enfant*. Quadrige: Presses Universitaires de France.



## 2

# How to return to subjectivity: Natorp, Husserl and Lacan on the limits of reflection<sup>7</sup>

---

### 1. Introduction

In recent years we have witnessed a resurgence of the issue of subjectivity in different research domains ranging from cognitive (neuro)science, analytic philosophy of mind to psychopathology and the so-called consciousness studies. One of the central impetuses for this return to subjectivity is related to a growing awareness on the part of philosophers and scientists alike, working within these various related research domains, of a certain *Unbehagen* with what is widely considered to be the dominant paradigm in the theorization of the mind from the 1960's onwards: i.e. cognitive science. Although it was initially expected that cognitive science would be able to fulfill Hume's famous ambition to become 'the Newton of mental phenomena', many researchers nowadays agree that some foundational issues still stand in need of further clarification (Parnas, Sass & Zahavi, 2013). Preeminent among these issues is the question about the relation between, on the one hand, cognitive descriptions of the mind's functioning, typically conceived as some kind of 'information processing' activity within our black box, responsible for generating various kinds of cognitive behavior, and on the other hand, subjectivity, loosely defined within the literature in terms of Thomas Nagel's (1974) expression of "what it is like" to be a cognizing mind, that is, the way things 'seem', 'appear' or are 'experienced' from a subjective or first-person perspective. Moreover, the core of the argument raised against cognitive science is that it would suffer from an 'explanatory gap' vis-à-vis subjectivity; that although it has provided us with ever-increasing complex descriptions of the cognitive dimension of the mind 'as it is in itself', it nevertheless fails to account precisely for how 'this mind' appears to itself, how the functioning of the mind is experienced subjectively. Or in the concise wording of Roy et al. (1999): "explaining what is

---

<sup>7</sup> This chapter is published as an article – Feyaerts, J. & Vanheule, S. (2015). How to return to subjectivity? Natorp, Husserl and Lacan on the limits of reflection. *Theory & Psychology* 25 (6), 753-774. An extended and translated version appeared as Feyaerts, J. & Vanheule, S. (2016). Verklaringskloof en subjectiviteit: Natorp, Husserl en Lacan over reflectie. *Tijdschrift voor Filosofie* 78 (3), 583-625.

happening *in* the black box is not explaining what is happening *for* the black box” (p. 12, italics added). According to advocates of this explanatory gap argument, it is thus one thing to try to account for what is going on in our brain – at whatever level of explanation, whether neurobiologically or functionally defined – when we engage in various kinds of cognitive behavior (e.g. memory, perception, attention, ...), and something fundamentally different to try to account for what appears or seems to be going on when we engage in those types of behavior from a first-person point of view.

Hence, what the explanatory gap argument vis-à-vis the current status of cognitive science amounts to can be succinctly formalized as follows:

- P1. Cognitive science purports to give a complete account of the cognitive mind.
- P2. An essential feature of creatures endowed with such a cognitive mind is that they have the capacity for phenomenality: in clear contradistinction with tables or stones, there is something ‘it is like’ to be a cognitive agent.
- P3. Cognitive science does not attempt or (up until now) fails to account for this subjective dimension.
- C. Hence, in its current status, cognitive science is incomplete and found wanting with respect to P1.

It is this remarkable reversal in a sort of strange reading of Kant’s scientific ideal – in which cognitive science does succeed in charting the cognitive/noumenal ‘Ding an sich’, yet doesn’t manage to account for its phenomenal appearance – which has led researchers to argue that a mere functional analysis of the inner workings of the mind does not provide an exhaustive nor adequate explanation as long as the first-person dimension of subjective experience is not taken into consideration as an explicit area of investigation (Flanagan, 1992; Strawson, 1997; Parnas, Sass & Zahavi, 2013). Furthermore, in light of this explanatory gap and the related phenomenological critique, several authors have argued for a return to classical phenomenological approaches of subjectivity in the hope to enhance our understanding of the relation between cognitive processes and their phenomenal manifestations. Still according to these authors, phenomenology, as originated in the work of Husserl and subsequently developed by e.g. Heidegger, Merleau-Ponty and Sartre, could provide us with a refined theory and descriptive approach of subjectivity as the necessary requirement prior to any naturalistic effort of ‘bridging the gap’.

Of course, in its current form, the discussion raises many questions which are for the most part related to some of the assumptions mentioned above – i.e. (1) whether there is indeed an explanatory gap, and if so, (2) whether it should be the goal of cognitive science to bridge this gap; (3) whether phenomenological data, given their subjective nature, are compatible with a scientific approach of mental functioning; (4) ... – or lead to the question whether phenomenology, in the form of Husserl’s transcendental-idealistic philosophy, is open to a naturalization in the first place (see Rinofner-Kreidl, 2005; De Preester, 2006; Zahavi, 2010). Yet, what is less frequently discussed within this Sellarsian

opposition between the defenders of our manifest first-person perspective and the advocates of an all-encompassing scientific ‘view-from-nowhere’ are the difficulties which have traditionally surrounded all efforts of referring to, or of speaking from, this subjective-perspective-supposed-to-know. In that way, the discussion is all too often reduced to the simple opposition between, on the one hand, what was *already known* from the first-person point of view, and on the other, what the ongoing developments in neurocognitive science could adduce to this knowledge qua explanatory account. Although Francisco Varela, one of the originators of this effort to conjoin phenomenology and cognitive science, originally spoke about a “methodology of reciprocal constraints” (1996, p. 343), we can read about this relation in, for example, Gallagher that “[...] if the phenomenology indicates that we are not conscious of X, or that our conscious is not X-like, but the neurological explanation inserts X into the picture, so to speak, then the neurological explanation is not explaining consciousness as it is” (2007, p. 303). In the same breath, Weisberg claims: “If a theory cannot explain why the explanandum appears as it does to common sense – that is if it cannot ‘save the common-sense appearances’ – we do not count it as a successful theory” (2005, p. 6). Differently put: if neurocognitive science does not join or confirm what phenomenology already revealed about our subjective self-appearance, then it is apparently explaining something else. Perhaps we are missing out on something here, but it seems that Varela’s reciprocity is still nowhere to be found.

It is important to point out that this peculiar epistemic constellation in which phenomenology unilaterally dictates which ‘gap’ neurocognitive science is supposed to bridge, immediately follows from a certain conception about the sort of knowledge which is presumed to be at stake in this first-person perspective. According to this intuitive conception, when something appears in a certain way within this subjective sphere (say, that my partner *seems* angry with me for some reason or other), then we might be possibly mistaken about the reality of this appearance (i.e. whether my partner *is* indeed angry with me), but not about the appearance of this appearance itself (i.e. whether it *seems* that my partner *seems* angry with me). Appearing or seeming is, according to this idea and as Descartes already remarked, not recursive: the conceptual content of ‘it seems that my partner is angry’ is not altered in a further iteration of the ‘seems’-operator: i.e. ‘seems [mad]’ is trivially confirmed in ‘seems [seems [mad]]’. ‘Appearances’ or ‘seemings’ possess a certain epistemic attractiveness because they are considered able to put a halt on possible (infinite) regressions, precisely because of the assumption that their reality coincides with their appearance. This was, moreover, one of the important motivations behind Husserl’s effort to study subjective phenomena in their own reality, in so far as we can describe the “appearing as such” (Husserl, 2009, p. 400, our translation) in complete evidence.

However, there are some exceptions to this intuitive conception. A classical difficulty for the phenomenological effort to return to our most immediate, lived subjectivity – regardless of the sort of motivation behind this return: i.e. subjectivity as transcendental explanans of intentional object-givenness in Husserl’s project or rather, like in the contemporary debate, as an empirical explanandum

in the effort to explain consciousness – is related to the way in which we have to account for its presumed self-givenness. That is, is it really that evident to speak with certainty from within this subjective perspective? And in what way does this subjectivity manifest itself? If subjectivity, rather than being a transcendent object we occasionally encounter within our world, is the very perspective from which we approach this object-world, in what way could it then itself be made accessible to phenomenological reflection? Furthermore: what is the nature of this subjective self-givenness if we seem to have to ‘return’ to it every so often? Is the epistemological duality we assume between knower and known compatible with a conception of subjectivity as an a-positional and pre-reflexive self-consciousness? In the current article, we will elucidate how these different questions have been approached in the works of respectively Natorp, Husserl and Lacan, in order to problematize the previous simple opposition between the first- and third-person perspectives.

In the first part, we will present Natorp’s neo-Kantian challenge to a phenomenological approach as a reflective investigation of subjectivity in his *Allgemeine Psychologie nach kritischer Methode* from 1912. According to Natorp’s classical argument, every effort to grasp subjectivity by means of reflection is always a step too late vis-à-vis subjective experience as it is lived before this reflection. This will lead him to deny the possibility of attaining a direct access to subjective experience and to refiguring his method as a *reconstruction* with subjectivity hence considered as an unattainable limit-case.

Next we will explore Husserl’s answer to Natorp’s critique of reflective phenomenology as laid out in his *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy* (Husserl, 1913/1983). Contra Natorp, Husserl will defend the phenomenological method of *reduction* as being based on direct evidence concerning the description of lived experience in its intentional functioning. This method, Husserl claims in 1913, is in no way indirect or reconstructive, but directly based on evidence as obtained in phenomenological reflection.

Finally, we will shed a new light on the debate between Natorp and Husserl concerning the status of phenomenological reflection by relating it to the distinction Lacan draws in his 11<sup>th</sup> seminar on *The Four Fundamental Concepts of Psychoanalysis* between “the subject of enunciation” and “the subject of the utterance” (Lacan, 1964/1973). As will become clear, Lacan’s minimal conceptual apparatus is an alternative way of representing Natorp’s and Husserl’s ongoing debate concerning the relation between subjectivity and objectivity and as such it will prove helpful to tackle our question regarding the possibility of investigating subjectivity by means of phenomenological reflection. In our conclusion, we will lay out some possible consequences of this cross-reading for the contemporary debate on the explanatory gap and the consequent call for a return to subjectivity.

## **2. Natorp’s critical psychology: subjectivation versus objectivation**

It is a commonplace within 19<sup>th</sup> century German intellectual history writing that philosophy, after the exotic wanderings of the speculative systems of post-Kantian idealism, had to cede her claim of scientific validity to the positive sciences. Whereas these scientific disciplines exhibited a steady and continual progression, philosophy appeared to be left behind rather disoriented and aimless. For the so-called Neo-Kantian Marburg School - with Cohen, Natorp and Cassirer as its main representatives – this meant that the only hope for a rehabilitation of philosophy as an autonomous discipline consisted in a return to Kant. Yet, the Marburgers interpreted this return primarily as a call for philosophy to orient itself towards the sciences, rather than of trying to develop its own scientific system independent of the results of the positive sciences. However, what did this “orientation” effectively imply? For the Marburg School, it was only in and through the *transcendental method* that philosophy could confirm and regain its autonomy, since the latter firmly grounded philosophy in the accomplishments and facts of the positive sciences (Cohen’s *Faktum der Wissenschaften*), of which transcendental philosophy aimed to elucidate the a priori conditions and justification (*Rechtsgrund*). By limiting its task to questioning the transcendental conditions of possibility of the sciences, philosophy could avoid both the speculative construction of ideas which are not immanent in the facts of scientific knowledge production as well as the reduction of philosophy to an empirical discipline. More specifically, Marburg Neo-Kantianism set out from the *Faktum* of mathematized natural science, by which was meant science’s progressive determination of an objective order as it is presumed to appear through or in sensible phenomena, via a method of hypothesizing laws that are “valid” (*gültig*) or “hold” (*gelten*) of these phenomena, in order to inquire subsequently into the a priori conditions that factor into this construction of objective reality and which hence could explain how such a constructive determination is attained. As such, the goal of the Marburg logic was to determine, by means of an analytic reconstruction of thinking’s synthetic construction of objectivities, how scientific thinking lawfully generates or synthesizes the unities that are its objects of knowledge. It is consequently *in science*’ progressive objectifications that we can find subjectivity at work, where it is seized as the lawful activity in accordance with which the subsequent hypothesizing of science occurs.

Natorp’s interesting addition to this standard rendition of the transcendental project and at the same time the problematic nexus running through both his *Allgemeine Psychologie* (1912) and his discussion with Husserl<sup>8</sup>, is that not only cognizing activity in science is objectifying; *subjective life as such appears to be objectifying*, whether it leads towards objective knowledge or not. According to

---

<sup>8</sup> Natorp’s *Allgemeine Psychologie* from 1912 is to an important extent a resumption and further elaboration of the problem (i.e. the possibility of a ‘subjective grounding’ of scientific cognition) he already discussed in two earlier publications, namely in the article “Über objektive und subjektive Begründung der Erkenntnis” (1887) and the book *Einleitung in die Psychologie* (1888), which appeared nearly simultaneously. At the time of the publication of the *Ideen*, Husserl had only read and taken account of these works, most explicitly in his *Logische Untersuchungen* (cf. Husserl, 1900-1901/1970a, pp. 91-93). For a historical and bibliographic overview of the relation between Natorp and Husserl, see Luft (2010); see also the recent publication by Dahlstrom (2015) on the same issue. The best and most extensive work on the relation between Husserl and Neo-Kantian thought still remains Kern (1964).



Natorp, this already becomes clear in ordinary subjective enunciations in that they always claim something. As such, every statement, expression or reflexive thought (*Äußerung*) is an objectifying externalization (*Entäußerung*) vis-à-vis the subjective experience which preceded and motivated the subsequent utterance: “Spricht die Seele, so spricht, ach, schon die Seele nicht mehr” (Ibid., p. 98). As Natorp explains, this is the character of “objectifying cognizing (*Erkenntnis*), scientific as well as pre-scientific: [...] to make objects out of appearances (*Erscheinungen*)” (Ibid., p. 193, our translation). But if it is characteristic for all reflection to make objects out of appearances, what consequence will this have on the study of subjectivity as such? If all subjective acts (reflection, judgments, enunciations, ...) in one way or another objectify this pristine (*ursprünglich*) subjectivity, it is both obvious and inevitable that the acts pertaining to the study of subjectivity will be objectifying too. Hence Natorp’s following conclusion:

If one were oneself to try, if it were at all possible, to somehow grasp the content of immediate experience purely as it is in itself – far from every expression, every judgment, every intention – would one then not somehow be forced, nevertheless, to delimit it, to raise it above the mesh of experiencing, be it with the pointing of a finger, with a blink of an eye; would one not be forced to artificially still and interrupt the continuous stream of becoming, which surely is how inner life presents itself, to isolate the individual finding, to fixate it with the isolation in mind, to sterilize it, like the anatomist does with his specimen? But doesn’t one then detach it from the experienced, from the subjective, and doesn’t one then, nevertheless, make it into an object? In the end, one apparently never grasps the subjective, as such, in itself. On the contrary, in order to grasp it scientifically, one is forced to strip it of its subjective character. (Ibid., pp. 102-103)

So, whereas everything else can be made into an object (*Gegenstand*) of reflection, subjectivity as such necessarily eludes this reflection (*nicht sich selber Gegenstand sein*). The moment we start to investigate subjectivity reflectively, we transform it into its very opposite. Therefore, in reflection, the subject of consciousness necessarily makes way for an object of consciousness. Furthermore, according to Natorp, this transcendental clarification with regard to the limits of reflection has important repercussions for the discipline *par excellence* that ought to be devoted to the scientific study of subjectivity, i.e. psychology. To the extent that psychology is a discipline devoted to disclosing scientific facts about subjectivity, it treats its object in the same objectifying way as other scientific disciplines do with regard to their respective objects of study and it will subsequently disregard the ontological difference that sets apart subjectivity from objectivity.<sup>9</sup> In so doing, the particular element that makes up for the subjectivity

---

<sup>9</sup> Natorp’s analysis with regard to this asymmetrical relation between subjectivity and objectivity anticipates, on a transcendental level, similar conclusions that are more recently discussed within analytical philosophy of mind regarding the possibility of reconciling first-person and third-person perspectives on subjectivity by, most notably, Thomas Nagel. Like Natorp, Nagel suggests that there will always be a dimension of subjectivity that will exceed the limits of objectification (see Nagel, 1986, pp. 25-27).

of the subject and that is concomitantly the actual *raison d'être* of psychology as a scientific field of study in the first place – the dynamic, concrete, fluctuating life of subjective experience – is lost. As Natorp adds in a Hegelian way, that which makes the subject a subject is “killed”, mortified under its objectifying representations (Ibid., p.191). All traditional psychology supposedly proceeds in this way: psychology “[...] kills subjectivity in order to dissect it, and believes the life of the soul is on display in the result of the dissection” (Ibid., pp. 102-103).

Thus far Natorp’s macabre conclusion with regard to the possibility of investigating subjectivity has not been very optimistic. Since every effort to investigate subjectivity amounts to an objectifying *Entäußerung* of its point of departure, what other options do we have left? The title of Natorp’s work – *Allgemeine Psychologie nach kritischer Methode* – nevertheless seems to suggest that there is a ‘critical method’, able to overcome the pitfalls of traditional psychology which merely deals with the subjective as a factual state of psychic affairs. Indeed, this is the task Natorp sets out for his ‘transcendental psychology’. It has become clear that the method of this transcendental psychology should allow for a clarification of subjectivity that is not objectifying, one that doesn’t transform our subjective life into an artificial object. Hence, Natorp suggests that we will never be able to do justice to the ontological difference if we operate within an object-oriented methodological monism. Differently put: *ontological difference should be met by methodological difference*. Therefore, whereas, according to a transcendental perspective, objectivity is attained by a *construction* of objective idealities out of the immediate flux of subjective appearances, subjectivity should be investigated by going back from these objectifications to the immediacy of subjective appearances by a method of *reconstruction*.<sup>10</sup> Methodologically speaking, reconstruction should therefore be regarded as the inverse method of objectification, the “turning inside out” (*Umstülpung*) of its constructive effects towards that from which reality is progressively built up in the first place: the concrete and dynamic conscious life of the subject (Ibid., p.20). Such a reconstruction should offer us a kind of genetic-archeological picture of how our direct and pre-reflexive experience ‘would have been like’ before the *rigor mortis* of its objectifying representation set in, this however unavoidably starting from the clues and indications that are at our disposal *post mortem*.<sup>11</sup> Ultimately, however, the dimension of pure subjectivity as the lived immediacy

---

<sup>10</sup> However, as Heidegger rightfully pointed out in response to Natorp’s critique of phenomenological reflection, re-construction is of course itself also a form of objectivating construction and is therefore susceptible to the same criticism Natorp offers towards phenomenology (Heidegger, 1919/1982, 56/57, pp. 104-107). For a further reading of the discussion between Heidegger and Natorp and the former’s methodological defense of hermeneutical intuition to escape this deadlock, see Zahavi (2003).

<sup>11</sup> Formulated that way, Natorp’s somewhat paradoxical critique of traditional psychology is therefore *not* that the latter is only occupied with studying the objectivized echo of our subjective life instead of focusing on its immediate disclosure, since this will inevitably be the case and hence also in Natorp’s transcendental alternative. Natorp’s critique is rather that psychology is itself unaware of this fact and therefore considers the reflexive mirror image to be the pre-reflexive original. Differently put: according to Natorp, psychology operates according to a double negation in which the first negation is characteristic to our own reflexive self-relation and the second ‘negation of the negation’ is specific to psychology as an empirical discipline.

of conscious phenomena remains an unreachable ideal and will hence forever elude our grasp (Ibid., p.233).

### 3. Husserl's response: "the principle of all principles"

One does not have to be an erudite phenomenologist to see that Natorp's methodological reservations concerning the possibility of a direct description of subjective experience through reflection depart significantly from phenomenology's *modus operandi* with regard to its field of study. A common way to define phenomenology is by determining its task as a description of our lived experience as it is given, precisely in the way it is given, in its pre-theoretical immediacy. Indeed, as Husserl famously holds in his *Ideas I*: "the principle of all principles" which grounds phenomenology as an eidetic enterprise – as a pure description of the essential (eidetic) structures of consciousness - is that it builds upon that which is present "in self-giving evidence and intuition" (1913/1983, p. 44). Therefore, Natorp's challenge vis-à-vis phenomenology as a reflective investigation of subjectivity is leveled at nothing other than what Husserl – from the publication of his *Logical Investigations* in 1900/1901 onwards - considered its most basic methodological principle, namely that of *reflexive intuition* which brings phenomena to clear and apodictic *evidence*. As Husserl explains in the section *Critical Excursus. Phenomenology and the Difficulties of "Self-Observation"* in response to similar challenges<sup>12</sup> to his reflective paradigm as those raised by Natorp:

The phenomena of reflection are, in fact, a sphere of pure and possibly perfectly clear data. It is an *eidetic insight*, always attainable because immediate, that starting from the objectively given as objective reflection on the presentive consciousness and its subject is always possible: starting from the perceived, from what is "there 'in person'," a reflection on the perceiving; starting from the remembered, just as it "hovers before us" as remembered, as "having been," a reflection on the remembering; *starting from the statement in the flowing off of its being given, a reflection on the stating*, etc. It is evident that by virtue of its essence – [...] – something such as consciousness and consciousness-content [...] is cognizable by reflection (Ibid., p. 187, italics added).

So for Husserl, the possibility of grasping subjectivity by means of reflection is indispensable for the methodological grounding of phenomenology to its subject matter. It is furthermore essential for, so to speak, properly essential reasons: reasons pertaining to the essence of consciousness itself as revealed in eidetic insight. Let us be precise here with regard to Husserl's more or less Cartesian argumentation: according to Husserl, it is *widersinnig* to deny phenomenology's 'principle of all principles' and thus also to doubt the givenness of subjective processes and their accessibility through reflection, since this

---

<sup>12</sup> More specifically the critique by H.J. Watt in the article "Über die neuern Forschungen in der Gedächtnis- und Associationspsychologie aus dem Jahre 1905" (1907).

doubt itself would again be asserted from the initial givenness of reflective intuition. Doubts concerning the possibility of a reflective investigation of subjectivity thus always-already *presuppose* knowledge of the so-called “reflectionally unmodified mental processes” which is at the same time and erroneously put into question (Ibid., p.185). Consequently, the methodological critique concerning the possibility of reflective intuition culminates according to Husserl in doubting the possibility of claiming anything whatsoever on the basis of reflectionally presentive intuition and is thus essentially, like all forms of genuine skepticism, self-refuting. As Husserl contends:

All genuine skepticism of whatever kind and persuasion is indicated by the essentially necessary countersense that, in its argumentations, it implicitly presupposes as conditions of the possibility of its validity what it denies in its theses. [...] He who also says: I doubt the cognitive signification of reflection, asserts a countersense. For as he declares his doubt, he reflects, and setting down this statement as valid presupposes that reflection actually and without doubt (scil. for the cases present) has the cognitive value doubted, that it does *not* change the relation to something objective, that the reflectionally unmodified mental process does *not* forfeit its essence in the transition to reflection (Ibid., pp. 185-186).

Phenomenology’s methodological procedure of reflective intuition is, in sum, not a consequence of any strategic or methodological argumentation which transcends the limits of ‘immediate intuition’ within conscious experience (e.g. God’s incapacity to mislead the subject), but is for Husserl essentially grounded and given as something obvious within the stringent confines of absolute consciousness itself.<sup>13</sup>

For the time being, however, let us momentarily ‘bracket’ this argumentation and focus instead on Husserl’s further elaboration of the nature of this sort of reflection and what it should allow us to describe. As might have become clear from the foregoing, the foundational principle that grounds phenomenology as a reflective investigation of subjectivity does not entail just any intuition. As Husserl explains: “Of course reflection can be effected by anyone and anyone can bring consciousness within the sphere of his seizing regard” (Ibid., p.114). Rather than simply directing the introspective gaze at ourselves and describing what is or is not presented there, phenomenology, by contrast, wants to disclose “a new region of being never before delimited in its own peculiarity” (Ibid., p.63). This new region of being, which will become the fundamental field of phenomenological investigation, is the field of “pure consciousness”, “the sphere of being of absolute origins” (Ibid., p.129), which is allegedly only seized upon by phenomenological reflection. But now the obvious question suggests itself: if this purportedly new region of being is the field of consciousness, why not turn to the empirical findings psychology

---

<sup>13</sup> It is of course this epistemic appeal on givenness, and more in particular, on consciousness as the transcendental condition of the ‘givenness of the given’ which will become the subject of Sellars’ critique in his famous analysis of “the myth of the given” (see Sellars, 1991, pp. 127-196). For an analysis which focuses on both Husserl’s and Merleau-Ponty’s phenomenology and Sellars’ critique, see Sachs (2012). For a defense of Husserl, see Roy (2003).

discloses about this object of study? In what way does this sphere of absolute consciousness differ from the region that is supposed to be already well known, that of the mind or the psyche, which is extensively dealt with by its own regional science, that of psychology? This is of course an important question for Husserl not least because he was especially concerned that phenomenology would not be confused with any kind of psychology whatsoever. Since Husserl's answer to this question is as well-known as it is contested – especially in the discussion on the alleged (non)sense/(im)possibility of a naturalized phenomenology we briefly considered in our introduction –, we will only sketch out the main points we will need for our present inquiry. The principal reason why, according to Husserl, “pure phenomenology is not psychology” (Ibid., p.xix) has nothing to do with accidental delimitations of its field of study, i.e. in the same way as ‘physics is not psychology’ because it treats another ‘province of reality’, nor is this faithful separation between phenomenology and psychology altered by the seemingly obvious fact that both have to do with consciousness. What, according to Husserl, prevents phenomenology from being reduced to psychology, an empirical science dealing with psychological ‘matters of fact’, is that the latter merely deals with consciousness or subjectivity as *real*, actually existing occurrences belonging to *real* empirical subjects in the spatiotemporal world as “omnitude realitatis” (Ibid., p.xx). As such, psychology operates from within the naturalistic attitude wherein the givenness of the actual empirical reality of its object of study – in *casu* consciousness, psychic reality, behavior, ... – is always-already tacitly assumed and accepted as pre-existing in a world where things just are what they are, independent of the performative consciousness of the empirical psychologist as scientific investigator of Nature. It is mainly because of this attitude that Husserl turns away from what he calls “that philosophical poverty in which, under the fine name of a world view founded on natural science, we are vainly fatiguing ourselves” (Ibid., p.115). On the one hand, psychology is precisely a “philosophical poverty” because it only studies consciousness as a component part of Nature and does not reflect on the conditions of possibility of its givenness. Psychology merely takes this givenness of consciousness as an object of scientific inquiry for granted without effectively inquiring *how this is possible*, without answering the transcendental question of the constitutional conditions of its own object of experience, a question which is nevertheless indispensable for grounding and justifying its scientific activity. Or, to slightly reduce the foregoing formulations, psychology only treats consciousness as an object *in* the world, in the sense of a ‘part-whole’ relation, and not as a subject *for* the world, in the sense of a ‘constituting-constituted’ relation.<sup>14</sup> We are on the other hand “vainly fatiguing ourselves” if the effort to present a

---

<sup>14</sup> To be sure, this constitutional problem is not the same as the Cartesian effort to ‘secure’ the world by means of some sort of miraculous deduction starting from the subjective sphere. After Husserl complained about the fact that many of his contemporaries rejected phenomenology as yet another form of Cartesianism, he wrote the following: “The point is not to secure objectivity but to understand it” (1970b, p. 193). In other words, if the so-called ‘natural attitude’ is indeed the focus of Husserlian phenomenology, it is not because the world or any of its objects are in need of being secured or confirmed in their existence, but rather that the possibility of the certainty of their existence still needs to be *understood*. This understanding of how it is possible that the world and all of its entities are given in our subjective experience is for Husserl identical to the question concerning the *constitution* of objects. For a clear presentation of this crucial point, see Mohanty (1985).

phenomenological clarification of pure consciousness is once again confused with an investigation of Nature since the latter, as intentional correlate of meaning bestowing consciousness within the a priori correlation between ‘constituting subjectivity – constituted objectivity’, depends for its sense as a transcendent objective unity precisely on the constitutional acts of pure consciousness. This does not mean however that something is ‘wrong’ with psychology, nor with the other sciences that operate within this attitude, as long as we understand its scientific activity and its accomplishments as, indeed, always starting from and situated in this attitude. On the contrary, Husserl recognizes in this “uncritical presupposition” one of the pillars that carries the success of those sciences (1925/1997, p.191). Husserl’s point is simply that, when aspiring to understand the “general thesis” itself – the experiential fact of there-being-a-world-for-us – psychology does not have anything to tell us, nor does any other science that studies objects in this given world. To schematize the foregoing argument: if it is (1) the task of transcendental phenomenology to understand the thesis of the naturalistic attitude – this is its explanandum – and if (2) the sciences always already operate from the presupposition and pre-givenness of this explanandum, then it follows (3) that phenomenology should avoid any reference to those sciences because otherwise (4) elements from the explanandum would reappear in the phenomenological elucidation of the explanans. According to Husserl, this would amount to what he calls “the transcendental circle, which consists in presupposing something as beyond question when in fact it is encompassed by the all-inclusiveness of that very question” (Ibid., pp. 249-250).

Thus, like Natorp, Husserl stresses the necessity of a radically different approach to consciousness, one that does not conflate subjectivity and objectivity but is instead able to reckon with this ontological difference on a methodological level. Since from a transcendental phenomenological perspective, objective reality is only possible and comprehensible as an intentional unity within the confines of a transcendently pure consciousness, the phenomenological clarification of subjectivity should accordingly shift from an object-oriented to a subject-oriented investigation of consciousness. This is indeed the role to be played by, in a first movement, the famous phenomenological *epoché* through which the thesis of the natural attitude is ‘bracketed’ or ‘put out of action’ in order to avoid that we would rely on knowledge obtained within this attitude to tackle the question of constitution (i.e. on pains of the ‘transcendental circle’) and to get a clear picture of the phenomenon that we would like to explain, i.e. the world as it appears to me in my experience. Yet, here it is important to clear up a recurrent misunderstanding with regard to what the procedure of epoché precisely accomplishes and what the “putting into brackets” does or does not permit. To the extent that the distinction between epoché and reduction (cf. infra) is made at all, several authors (e.g. Spiegelberg, 1965; Hintikka, 1995) seem to understand the notion of epoché in the sense of an ‘exclusion’, ‘elimination’ or ‘negation’ of the transcendent world in favor of an exclusive focus on immanent consciousness. What is subjected to the epoché is hence forever eliminated from further phenomenological research. As is the case in the Cartesian epoché, the world and its objects are removed from focus in order to arrive at the Archimedean zero-point of immediate self-consciousness. However, in an interesting passage in the *Ideen* Husserl

seems to contradict this idea about the epoché as “elimination” explicitly: “Figuratively speaking, that which is parenthesized is not erased from the phenomenological blackboard but only parenthesized, and thereby provided with an index. As having the latter it is, however, part of the major theme of inquiry” (1913/1983, p. 159). So instead of being definitely banished from phenomenological research, Husserl argues that the ‘parenthesized’ becomes in fact its central focus. Yet, how does this add up with the previous claim that, with the entry of the epoché, the natural attitude and its accomplishments could no longer be used in the phenomenological elucidation of our subjectivity? Differently put: if the epoché appears to eliminate nothing at all, why should it be important? The answer consists in a precise reading of the previous quote in which Husserl claims that the epoché provides an ‘index’ or ‘label’ to the world of the natural attitude. This suggests that what has been subjected to the epoché does not disappear from research, but is nevertheless altered. On the one hand, “everything remains as of old” (Ibid., p. 216), yet on the other hand, our reflexive relation towards this ‘everything’ has changed:

[...] through the epoché a new way of experiencing, of thinking, of theorizing, is opened to the philosopher; here, situated above his own natural being and above the natural world, he loses nothing of their being and objective truths and likewise nothing at all of the spiritual acquisitions of his world-life (...); he simply forbids himself – as a philosopher [...] – to continue the whole natural performance of his world-life (1970b, p. 152).

So the crucial importance of the epoché consists in the way it transforms our natural attitude, yet without eliminating or skeptically denying anything which was given within this attitude. It is because of the reflexive epoché that I become conscious of the natural attitude *as* natural attitude, this precisely in order to emphasize how the world manifests itself within this attitude. Suppose, for example, that I have the perceptual experience of a cup of coffee. In the natural attitude, my interest in the latter is related to the fact that I made coffee and that I want to drink coffee in the morning. I am not primarily interested in the experience of the coffee or how it appears to me. In the phenomenological attitude, on the other hand, I become interested in the perceptual experience as such and how it is part of a broader world-revealing experience. The methodological function of the epoché consists in enabling this type of questioning. Therefore, it becomes clear why it cannot be a part of my phenomenological agenda to change anything about this experience. On the contrary, I have to describe this experience precisely in the way it presents itself. As Sokolowski put it: “We must leave everything as it was, for otherwise we would change the very thing we wish to examine” (2000, p. 190). So, in itself, the epoché does not contribute anything with regard to answering the question of world-constitution. It is only, in Husserl’s words, “the gate of entry through which one must pass in order to be able to discover the new world of pure subjectivity” (1970b, p. 260).

Passing through that door, we have arrived at the second methodological movement, that of the phenomenological or transcendental reduction, through which we are lead back (from the Latin:

*reducere*) from the constituted world – as it manifests itself as object-pole within my intentional experience – to that “place” where the world constitutes itself in its givenness-for-me, i.e. transcendental subjectivity. In this way, the world placed within brackets functions as the transcendental guiding clue (*Leitfaden*) starting from which the regressive<sup>15</sup> constitution-question can be pursued in a reflexive way: “Starting from the experiential world given beforehand as existent and [...] from any experiential world whatever, conceived as given beforehand as existent, we exercised transcendental reduction — that is: we went back to the transcendental ego, who constitutes within himself givenness-beforehand and all modes of subsequent givenness [...]” (1960, p. 136). It is this *Rückfrage* which, according to Husserl, makes it possible to answer the specific phenomenological question, that is, the question with regard to the possibility of world-manifestation. Since the reduction leads us from the intentional objectivity to its subjective point of departure, we are enabled to say something about the nature of subjectivity in its intentional performativity (*Leistung*). In this reduced status, I am no longer an object next to other objects in the world *partes-extra-partes*, but a subject of the world as ‘dative’ of manifestation. In the first case, for example, I am someone who sits behind his table and busy writing an article, after the reduction, however, I view myself as the one to whom or for whom this appears, the one to whom I appear as sitting behind that table. Yet, this does not mean, as Derrida rightly remarks<sup>16</sup>, that there is something like an ontological duplication taking place into a transcendental and an empirical part between which we have to fathom some sort of mysterious relationship. It is the same subject which simultaneously appears as both a worldly entity, a human being, and a transcendental subjectivity, depending on the reflexive position I assume with regard to myself, or more precisely, depending on the nature of my reflection: transcendental vs. empirical. In Husserl’s formulation:

Transcendental subjectivity, which is inquired into in the transcendental problem, and which subjectivity is presupposed in it as an existing basis, is none other than again “I myself” and “we ourselves”; not, however, as found in the natural attitude of everyday or positive science; i.e., apperceived as components of the objectively present world before us, but rather as subjects of conscious life, *in* which this world all that is present – for “us – “makes” itself through certain apperceptions (1968, p. 292).

In sum: according to Husserl, an adequate, reflective description of subjectivity is possible if one refrains from conceiving it as a natural object. Through the double methodological procedures of epoché and

---

<sup>15</sup> Here, however, we should distinguish between the kind of regressive method we find, for example, in Kant, who, according to Husserl, lost himself in “a mythical concept-construction” (1970b, p. 199). The phenomenological, reflexive regression is, on the other hand, a “thoroughly intuitively disclosing method, intuitive in its point of departure and in everything it discloses” (Ibid., pp. 115-116). Undoubtedly, Husserl would have placed Natorp in the same ‘mythical-constructive’ camp.

<sup>16</sup> Cf.: “The difficulty is based on the fact that this doubling of sense must not correspond to any ontological double. For example, and briefly put, my transcendental I is radically different, Husserl explicitly states, from my natural and human I by nothing, by nothing that might be determined by the natural sense of distinction. The (transcendental) I is not an other” (Derrida, 2011, p. 10). For a further elaboration of this relation between transcendental and empirical forms of subjectivity, see Carr (1999).



phenomenological reduction, a radically different perspective on subjectivity is attained, one that is committed to the radical opposition between (constituting) subjectivity and (constituted) objectivity, both of which are related to each other in a transcendental ‘correlational a priori’ and accessible to phenomenological reflection.

#### 4. Natorp and Husserl reconsidered

The foregoing analysis has presented us with two important points as to the relation between this preliminary outline of Husserl’s phenomenology and Natorp’s challenge of a reflective investigation of subjectivity.

Firstly, Husserl fully agrees with Natorp’s critique of psychology as being a discipline ill-suited to treat its subject matter in a methodologically appropriate way. Like Natorp, Husserl critically approaches psychology since it is, and will be, unable to teach us anything about the proper *essence* of subjectivity because of two closely related reasons. On the one hand, psychology remains stuck in the naïveté of the naturalistic attitude and consequently fails to recognize the difference between transcendental subjectivity and mundane empirical subjectivity. On the other hand, and as a direct consequence of the first, it has conceived of its object (the ‘psyche’ or the ‘soul’) as something real, similar to any object of the natural sciences, and hence follows the path of an objectivistic science *more geometrico*. In contrast, Husserl resists any characterization of subjectivity in the terms used for the characterization of physical objects and hence considers its study according to the methods of natural science a category mistake.

Secondly, it is questionable however whether Husserl’s reconfiguring of phenomenology as a transcendental investigation of subjectivity really solves the radical nature of Natorp’s critique. Let us recall that Natorp’s critique is in the first place a *methodological* one, aiming for phenomenology as grounded in *reflective intuition*. Although it may well be the case that phenomenology no longer considers subjectivity along the lines of the natural sciences, that it attains a transcendental outlook on absolute consciousness and that it refrains from naturalizing it through the phenomenological reduction, the question remains if it is consequently able to avert Natorp’s logic according to which ‘every Äusserung is an Entäusserung’, i.e. according to which ‘subjectivity is objectifying as such’. Even more so, to the extent that the description of the by now phenomenologically reduced ‘pure consciousness’ still rests on the evidence-paradigm of reflective intuition, and that the possibility of this description is essential for grounding phenomenology as the transcendental ‘science of all sciences’, the question with regard to the ability of Husserl’s transcendental phenomenology to stand up against Natorp’s challenge retains its force, even, and maybe especially so, within the transcendental domain.<sup>17</sup> As Zahavi explains

---

<sup>17</sup> In his reading of the relation between Natorp and Husserl, Luft appears to assume that Natorp’s critique could simply be dissolved through the introduction of the phenomenological reduction: “Husserl, however, has shown that it is possible to trespass this threshold [i.e. of Natorp’s unobjectifiable subjectivity] when one practices

with respect to the importance for phenomenology as a reflective enterprise of clarifying this problematic within the transcendental domain: “To deny that transcendental subjectivity manifests itself is to deny the possibility of a phenomenological analysis of transcendental subjectivity. And to deny that is to deny the possibility of transcendental phenomenology altogether” (1999, p.51). Indeed, we have seen that the task of a transcendental phenomenology is not merely restricted to a meticulous description of the appearance of the world in all its ontic diversity, but precisely to unearth the appearance qua appearance and its conditions of possibility. When we begin to study the distinguishing features of this appearing, we will notice that it is characterized by an intentional and dyadic structure: appearance is an appearance of something for someone. If we acknowledge this dyadic and intentional structure for the appearance of e.g. tables and chairs, does it then equally hold for the manifestation of transcendental subjectivity itself? To put it differently: can the condition of appearance appear itself? Can that which forms the condition for phenomenal appearance also become a phenomenon in its own turn? Or do we have to reconstruct this dimension, as Natorp suggests, as that “what will have been” starting from the constituted objectivity?

In other words, Natorp’s critique is already situated on this side of Husserl’s critical-transcendental perspective which has as its essential task to question the naiveté of the natural attitude – in which everything that is given is dogmatically accepted as a pre-given datum – as to its possibility. Natorp shares the view that every given ‘Faktum’ should be understood as the accomplishment of a ‘Fieri’, as indeed for Husserl the noematic object-pole of my consciousness can only be comprehended as the intentional realization of my consciousness. Natorp writes in his recension of the *Ideen* that with respect to this issue, Husserl’s phenomenological project went through an important evolution in the period between the publication of his *Logische Untersuchungen* and the *Ideen*. Whereas, according to Natorp, the former still exhibited a naïve, descriptive approach of consciousness as a “faktische Daseinslehre”, Husserl’s ulterior position is more in accordance with his own because here, subjectivity is no longer considered as “a doubled objectivity” (Natorp, 1917-1918, p. 236) or a “second reality” (Ibid., p. 235), but as beyond “every objectivity” (Ibid., p. 225). Furthermore, Natorp also declares his agreement with Husserl’s claim that a reflexive thematization of intentional consciousness should be considered as an “act of the second degree”, since such acts are secondary with regard to the “primary acts” through which a transcendent object is constituted in the first place. Yet, despite these points of

---

the reduction and attains a different perspective from which to analyze subjectivity in its proper nature” (Luft, 2010, p. 77). The reason for this is, according to Luft, related to the fact that Husserl, thanks to the reduction, is able to break with the natural attitude and thereby attains a transcendental perspective (Ibid., p. 73). Yet, in our reading, we tried to make clear that the break with the natural attitude is not the consequence of the reduction, but precisely of the epoché. This means that it is the epoché which alerts us to how the world appears as it appears in our intentional experience. In Natorp’s terms, this is the description of the constituted outcome of transcendental activity, or better yet, of how the world-phenomenon appears as already constituted within a transcendental perspective. Natorp’s question therefore remains entirely unanswered: that is, whether or not I have a direct access to my intentional functioning within this transcendental perspective or whether I have to reconstruct this subjectivity starting from the constituted objectivity, as Natorp suggests. Differently put: the break with the natural attitude is in itself insufficient to ward off Natorp’s critique.

agreement, Natorp nevertheless argues that between Husserl's transcendental phenomenology and his own critical psychology, a "radical distinction" subsists (Ibid., p. 236). This distinction has, in Natorp's view, everything to do with the fact that Husserl's critical questioning of givenness is in fact not critical enough, since "for Husserl the absolutely given pure consciousness [...] is set forth as a self-evident "residue", and to be clarified as such" (Ibid.). In that way, Natorp argues, the naiveté of the natural attitude is simply substituted for a naiveté on a transcendental level<sup>18</sup>: instead of a dogmatic acceptance of the givenness of the empirical world, now we become confronted with a dogmatic acceptance of the givenness of givenness as such, which immediately reveals itself to the intuitive gaze of the phenomenological spectator. According to Natorp, Husserl confuses the way in which absolute consciousness is lived – its pre-reflexive immediacy – and the way in which it can be known and determined. In Natorp's formulation: "the 'immediate' of pure consciousness is not already as such also immediately known [*erkannt*] or knowable" (Ibid., p. 50). Ultimately, this claim with regard to the value of phenomenological reflection in relation to our lived subjectivity has two different implications:

(1) On the one hand, one could claim that phenomenological reflection does not alter the pre-reflexive givenness of the original, not objectified ego-life and is consequently able to grasp subjectivity in its transcendental functioning. Advocates of this view will either suggest that reflection, rather than being a falsifying mirror which transforms whatever it makes appear (to use Gashé's Derridean phrase: reflection as entailing a "tain in the mirror"), should be regarded as a kind of "disclosing intensification" or "consummation" (Merleau-Ponty, 1945/2002, p.207) of the pre-reflexively lived experience or they will appeal to a kind of hierarchization of the reflexive function in terms of a distinction between "pure" or "impure" (Sartre), "radical" and "less radical" (Merleau-Ponty) or "theoretical" and "hermeneutical" (Heidegger) forms of reflection. However, although it could be argued that one should make a distinction between on the one hand, forms of reflection that explicitly reify and objectify their subject matter in the sense of a naturalistic reduction of subjectivity to the world as *prima materia* and, on the other hand, reflection in the sense of the phenomenological epoché, one nevertheless cannot argue that reflection *as such* simply repeats or continues whatever was already given before the reflection. For if this should be the case, why would we engage in this reflective procedure, whose explicit aim was to lift the dogmatic slumber of the naturalistic self-forgetfulness of subjectivity, in the first place?

(2.) Phenomenological reflection, although it might escape a naturalistic objectification of its subject matter, is itself confronted with a redoubling or splitting of the Ego between an anonymously functioning transcendental life and a *reflected subject*. As Asemissen points out, every reflexive thematization of subjectivity is necessarily marked by a *self-fission* or *self-multiplication* of the

---

<sup>18</sup> For the way in which Husserl treats the problem of a 'transcendental naiveté', see Husserl (2002).

subject (1958/1959, p. 262) and, as Zahavi indicates, it is consequently characterized by an “internal division, difference and distance” (1999, p.188). One of the important consequences of this is that reflection, even when purified from its worldly elements, will always and inevitably fall short with regard to subjectivity qua anonymously functioning subject-pole and hence will encounter in itself a structurally inscribed moment of “self-forgetfulness”.<sup>19</sup>

In the following final section of this article we will examine how Lacan handles the question of this structural *Selbstentfremdung* from the perspective of a psychoanalytic theory of subjectivity and, more in particular, how he does so by means of the distinction drawn in his 11<sup>th</sup> Seminar on *The Four Fundamental Concepts of Psychoanalysis* between “the subject of enunciation” and “the subject of the utterance” (Lacan, 1964/1973).

## 5. The Lacanian ‘split subject’

Thus far our discussion of the phenomenological effort to return to subjectivity by means of a transcendently reduced reflective investigation of consciousness has come across a number of themes that can serve as subsequent points of entrance for our elucidation of Lacan’s position with respect to this problem. First, we have seen that both Natorp and Husserl, for closely related reasons, are critical of psychology understood as a specialized enterprise among the numerous established empirical sciences dedicated to the study of its own province of reality, i.e. the human psyche. Second, in spite of their shared rejection of psychology, Natorp and Husserl clearly differ when it comes to the outline of a methodological alternative which is able to meet psychology’s diagnosed shortcomings. Natorp’s ‘critical method’ tries to side-step the subject’s objectification by means of a genetic reconstructive attempt. Ultimately however, an immediate access to this original dimension is deemed structurally unattainable. For Husserl by contrast, phenomenology’s core methodological commitment is strictly linked to the possibility of a reflective thematization of subjectivity, although, as we have seen, the latter

---

<sup>19</sup> One possible imaginary solution to such an egological *folie-à-deux* consists in switching to an egological *ménage-à-trois* in which a third subject emerges to gaze at the original transcendental I. It goes without saying that this necessarily leads to an infinite regression, but this is indeed the phantasmatic solution we find in Eugen Fink’s notorious *VI: Cartesian Meditations* from 1932 in the figure of the “disinterested spectator”. This figure is obliged to become the spectator of the original, constituting consciousness in order to bring the latter to self-consciousness of its own constituting role (see also Bernet, 1989). Is it too farfetched to structurally identify this phenomenological fantasy with the well-known phantasmatic scenario in which we become the spectators of our own conception/origin? The “disinterested spectator” (however, is he really that disinterested?) within the phenomenological reduction will be substituted for the insistent gaze, purified from every empirical determination, which is able to contemplate its own constitution/birth from a safe distance. A rigorous description of this transcendental illusion is found in Zupancic: “The anteriority of the gaze in relation to consciousness is seen as something that could be suspended and synchronized with this consciousness by means of *staging* – in the present or the future, the hypothetical point of the successful encounter of the gaze and consciousness – of their mutual recognition. It is apparent that the ‘original fantasy’ is always the fantasy of origins” (Zupancic, 1996, pp. 47-48). Differently put: the reflexive impossibility of immediate self-apperception determines the appearance of the “disinterested” object-gaze within a phantasmatic staging in which the transcendental self-division is covered up.

should be purified from its naturalistic remainders through the phenomenological epoché. Furthermore, and somewhat at odds with phenomenology's methodological principle, we noted that the phenomenological reduction in its turn does not, so to speak, 'reduce' the egological structure which pertains to every reflective act, namely the reflectively induced self-fission or splitting of the subject between an anonymous transcendental life on the one hand, and a thematically 'objectified ego' as intended within the reflective act on the other.

With regard to our first question, i.e. Lacan's appreciation of psychology considered as an empirical discipline dedicated to the scientific study of the human psyche, it is clear that, to a large extent, Lacan agrees with Natorp's and Husserl's critical reservations towards psychology and, more in particular, with its inadequacy to give a rigorous account of subjectivity. As Lacan contends in his 1960 *Subversion of the Subject and the Dialectic of Desire in the Freudian Unconscious*:

I shall take advantage of your kindness in assuming we agree that a science cannot be conditioned upon empiricism. Secondly, we encounter what has already been constituted, with a scientific label, by the name of psychology. Which I challenge - precisely because, as I will show, the function of the subject, as inaugurated by Freudian experience, disqualifies from the outset what, going by the name "psychology," merely perpetuates an academic framework, no matter how one dresses up its premises (1960/2006, p. 672).

Lacan's critique of what can be called, in line with Husserl, 'psychology's naturalistic self-understanding' is by now a familiar one: no matter how much psychology self-confessedly asserts to rely solely on empirical data and unbiased observation, it nevertheless cannot function without a non-empirical foundation as its operational condition of possibility. For Lacan, psychology, in its attempt to don the clothing of the natural, observational sciences, feebly denies its necessary roots in a set of inadequately thematized assumptions – its own disavowed 'academic framework'. Any psychological school, in order to advance its empirical research, tacitly relies on a series of metaphysical (or, more properly, metapsychological) presuppositions – e.g. a conception of the mind-body relation, an understanding of the subject-object distinction, a particular notion of what constitutes consciousness/experience, and so on. These meta-level positions support the edifice of "what has already been constituted by the name of psychology" (Ibid., p.672), and yet because of its pretension to be grounded strictly in 'the facts themselves' under the banner of its avowed 'empiricism', psychology, according to Lacan, "tends to forget the circuitous path by which it came into being; otherwise stated, it forgets the dimension of truth that psychoanalysis seriously puts to work" (1966/2006, p. 738). Hence, in a quote that could have been easily written by Husserl instead, Lacan concludes: "science, if one looks at it closely, has no memory" (1966/2006, p.738). Moreover, and further pointing towards a principal agreement between Husserl and Lacan in their respective understanding of the relation between psychology and subjectivity, what is precisely forgotten by scientific psychology is, as Lacan contends,

“the function of the subject” (1960/2006, p. 672).<sup>20</sup>

Now Lacan’s unremitting insistence on the function of the subject as being of central importance to his reworking of psychoanalytic theory in his famous *retour à Freud* may come as a surprise for both those who, on the one hand, tend to consider Freud’s notorious ‘discovery’ of the unconscious as sharply opposed and even antithetical to any philosophical account which stresses the importance of the subject (as the much celebrated story goes, after Copernicus and Darwin, Freud delivered the third and most bitter blow to the subject’s narcissistic self-understanding when he showed that ‘the ego is not even master in its own house’) and, on the other hand, those who tend to downplay Lacan as a mere child of his time (e.g. Manfred Frank, 1989; for a different reading, see Hallward & Peden, 2012), i.e. the time of a strict and fierce structuralism that seemed to have done away with any notion of the subject whatsoever and for once and for all assigned it to its proper place: namely, that of an imaginary epiphenomenon of an underlying determinative ‘a-subjective’ structure functioning ‘behind’ or ‘beneath’ or ‘anterior’ to the subject and at the same time rendering the subject the ignorant place of self-deception, blind to the structures of which it was essentially the outcome (Dolar, 1998). But if so, what then to think about Lacan’s call, after his insistence in the beginning of his teaching on a ‘return to Freud’ in order to avoid the misunderstandings of the ego-psychological obfuscation of the Freudian heritage, for a second ‘return’, yet this time *à Descartes* (1946/2006, p. 133)? Does this mean that Lacan eventually realized that Freud had perhaps been a little too hasty in his psychoanalytic burial of the self-conscious subject of German philosophy and consequently subscribed to that project with which we already became familiar through our reading of Natorp and Husserl?

Not quite, as should become clear from Lacan’s take on the possibility to ‘return to subjectivity’ as an alternative for the naturalistic and psychological objectifications, which was the second theme we deduced from the discussion between Natorp and Husserl. In both his 11<sup>th</sup> Seminar (1964/1973) as in the *Ecrits*-article *The Subversion of the Subject and the Dialectic of Desire in the Freudian Unconscious* (1960/2006), Lacan resumes the egological rest-structure that was the outcome of the phenomenological reduction, i.e. the field of tension surrounding the reflectively induced self-fission of the subject, albeit in structuralistic terms, as the difference between the ‘subject of enunciation’ and the ‘subject of the utterance’.<sup>21</sup> What is, according to Lacan, the relation between these two subjects? And are we able to

---

<sup>20</sup> It should be clear from the following discussion that for Lacan, this does not mean that psychoanalysis should therefore provide science with a transcendental ground *à la* Husserl, that science should become conscious of its own implicit assumptions which enable its constitution *qua* science and which would make its necessary transcendental justification possible. In that respect, psychoanalysis is in our view a rather helpless discipline. The only thing psychoanalysis should (and ought to) do, is to think through the ‘function of the subject’ which is both the starting point and result of the scientific world view. This subject is in Lacan’s version purified from the wealth of empirical determination and subsists as the ‘empty place’ forever in tension with this determination, or in the words of Alain Badiou, as “the pure void of its subtraction” (see Badiou, 2006). For an interesting reading of this subject “in times of psychologization”, see Devos (2013).

<sup>21</sup> Lacan resumes a distinction which originated from the work of the French linguist Emile Benveniste who, against the current of the structuralist reading of language as a formal system of oppositional differences, argued for the necessity to take into account the *spoken* dimension of language (see Benveniste, 1966). For Benveniste’s position within structuralism and psychoanalysis, see Dosse (1997).

circumvent, through a yet unknown methodological *tour de force*, the difference between both in order to accomplish the dazzling task set by both Natorp and Husserl of returning to our most pristine, lived subjectivity? According to Lacan, as indeed was the case for Natorp, the answer to this question must be negative, yet for a different reason. Lacan confirms the irreducible difference between the “subject of enunciation”, which tries to represent itself by means of the diachronic chain of unfolding signifiers, and the “subject of the utterance”, which is to be regarded as the objectified sense-result of the linguistic closure or, to put it more technically, “punctuation” of the anticipated signification that was set in motion through the use of signifiers (Lacan, 1964/1973, p. 140). Or, as Lacan later formulates this distinction in his 1973 essay *L’étourdit*, there is an irrevocable difference between the “saying” (*dire*) of enunciation and the “said” (*dit*) of the utterance (Lacan, 1973/2001, pp. 452-453). Lacan terms this process, wherein the subject of enunciation tries to represent itself but at the same time necessarily loses itself within language, the “fading” of the subject (1964/2006, p. 709; 1964/1973, p. 208) and provides it with the following definition: “The subject is this emergence which, just before, as subject, was nothing, but which, having scarcely appeared, solidifies into a signifier” (Ibid., p.199). Thus far, however, we did not do anything more than providing a reformulation of the foregoing neo-Kantian and phenomenological analyses within a Lacanian framework: reflection within and throughout the use of language thereby replaced reflection within the intentional act. However, what separates Lacan’s account of the preceding analyses is to be read in his following reformulation of the temporal and logical dynamic that inheres in the same process of “fading”:

Produced in the locus of the yet-to-be-situated Other, the signifier brings forth a subject from a being that cannot yet speak, but at the cost of freezing him. The ready-to-speak that *was to be* there—in both senses of the French imperfect “*il y avait*” placing the ready-to-speak an instant before (it was there but is no longer), but also an instant after (a few moments more and it would have been there because it could have been there)—disappears, no longer being anything but a signifier (Lacan, 1964/2006, p.713).

As is often the case with Lacan, he uses an almost disorientating terminology to describe the process wherein the subject of enunciation is being rewritten or reflected in the form of its linguistic counterpart. Can we put this a bit clearer? The “ready-to-speak that was to be there” can be taken as roughly equivalent to the original subject of enunciation, whereas the subsequent condition of “freezing”, or in Natorpian terms, ‘mortification’ stands for the objectified subject of the utterance. Important however is Lacan’s additional suggestion that, in this by now familiar process of ongoing ‘corpsification’, there is a fundamental illusion to be aware of: it is not that I, as a fully constituted subject, always-already self-consciously precede or preexist the unfolding of the signifying chain and concomitant process of signifying closure leading to the articulation of the “subject of the utterance”, and that I, as Natorp would say, become betrayed by the objectifying dimension of the *Entäusserung* implied by language use. No, in Lacan’s view, this idea that “there was to have been” a pristine subjectivity that is subsequently

betrayed by its necessary flawed translation into the “subject of the utterance” is an *après-coup* illusion of perspective. Or to put it differently: the ever elusive, nostalgic returning point of full-blown immediacy which allegedly precedes the deplorable *Enttäusserung* of the subject is for Lacan the *dynamic outcome* (i.e. placing it both “an instant before” as “an instant after” the process of reflection) rather than the point of departure, of the fading of the subject.<sup>22</sup> Furthermore, this paradoxical logic according to which the impediment to reflexive identification simultaneously generates that to which it is an impediment, is the defining feature, or perhaps the Husserlian essence, of Lacanian subjectivity as such. As Slavoj Žižek has noted time and again in his instructive cross-reading of Lacanian psychoanalysis and German Idealism, the Lacanian split subject is the constituted correlate of the *failure* of its own representation (1998, p.263). Or to revert to Husserlian terminology: the Lacanian split subject (\$) or subject of the unconscious makes its home within the gap separating the “I that I am on the subject side” and the “I that I am as Object for myself” (Husserl, 1952/1989, p. 253), within the irreducible difference that is operationally inscribed in the very surface of reflexive consciousness itself. This subject is thenceforth no longer, as it has traditionally been conceived, the mysterious outside or negative counterpart of consciousness, residing in some sort of unfathomable depth or irrational container lurking beneath the smooth surface of its transcendental counterpart and secretly pulling the strings, but a *performative* dimension of subjectivity and correlated outcome of consciousness’ inability to close up on itself. Or as Lacan prosaically put it: “Now if, turning the weapon of metonymy against the nostalgia that it serves, I stop myself from seeking any meaning beyond tautology, and if, in the name of “war is war” and “a penny’s a penny,” I resolve to be only what I am, how can I escape here from the obvious fact that I am in this very act?” (1957/2006, p. 430). In this very performative act, where I finally renounce every effort to reflexively determine my own identity by retreating in the tautological statement “I am I”, hence simultaneously acknowledging that the enunciation of the first I is apparently insufficient to secure my own identity, is where the minimal level of temporal negativity concomitant of the Lacanian subject inevitably makes its entrance. A further illustration of this peculiar logic is to be found in Freud’s famous example on negation, i.e. the notorious “You ask who this person in the dream can be. It’s *not* my mother” (Freud, 1925, p. 235). In which case, adds Freud, the question

---

<sup>22</sup> A possible phenomenological rejoinder to such a conception of the dynamic process that in Lacan’s view provides the basic structure of the subject’s self-relating is the following one: if we concede to Lacan’s view that there is indeed an irreducible difference between the subject of enunciation and the subject of the utterance, and furthermore, that this difference is constitutive of Lacanian subjectivity as such insofar as it is marked by a temporal self-relating negativity, how then are we able to explain the experiential phenomenon that, despite this fracture in the heart of subjectivity, I nonetheless remain the same throughout this process of failed translation? To put this differently, what remains to be explained in such a conception of subjectivity where it is considered the dynamic outcome of a reflexive failure is twofold: (i.) the irreducible ‘my-ness’ of subjective experience that is the basic feature of subjectivity within a phenomenological account, and (ii.) the experiential (or other) standard against which reflection is precisely to be considered as a *failed* reflection. This last question is akin to the critique offered by Manfred Frank in his analysis of Lacan in his “What is Neo-Structuralism?” (1989, pp. 279-294). Without going into further detail, we had preferred to see Žižek, in his response to Frank’s critique, actually engage with this crucial issue instead of simply “staring at this line of argumentation” (see Žižek, 1998, p.271).



is settled; we can be sure that it is indeed her. In that case too, this rather remarkable epistemic demarche rests on the gap which emerges between the two levels of subjectivity in and through the negation. On the noematic level of the statement – or again, on the level of how the dream appears to the patient – everything seems to be in order: this is simply how it seems to the patient. Yet the strange thing and at the same time the reason for the Freudian suspicion is the fact that this is being said *at all*. On the level of the enunciation, the utterance “it is *not* my mother” leads to the annoying question: if nobody suggested that it is the mother, why then should it be explicitly negated? In that case, a second message is implicit in the fact that a first explicit message has been offered to the analyst and we become confronted *après coup* with our own, elusive enunciative dimension: divided between *what* we say (subject of the utterance) and the fact *that* we say it (subject of enunciation). Again: this does not mean that for Lacan, the true, original subject - as was the case for Natorp and Husserl - is to be situated in one of the two poles. As Lacan remarks:

The core of our being does not coincide with the ego. That is the point of the analytic experience, and it is around this that our experience is organized [...]. But do you think that we should be content with that, and say – the *I* of the unconscious subject is not *me* [*moi*]? That is not good enough, because nothing [...] implies the inverse. And normally you start thinking that this *I* is the real ego. You think that the ego is nothing but an incomplete, erroneous form of this *I*. In this way, you have accomplished the decentring essential to the Freudian discovery, but you have immediately reduced it (Lacan, 1988, p. 59).

Rather, this means that psychoanalytic subjectivity is, per definition, divided without a center and is able to arise momentarily – as an insubstantial cogito – whenever the discordance of our own speech becomes apparent. In our conclusion, we will discuss which implications such an idea could have for the contemporary discussion on the so-called explanatory gap.

## 6. Conclusion

We started our discussion by revisiting the contemporary debate on the so-called ‘explanatory gap’ between cognitive and naturalistic explanations of the mind and the first-person dimension of subjective experience. As has been rightfully argued by several authors, a necessary requirement for the effort to start bridging this gap is to take this first-person perspective seriously in order to grasp adequately what has to be explained in the first place. It is precisely in this effort to get a firm grasp on the ‘explanandum’, i.e. what it is like to be a subject with a first-person experience, that our discussion of Natorp, Husserl and Lacan has its relevance.

First, we have seen that Natorp points towards a dimension of subjectivity that eludes every effort to grasp it reflexively. Second, Husserlian phenomenology argues for a different approach of subjectivity as not being amenable to an unproblematic process of empirical individuation and

description akin to the measurement of middle-sized, three-dimensional objects in the external world, but as the transcendental condition from where the empirical world is understood in the first place. Furthermore, as we have pointed out in our analysis of Husserl's transcendental phenomenology, besides the explanatory gap which separates naturalistic from phenomenological accounts of subjectivity, there seems to be another 'gap' which inheres within the reflexive first-person account itself. Whereas this gap within the reflexive surface of consciousness is traditionally understood as an *epistemological* stumbling block in our perpetual effort to 'return to subjectivity', we have seen that Lacan treats it as an *ontological* dimension of subjectivity that marks and determines our first-person experience as such. To designate the paradoxical nature of such a subject, Lacan famously coined the term 'split subject', which he, in a punning reformulation of the Cartesian phrasing, characterized as follows: "I am thinking where I am not, therefore I am where I am not thinking" (1957/2006, p. 430). This means above all that the very relation it entertains with itself must be rooted in a fundamental impossibility of coinciding with itself. And although this may sound like an odd definition of subjectivity, certainly when placed in the light of the contemporary portrayal of the naturalization debate (cf. *infra*), it must be remarked that it is also a perfectly classical one: from Hegel to Sartre, passing via Kierkegaard and Heidegger, the subject has long been defined either as an impossible being that is what it is not and is not what it is, or one for whom identity and difference are identical, a paradoxical entity which is at the same time a given and a task, and so on (Maniglier, 2012). Furthermore, we think that such a conception of subjectivity is also consistent with more recent phenomenological readings that emphasize the "internal differentiation" and "immanent alterity" at the very heart of the phenomenon of self-manifestation, at the expense of the usual portrayal of the phenomenological transcendental subject as a self-enclosed entity marked by an unmediated, pure and apodictic self-presence (see Zahavi, 1998).

So, to conclude our discussion, what then are the possible consequences of such a paradoxical definition of the subject as, in the words of Lacan, "caught up in a constituting division" (1965/2006, p. 727) for the contemporary naturalization debate and the explanatory gap-argument we briefly touched upon in our introduction? Although a full exposition of this matter would evidently exceed the limits of this article, we will nevertheless point out what we consider to be two of its most interesting and promising implications in terms of their potential to provide a new perspective on how this debate is currently understood. By this we mean, firstly, that the Lacanian reformulation of subjectivity in terms of a constitutive 'méconnaissance', while at a first glance seeming to undermine our very idea of subjectivity both as it appears to us in our most immediate self-experience and as it is traditionally conceptualized within the phenomenological first-person perspective, i.e., as a form of intimate self-acquaintance or immediate self-presence, actually reinforces the idea of a subjective first-person dimension that is to be taken seriously in its own right above and beyond certain reductionist strategies for which subjective appearances are indeed merely appearances. It could be argued that it is precisely this overly facile philosophical opposition between *reality* and *appearance* that figures predominantly in the contemporary naturalization debate: where the former is the domain of neurocognitive research

in that it tries to give an account of the naturalistic underpinnings of our manifest self-experience, whereas the latter falls entirely within the jurisdiction of phenomenology as a propaedeutic descriptive investigation of ‘the way things appear to me in the way they appear to me’. Depending on philosophical allegiance, this peculiar distribution of the reality/appearance-doublet has been taken to mean that

1. the success of neurocognitive science should be measured against its ability to ‘save’ the appearances as they were previously defined in the phenomenological first-person perspective (see Roy et al., 2000)<sup>23</sup>; or
2. subjectivity can be entirely reduced to the status of “mere seeming” so that the reality of consciousness will turn out to be independent of the subjective appearance of consciousness.<sup>24</sup>

What is nevertheless shared by these two seemingly radically opposite research agendas is precisely the assumption that is overturned in the Lacanian approach of subjectivity, i.e., the idea that subjective appearances are entirely clear for the subject of those appearances or, to use Searle’s formulation, the idea that “where appearance is concerned we cannot make the appearance-reality distinction because the appearance is the reality” (1992, p. 121). When we said earlier that Lacan’s contestation of such an effacement of the reality-appearance distinction within the very field of appearances itself, common to both phenomenological and reductionist approaches, actually strengthens rather than undermines the possibility of formulating an adequate account of subjectivity, this is primarily due to the fact that it can account for, and makes room for, the notion of *subjective truth*, i.e. the way things actually seem to me even if they don’t seem that way to me. Indeed, it is precisely because I can be wrong about myself with respect to my very subjectivity, that the reduction of subjective appearance as mere appearance is warded off in any subsequent research effort to give a naturalistic explanation of subjectivity. This is the original meaning of the psychoanalytic ‘subversion’ of the subject: the point is not that my subjective experience qua experience is regulated by objective and unconscious mechanisms which are entirely unknown to me (say, for example, that my crush on someone is the immediate effect of a neurochemical imbalance in my brain) and as such shouldn’t concern me too much: i.e. whatever the ‘recent developments’ in neurocognitive science could offer me as an explanation of my phenomenal self-

---

<sup>23</sup> As Dan Hutto puts it: “Experience cannot simply be explained away. If one admits that experiences appear to exist [...] then they must be accounted for. We are owed an explanation of what accounts for the fact that experiences appear to be as they are. [...] Thus [...] to say that experiences are mere appearances hence not real does not reduce the materialist’s explanatory burden” (Hutto, 2000, p. 5). See also Hutto’s critique of Dennett in which he makes the same point (Hutto, 1998, p. 330). What all these arguments have in common is that they reduce subjectivity to something which is simply ‘given’. The rightful objection against such arguments is that the appeal to ‘givenness’ as such is not sufficient to justify a claim (i.c. about subjectivity), since whatever is assumed to be ‘given’ could itself be subject to change depending on the developments within neurocognitive research. Cf. Patricia Churchland for this last point (1996, p. 398).

<sup>24</sup> In Dennett’s formulation: “There seems to be phenomenology [...]. But it does not follow from this undeniable, universally attested fact that there really is phenomenology” (1991, p. 336).

experience, no one could ever take this appearance qua appearance away from me. ‘Subversion’ means that this standard opposition is effectively transcended and that, even within this field of appearance, I am no longer ‘master in my own house’<sup>25</sup>: for example, it seems that I am in love and that I am doing everything within my power to make sure that the relation will work out, yet in various ways and to my own surprise, this effort seems to be counteracted by myself. Does this not imply that in that case, we should say that the amorous person doesn’t know how his loved one actually appears to him?

This brings us to our second and last point: if the contemporary naturalization project aims at closing the gap between neurocognitive levels of explanation and the first-person dimension of subjective experience, the question we can ask with regard to this effort becomes the following one: *what must the ontology of our neurocognitive organization be, so that subjects which are decentered from themselves can emerge as an effect of that level of organization?* Although it may seem that such a question is fundamentally at odds with how psychoanalysis today in general responds to – to use the accustomed turn of phrase – ‘the latest advances in brain sciences’, i.e. by proclaiming to be the last safe haven for the subject in its non-reducible singularity against the objectifying clutches of all too greedy neuroscientists (for a recent example of this strategy, see Laurent, 2014). However, contrary to this worn-out theme of strategic sacralization of the subject against the ongoing scientific disenchantment of the world, it could be argued that it was precisely this question which stood at the forefront of Lacan’s structuralistic *retour à Freud*. As has been noted by Maniglier (2012), it was Lacan’s original idea that this paradoxical element always at a distance from itself, viz. the split subject, far from being only accessible through a transcendental philosophical a priori approach, could be approached scientifically by the then currently in vogue ‘structural methods’ in the field of the human sciences. And although space prevents us from fully substantiating this claim, by, for example, pointing out the similarities and divergences between neurocognitive and structuralistic approaches to the question of subjectivity, it is our view that this question remains one of the most fruitful to investigate in the years to come.

---

<sup>25</sup> In short: the neurocognitive unconscious is not the same as the psychoanalytic unconscious.

## References

- Asemissen, H.U. (1958-1959). Egologische Reflexion. *Kant-Studien* 50, 262-272.
- Badiou, A. (2006). *Being and Event*. New York : Continuum.
- Benveniste, E. (1966). *Problèmes de linguistique générale*. Paris: Bibliothèque des sciences humaines.
- Bernet, R. (1989). “Différence ontologique et conscience transcendentale : La réponse de la Sixième Méditation Cartésienne de Fink”, in E. Escoubas & M. Richir (Eds.), *Husserl*. Grenoble : J. Millon, 89-116.
- Carr, D. (1999). *The Paradoxes of Subjectivity*. New York: Oxford University Press.
- Churchland, P. (1996). The Hornswoggle Problem. *The Journal of Consciousness Studies* 3(5-6), 402-408.
- Dahlstrom, D. (2015). “Natorp’s Psychology”, in N. De Warren & A. Staiti (Eds.), *New Approaches to Neo-Kantianism*. Cambridge: Cambridge University Press, 240-260.
- Dennett, D. (1991). *Consciousness Explained*. Boston: Little Brown.
- De Preester, H. (2006). Naturalism and transcendentalism in the naturalization of phenomenology. *New Ideas in Psychology* 24, 41-62.
- Derrida, J. (2011). *Voice and Phenomenon: Introduction to the Problem of the Sign in Husserl’s Phenomenology*. Evanston, Illinois: Northwestern University Press.
- De Vos, J. (2013). *Psychologization and the Subject of Late-Modernity*. London: Palgrave Macmillan.
- Dolar, M. (1998). “Cogito as the Subject of the Unconscious”, in S. Zizek (Ed.), *Cogito and The Unconscious*. Durham, N.C.: Duke University Press, 11-40.
- Dosse, F. (1997). *History of Structuralism, Vol. 2, The Sign Sets: 1967-Present*. Minneapolis: University of Minnesota Press.
- Fink, E. (1932). *VI: Cartesianische Meditationen, Teil I, Die Idee einer transzendentalen Methodenlehre*. Dordrecht: Kluwer Academic Publishers.
- Flanagan, O. (1992). *Consciousness reconsidered*. Cambridge: MIT Press.
- Frank, M. (1989). *What is Neostructuralism?* Minneapolis: University of Minnesota Press.
- Freud, S. (1925). “Negation”. *Standard Edition* 19, 235-239.
- Gallagher, S. (2007). “Neurophilosophy and Neurophenomenology”, in L. Embree & T. Nenon (Eds.), *Phenomenology 2005, Vol. 5, Selected Essays from North America*. Bucharest: Zeta, 293-316.
- Heidegger, M. (1982). *The Basic Problems of Phenomenology*. Bloomington, IN: Indiana University Press.
- Husserl, E. (1960). *Cartesian Meditations: An Introduction to Phenomenology*. The Hague: Martinus Nijhoff.
- Husserl, E. (1970a [1900-1901]). *Logical Investigations*. New York: Humanities Press.
- Husserl, E. (1970b). *The Crisis of the European Sciences and Transcendental Phenomenology: An Introduction to Phenomenological Philosophy*. Evanston: Northwestern University Press.
- Husserl, E. (1983 [1913]). *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy. First Book: General Introduction to a Pure Phenomenology*. Dordrecht: Kluwer.
- Husserl, E. (1989 [1952]). *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy. Second Book: Studies in the Phenomenology of Constitution*. Dordrecht: Kluwer.
- Husserl, E. (1997 [1925]). *Psychological and Transcendental Phenomenology and the Confrontation with Heidegger*. Dordrecht: Kluwer Academic Publishers.
- Husserl, E. (2009 [1893-1918]). *Untersuchungen zur Urteilstheorie. Texte aus dem Nachlass*. Dordrecht: Springer.
- Hutto, D. (1998). An Ideal Solution to the Problem of Consciousness. *The Journal of Consciousness Studies* 5(3), 328-343.
- Hutto, D. (2000). Author’s Précis of *Beyond Physicalism*. *SWIFT Philosophy of Mind Review*.
- Kern, I. (1964). *Husserl und Kant: Eine Untersuchung über Husserls Verhältnis zu Kant und zum Neukantianismus*. Den Haag: Martinus Nijhoff.
- Lacan, J. (1973 [1964]). *The Seminar 1964, Book XI, The Four Fundamental Concepts of Psychoanalysis*. New York: Karnac.
- Lacan, J. (1988 [1954-1955]). *The Seminar of Jacques Lacan, Book II, The Ego in Freud’s Theory and*

- in the Technique of Psychoanalysis*. Cambridge: Cambridge University Press.
- Lacan, J. (2001 [1973]). "L'Étourdit", in J-A. Miller (Ed.), *Autres Ecrits*. Paris : Editions du Seuil, 449-495.
- Lacan, J. (2006 [1946]). "Presentation on Psychological Causality", in J. Lacan & J-A. Miller (Eds.), *Écrits*. New York and London: W.W. Norton, 123-160.
- Lacan, J. (2006 [1957]). "The Instance of the Letter in the Unconscious", in J. Lacan & J-A. Miller (Eds.), *Écrits*. New York and London: W.W. Norton, 412-444.
- Lacan, J. (2006 [1960]). "The Subversion of the Subject and the Dialectic of Desire in the Freudian Unconscious", in J. Lacan & J-A. Miller (Eds.), *Écrits*. New York and London: W.W. Norton, 671-701.
- Lacan, J. (2006 [1964]). "Position of the Unconscious", in J. Lacan & J-A. Miller (Eds.), *Écrits*. New York and London: W.W. Norton, 703-721.
- Lacan, J. (2006 [1965]). "Science and Truth", in J. Lacan & J-A. Miller (Eds.), *Écrits*. New York and London: W.W. Norton, 726-745.
- Laurent, E. (2014). *Lost in Cognition: Psychoanalysis and the Cognitive Sciences*. New York: Karnac.
- Luft, S. (2010). "Reconstruction and Reduction: Natorp and Husserl on Method and the Question of Subjectivity, in R.A. Makkreel & S. Luft (Eds.), *Neo-Kantianism in Contemporary Philosophy*. Bloomington: Indiana University Press, 82-83.
- Maniglier, P. (2012). "Acting Out the Structure", in P. Hallward & K. Peden (Eds.), *Concept and Form. Volume 2: Interviews and Essays on the Cahiers Pour l'Analyse*. London: Verso Books, 25-46.
- Merleau-Ponty, M. (2002 [1945]). *Phenomenology of Perception*. London: Routledge.
- Mohanty, J.N. (1985). *The Possibility of Transcendental Phenomenology*. Dordrecht: Martinus Nijhoff.
- Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review* 83(4), 435-450.
- Nagel, T. (1986). *The View from Nowhere*. New York, Oxford: Oxford University Press.
- Natorp, P. (1887). Über objektive und subjektive Begründung der Erkenntnis. *Philosophische Monatshefte* 23, 257-286.
- Natorp, P. (1888). *Einleitung in die Psychologie nach Kritischer Methode*. Freiburg: J.C.B. Mohr.
- Natorp, P. (1912). *Allgemeine Psychologie nach Kritischer Methode*. Tübingen: J.C.B. Mohr.
- Parnas, J., Sass, L.A. & Zahavi, D. (2013). Rediscovering psychopathology: the epistemology and phenomenology of the psychiatric object. *Schizophrenia Bulletin* 39(2), 270-277.
- Rinofner-Kreidl, S. (2005). The limits of representationalism: a phenomenological critique of Thomas Metzinger's Self-Model theory. *Synthese Philosophica* 40(2), 355-371.
- Roy, J.M., Petitot, J., Pachaud, B. & Varela, F.J. (1999). "Beyond the Gap: An Introduction to Naturalizing Phenomenology, in J. Petitot, F.J. Varela, B. Pachaud & J.M. Roy (Eds.), *Naturalizing Phenomenology*. Stanford (California): Stanford University Press, 1-82.
- Roy, J.M. (2003). Phenomenological Claims and the Myth of the Given. *Canadian Journal of Philosophy* 33, 1-32.
- Sachs, C.B. (2012). "Phenomenology and the Myth of the Given: Sellars, Merleau-Ponty, and some Myths about the Given", lecture Inaugural Meeting of the Wilfrid Sellars Society, Eastern APA.
- Searle, J. (1992). *The Rediscovery of the Mind*. Cambridge: MIT Press.
- Sellars, W. (1991). "Empiricism and the Philosophy of Mind", in *Science, Perception and Reality*. Atascadero: Ridgview Publishing Co., 127-196.
- Strawson, G. (1997). The Self. *Journal of Consciousness Studies* 4(5/6), 405-428.
- Varela, F.J. (1996). Neurophenomenology: A Methodological Remedy for the Hard Problem. *Journal of Consciousness Studies* 3, 330-349.
- Watt, H.J. (1907). Über die neuern Forschungen in der Gedächtnis- und Associationspsychologie aus dem Jahre 1905. *Archiv für die gesamte Psychologie* 11, 1-10.
- Zahavi, D. (1999). *Self-Awareness and Alterity: A Phenomenological Investigation*. Evanston: Northwestern University Press.
- Zahavi, D. (2003). How to investigate subjectivity: Natorp and Heidegger on reflection. *Continental Philosophy Review*, 155-176.
- Zahavi, D. (2010). "Naturalized Phenomenology", in S. Gallagher & D. Schmicking (Eds.), *Handbook*

- of Phenomenology and Cognitive Science*. Dordrecht: Springer, 3-19.
- Zizek, S. (1998). "The Cartesian Subject versus the Cartesian Theater", in S. Zizek (Ed.), *Cogito and The Unconscious*. Durham, N.C.: Duke University Press, 247-274.
- Zupancic, A. (1996). "Philosophers' Blind Man's Buff", in R. Salecl & S. Zizek (Eds.), *Gaze and Voice as Love Objects*. Durham N.C.: Duke University Press, 32-58.





### 3

## Cartesian meditations: Lacan and the question of enunciation<sup>26</sup>

---

### 1. Introduction

In *The Question of Lay Analysis* from 1926, a text intended as an exposition of the peculiarities of psychoanalytic treatment before an imaginary ‘impartial person’, Freud lets his interlocutor make the following observation: “You assume that every neurotic has something oppressing him, some secret. And by getting him to tell you about it you relieve his oppression and do him good. That of course, is the principle of confession, which the Catholic Church has used from time immemorial in order to make secure its dominance over people’s minds” (p. 189). To which Freud responds in order to distinguish the originality of his talking cure from the catholic principle of confession: “In confession the sinner tells what he knows; in analysis the neurotic has to tell *more*” (Ibid.). Instructive proposal, but a fairly ambiguous therapeutic precept, to say the least.

For is Freud suggesting that one should tell things one doesn’t know, but merely pretends that one knows? Say, for example, by telling, in a pretentious anticipation of the psychoanalytic insight, that I actually secretly enjoy what I complain about? On closer inspection, this cannot be the case, since by telling what one doesn’t know *as if* one knows, under the propositional attitude ‘pretense’, one tells precisely what one knew all along: I already know that I am telling more than I actually know, i.e. that I know nothing about my secret enjoyment. As Sartre would say: consciousness of telling what I do not know is already accompanied by a non-thetic consciousness of my pretending that I do not know. So while I can deceive the analyst by telling what I pretend to know but do not know, it seems that I cannot deceive myself with respect to the knowledge of my ignorance. On this account, pretense is necessarily intersubjective: in the absence of an other for whom I can pretend to know, or even pretend to pretend to know, the iteration of pretense collapses into the primitive project of Sartrean ‘bad faith’. Obviously, the situation will not be not altered by subsequently telling that I was merely pretending to know, since this would lead us back the principle of confession.

But perhaps we can get closer to the Freudian injunction by leaving behind our folk-psychological intuitions according to which speech is merely a verbal mirroring of our anterior

---

<sup>26</sup> A translated version of this article will appear in the forthcoming issue of the journal *Psychoanalytische Perspectieven*.

knowledge of an ‘intentional state’. Instead of that idealistic *méconnaissance*, we should turn the tables and recognize that there are no determinate intentional states, no knowledge of my subjective condition, prior to the occurrence of a verbal expression (see Dennett, 1991, chapter 8). Thus considered, expression establishes *nachträglich* what ‘will have been’ the knowledge of my neurotic predicament. But yet again, on closer inspection, this cannot be sufficient, since although the order of the terms got reversed in this materialistic proposal (i.e. from knowledge determining expression to expression determining knowledge), I nevertheless still tell precisely what I know. That is, although I now no longer know before actually telling it, my telling does not outstrip my knowledge. What I say still matches exactly the knowledge of the intentional state I am in, because my speech fully determines any knowledge I might have of myself. Hence, judged according to Freudian standards, I still remain a sinner engaged in confession.

## **2. Enunciation versus enunciated**

From the provisional analysis just offered, it has become clear that in order to be able to maintain the Freudian precept without succumbing to either bad faith or confession, the subjective logic subtending my speech should satisfy at least two minimal conditions:

- (1) I should tell more than I know;
- (2) without knowing that I am about to tell more than I know.

Formulated in Lacanian terms: there should be a difference between the enunciated content (*sujet de l'énoncé*) and the position of enunciation (*sujet de l'énonciation*). More precisely: what is said on the level of the enunciated content should differ from the position of enunciation (i.e. condition 1), *in such a way* that this difference was neither expected nor predictable in advance from the position of enunciation (i.e. condition 2). An example: imagine a modest person posing a question during a lecture prefaced by the remark “my question is probably not that interesting but ...”, and then imagine his reaction when the addressee immediately interrupts to respond that it would then be best to move on to the next question ... By being insulted, Lacan would say, “the subject receives its own message in an inverted form” (1956/2006, p. 30). It was not that he already accounted for his own position of enunciation and was merely pretending to be the modest fool at the level of his statement. No, in this case, the feeling of being insulted is the phenomenological correlate and witness of the truth of its own enunciation, i.e. that of being superior. From this short example and the minimal definition of the unconscious we gave above, some further important consequences follow.

## **3. Unconscious, truth, subjectivity**

*First*, that the unconscious is necessarily performative, since it can only arise in and by an act of speech. This means that there ‘is’ no unconscious, no substantial unconscious reality, independent from the performative (f)act of telling more than one knows. The unconscious is not, to use Husserl’s expression, “a little tag-end of the world” (1995, p. 30), nor of any kind of ‘underworld’. It is made existent or ‘realized’ within the very opposition that is temporally established *in actu* between the performative and constative levels of speech.<sup>27</sup>

*Second*, while being performative, it is nonetheless impossible to perform, in exactly the same way, as it is impossible, for example, to be ‘spontaneously spontaneous’. Yet, despite being impossible to perform, it is at the same time impossible to avoid: according to Lacan (1964/2006, p. 834), the speaking subject constantly announces its elusive dimension on the level of enunciation without even wanting or noticing it. Consequently, one could say that, confronted with the Freudian directive to tell more than one knows, it is always a matter of ‘too little’ and ‘too much’. ‘Too little’ because I necessarily fail to uphold the directive whenever I consciously try or want to: within the field of conscious intention, I always end up with what I knew already. ‘Too much’ because I am constitutively unable to avoid telling more than I know, even in case I try to avoid telling more than I know precisely by telling more than I know. That is, in getting ahead of myself by telling more than I actually know — say for example, in commenting on some absurd obsession by adding the remark ‘that it surely has something to do with my unconscious’ —, I always-already unwittingly indicate a position of enunciation irreducible to this lofty self-accusation, in this case possibly, that I want to be ahead of myself in order to remain ahead of you, anxiously closing off any possible interpretation the analyst might make in my stead. In sum, regardless of my conscious intentions, the ‘I desire’ of enunciation *will* be able to accompany all of my representations, although it is not a position I can willfully adopt.

*Third*, in order to account for this performative fact of telling more than one knows in terms of its putative truth-conditions, a specific psychoanalytic version of the Cartesian-Husserlian procedures of (i) *epoché* and (ii) *reduction* is called for. As Freud put it (1911, p. 225): “one is bound to employ the currency that is in use in the country one is exploring – in our case a neurotic currency”. According to Lacan, this Freudian currency should be conceived along “a line of research which is not characterized by the same style as other sciences. Its domain is that of the truth of the subject” (Lacan, 1988, pp. 20-21). And furthermore, this “quest for truth is not entirely reducible to the objective, and indeed objectifying, research of the normal scientific method. It is a matter of the realization of the truth of the subject, as a specific dimension which must be detached in its originality from the notion of reality” (Ibid.). Let us take the example from above — i.e., the comment ‘it surely has something to do with my unconscious’ — to clarify Lacan’s remarks.

---

<sup>27</sup> Cf. Lacan’s: “Ontically, then, the unconscious is the elusive” (1964/1973 p. 32); “The gap of the unconscious may be said to be pre-ontological [...], it does not lend itself to ontology. It is neither being, nor non-being, but the unrealized” (Ibid., pp. 30-31)

What matters in a psychoanalytic approach of subjective truth in such an utterance is of course not its apparent truth-value with regard to empirical data — in this case, whether or not I am right in claiming that my obsessive idea has something to do with my unconscious. It might be true — and in our example, it is indeed true — but the enunciated content is not taken as a representation of an unconscious reality, nor is its truth-value conceived in terms of a representational correspondence theory of truth. This suspension or bracketing of the objective truth-claim of the enunciated content — in Lacan’s terms “the detachment from the notion of reality” (Ibid.) — is what we would call, in line with the phenomenological neutralization of the naturalistic attitude, a *psychoanalytic epoché*. Similar to Descartes’ famous transition in the Second Meditation from “I am now seeing light, hearing a noise, feeling heat” to “I certainly *seem* to see, to hear, and to be warmed” (Descartes, 1996, p. 19), the psychoanalytic epoché would take the form of a transition from

It surely has something to do with my unconscious.

to: **Said** [It surely has something to do with my unconscious.]

What falls within the scope of such a subtractive ‘said’-operator is an objective judgment regarding my neurotic condition. Its conceptual content is dependent upon, because identical to, the conceptual content of the objective truth-claim, commitment to which the operator then ‘brackets’ and suspends by only focusing on what is said. But contrary to a common reading of the Cartesian and phenomenological theses, however, to suspend objective commitment by an epoché of this sort is not, for psychoanalytic theory, to move from objectivity to subjectivity. Or more precisely: the subjectivity at which one arrives through a suspension of objectivity is not exhaustive for the kind of subjectivity that is at stake in psychoanalysis. Surely, by disregarding any possible relation the enunciated content could have to the reality of which it speaks, we move from an objective truth-claim concerning the nature of my obsession ‘as it is in itself’ to a corresponding trivial truth-claim, i.e. that it *seems* or *appears* to me that ‘it surely has something to do with my unconscious’. So on this account, for every objective truth claim, in which I am invariably fallible — i.e., whether or not it has something to do with my unconscious —, there would be a corresponding subjective truth claim, in which I am infallible, a truth claim fulfilled by the sheer fact that I claim so — i.e., that it seems to me that it has something to do with my unconscious. Yet again, and as we know from our previous analysis, this Cartesian indubitability would not be able to stand the Freudian-Lacanian Turing test for subjective truth. While it is indeed not a matter anymore of the “objective research of the scientific method” (cf. *supra*), my telling that ‘it surely has something

to do with my unconscious' because it appears or seems to me that way is still not a realization of the truth of the subject in the psychoanalytic sense.<sup>28</sup>

What is needed to account for this specific psychoanalytic dimension is a second procedure we described above as *reduction*. What counts in this reduction of the enunciated content is not its representational adequacy with regard to the objective nature of my symptom, nor the subjective nature of how it appears to me, but the relation to its *condition of production*. It tries to answer the question in what way the content of my enunciation is related to its position of enunciation. In Lacan's elusive phrasing: "The point is not to know whether I speak of myself in a way that conforms to what I am, but rather to know whether, when I speak of myself, I am the same as the self of whom I speak" (1957/2006, p. 430). Applied to our example: the point is not to know whether the proposition 'this surely has something to do with my unconscious' conforms to what I am, but rather to know, when saying this, in what way the act of saying is related to what is said. Like the method of reduction in Husserlian phenomenology, a (transcendental) correlation is presupposed between what is said about myself — the 'noematic' enunciated subject — and the one who is saying this — the 'noetic' subject of enunciation, but unlike the former, the psychoanalytic correlation is always a *failed* correlation. Not in some Schillerian-existential sense in which words always fail to articulate the pristine kernel of my being — cf. *Spricht die Seele, so spricht ach! schon die Seele nicht mehr* —, but in a much more subversive and precise sense: words *always* succeed in articulating the real of my desire, however, *never* in the right place. Again, in our example: while seeming to talk about myself by commenting on my obsession, I did in fact talk about myself in showing that I — not you — will be the one who is able to disclose the truth of my obsession. This then is the proper form of the psychoanalytic correlation between the subject of enunciation and its enunciated double: I did not speak about myself where I thought I was speaking; I speak about myself where I did not think I was speaking. And together with this formal delineation of the disjunctive logic subtending the speaking subject in psychoanalysis, we have the precise locus of its 'truth' as strictly co-determined by this form: I always speak the truth about myself, whatever that truth may be, where I did not think I was speaking.

*Fourth* and finally, what marks the difference between the representational *knowledge* of my obsessive nature put forth in the proposition 'this surely has something to do with my unconscious' and the obsessive *truth* expressed in my act of saying it, i.e. before the analyst got the chance? Did I say more than I knew if this 'more' turns out to be exactly what I said all along? Indeed, in both cases, so it seems, I told the truth, apparently disconfirming our previous disjunctive definition of subjective truth. The only possible way to resolve this apparent contradiction and to prevent collapsing the difference between knowledge and truth in a redundant tautology is by regarding the first representational proposition as an illusion in the precise Freudian sense. Freud writes:

---

<sup>28</sup> For more on the relation between lacanian psychoanalysis and transcendental phenomenology on the question of 'appearance', see Feyaerts & Vanheule (2015).

An illusion is not the same thing as an error; nor is it necessarily an error. (...) What is characteristic of illusions is that they are derived from human wishes. (...) Illusions need not necessarily be false — that is to say, unrealizable or in contradiction to reality. (...) Thus we call a belief an illusion when a wish fulfillment is a prominent factor in its motivation, and in doing so we disregard its relations to reality, just as the illusion itself sets no store by verification (1927, pp. 30-31).

Like in Freud's example (1927) of a middle-class girl who thinks that she will marry a prince — an idea which, however unlikely, may turn out to be true — the proposition 'this surely has something to do with my unconscious', despite being true, still remains an illusion because it is rooted in my obsessive desire to obliterate the position of the analyst. That is, while seeming to refer to myself as an empirical object-in-the-world from a pure and uncontaminated transcendental point of enunciation — as in Eugen Fink's figure of the "unbeteiligte und uninteressierte Zuschauer" (1995) —, I overlook the fact that this transcendental speech-act *is already obsessive*. Consequently, one completely misunderstands this illusionary (Freud) or imaginary (Lacan) idea with regard to my obsession if one confirms or criticizes it for its (in)adequacy at the constative level. At this level, the idea is indeed adequate to its object, but what counts is the fact that, at the performative level, this is not its object at all. At this level, it serves a different function than that of merely reporting knowledge about myself and, furthermore, it does not cease to fulfill this function when I happen to speak the truth. As in Spinoza's theory of imagination, it is only error, but not (my obsessive) desire, that is removed by the emergence of knowledge:

No positive quality possessed by a false idea is removed by the presence of what is true, in virtue of its being true (Spinoza, 1955, p. 191).

The disjunctive correlation between representational knowledge and performative truth in psychoanalytic theory is therefore essentially marked by what the classical psychiatrist Eugen Bleuler (1950) aptly described as a "double orientation" or "double bookkeeping".<sup>29</sup> That is, on the constative level, I can know very well that my obsession is absurd, that it "surely has something to do with my unconscious", and maintain an appropriate distance towards the obsessive absurdity. Like in Sartre's

---

<sup>29</sup> The concept of "double bookkeeping" was originally used by Eugen Bleuler to designate the peculiar "double" attitude schizophrenic patients maintain towards their delusions and hallucinations. On the one hand, they seem to be profoundly occupied with their fantastic delusions, holding them with an unshakeable degree of certainty irrespective of clear empirical proof to the contrary. On the other hand, despite the indubitable nature of their delusional experience, they nevertheless do not *act* as if they mistook their delusions for reality. Bleuler writes: "Kings and Emperors, Popes and Redeemers engage, for the most part, in quite banal work, provided they have any energy at all for activity. This is true not only of patients in institutions, but also of those who are completely free. None of our generals has ever attempted to act in accordance with his imaginary rank and station" (1950, p. 129). Interestingly, what we have here in Bleuler's account of the "double orientation" in schizophrenia seems to be a precise inversion of the disjunctive relation between knowledge and truth in our own account. Whereas in Bleuler's description schizophrenics do not act on beliefs they nevertheless hold to be most certain, in our case we precisely do act on those beliefs we do not really believe.

reformulation of the Cartesian cogito<sup>30</sup>, there is a minimal and ideal distance — ‘un néant’ — between the reflecting I which says “it surely has something to do with my unconscious” and the reflected I of which it is said that it is obsessive. It is this reflexive negativity within the immanence of my own self-relation that allows me to negate any supposed coincidence of myself with the obsession. As Sartre put it: “To believe is to know that one believes, and to know that one believes is no longer to believe” (1958, p. 69). However, if on this level to believe is effectively no longer to believe, that is to say, if the obsession is an illusion by which I don’t get duped and which I can maintain as an objective semblance, at the level of my performative position, I nevertheless *act* as if I they did not know it. At its most basic level, Bleulerian “double bookkeeping” in the disjunctive correlation between knowledge and truth entails that I do not really have to believe the obsessive belief in order for the belief to be effective. Put yet another way: believing is indeed no longer to believe, *mais ça n’empêche pas d’exister*.

The concept of imaginary misrecognition therefore does not refer to the possible sense or nonsense of what I claim at the level of my statement, it is not a measure of the difference between what I say and what is effectively the case – what I misrecognize, what I effectively overlook, is the illusion that is already on *this* side of reality, the reality which in this case structures my performative position as such. ‘Subject of the unconscious’ designates in the end that figure who is incapable to take itself into account in advance. Every saying is for such a subject an immediate showing of how I am reflexively inscribed towards what I say. For that very reason, the Lacanian subject is not to be considered as something to which I have to return in authentic way by, for example, confessing what I feel or think or by being faithful to the private depths of my soul – ‘subject’ is for Lacan that which will return all by itself.

---

<sup>30</sup> Cf. Sartre: “Now, my reflecting consciousness does not take itself for an object when I effect the Cogito. What it affirms concerns the reflected consciousness. Insofar as my reflecting consciousness is consciousness of itself, it is non-positional consciousness. It becomes positional only by directing itself upon the reflected consciousness which itself was not a positional consciousness of itself before being reflected. Thus the consciousness which says I Think is precisely not the consciousness which thinks” (2004, p. 44).

## References

- Bleuler, E. (1950). *Dementia Praecox or the Group of Schizophrenias*. New York, NY: International University Press.
- Descartes, R. (1996). *Meditations on First Philosophy*. Cambridge: Cambridge University Press.
- Feyaerts, J. & Vanheule, S. (2015). How to return to subjectivity: Natorp, Husserl, and Lacan on the limits of reflection. *Theory & Psychology* 25 (6), 753-774.
- Fink, E. (1995). *Sixth Cartesian Meditation: The Idea of a Transcendental Theory of Method*. Bloomington: Indiana University Press.
- Freud, S. (1911). Formulations on the Two Principles of Mental Functioning. *Standard Edition* 12, 218-226.
- Freud, S. (1927). The Future of an Illusion. *Standard Edition* 21, 5-56.
- Husserl, E. (1995). *Cartesian meditations: An introduction to phenomenology*. Dordrecht: Kluwer Academic Publishers.
- Lacan, J. (1973 [1964]). *The Seminar 1964, Book XI, The Four Fundamental Concepts of Psychoanalysis*. New York: Karnac.
- Lacan, J. (1988). *The Seminar 1954-1955, Book II, The Ego in Freud's Theory and in the Technique of Psychoanalysis*. Cambridge: Cambridge University Press.
- Lacan, J. (2006 [1956]). "Seminar on "The Purloined Letter", in J. Lacan & J-A. Miller (Eds.), *Ecrits*. New York and London: W.W. Norton, 6-50.
- Lacan, J. (2006 [1957]). "The Instance of the Letter in the Unconscious", in J. Lacan & J-A. Miller (Eds.), *Ecrits*. New York and London: W.W. Norton, 412-444.
- Lacan, J. (2006 [1964]). "Position of the Unconscious", in J. Lacan & J-A. Miller (Eds.), *Ecrits*. New York and London: W.W. Norton, 703-721.
- Sartre, J.P. (1958). *Being and Nothingness: An Essay on Phenomenological Ontology*. London: Methuen.
- Sartre, J.P. (2004). *The Transcendence of the Ego: A sketch for a phenomenological description*. London and New York: Routledge.
- Spinoza, B. (1955). *On the Improvement of the Understanding. The Ethics. Correspondence*. New York: Dover Publications.





## 4

# He or It (the Thing) thinks, not I Self-monitoring and verbal hallucinations<sup>31</sup>

---

### 1. Introduction

During the last twenty years, different theoretical models have been proposed, often inspired by developments in the neurocognitive sciences, to explain the phenomenon of verbal hallucinations (VHs). The most influential amongst those is undoubtedly the neurocognitive model which has been proposed by Christopher Frith (1991). In Frith's model, VHs are conceived in terms of a misidentification of one's own thoughts with those of another (other human beings, God, ...) or an external instance (machines, radiations, ...).<sup>32</sup> Frith's explanation more specifically consists in a neurocognitive modeling of the mechanisms that are potentially involved in the temporal process of 'self-monitoring' of our own thoughts. An adequate process of self-monitoring is thought to result in the phenomenological experience of being the active initiator or author of my own thoughts – commonly described in the literature as a 'sense of agency'. By contrast, the phenomenology of VHs is characterized by the absence of this agency: schizophrenic patients have the experience as if someone or something is thinking or talking instead of them. As a consequence, mental life becomes progressively marked by a feeling of radical self-alienation – as one of our patients conspicuously put it: 'My thoughts are no longer my thoughts. It's like putting on your coat: it's my coat, but not me'. By way of the latter example, we have arrived at a second important notion in Frith's model, i.e. 'sense of ownership'. That is, although it is indeed no longer 'I' who actively thinks, but rather is passively thought or spoken to, I am still to be

---

<sup>31</sup> A translated version of this article will appear in the forthcoming issue of *Tijdschrift voor Filosofie* as Feyaerts, J. & Vanheule, S. (2017). Hij of het (ding) denkt, niet ik: Frith over verbale hallucinaties.

<sup>32</sup> The idea according to which VHs would be the outcome of a misidentification of self-generated thoughts or inner speech is, of course, not unique to Frith's model, but shared by almost all contemporary theories of VHs. As Bentall remarks: "there is one fundamental assumption that all the theories [viz. of VHs] have in common: that hallucinators mistake their own internal, mental, or private events for external, publicly observable events" (1990, p. 88). Yet, although this idea excludes certain types of explanations (as, for example, empiristic perception-models, cf. infra), it still leaves a lot of issues untouched. These can be heuristically divided into two separate questions: (1) what has been misidentified (e.g. thoughts, memories, speech, ...)? (2) how is this misidentification itself to be conceived (top down vs. bottom up, cf. infra). For a recent overview, see McCarthy-Jones (2012).

considered as the one who experiences this or to whom this alienating phenomenon appears. That is to say, although I am no longer ‘the unmoved mover’ of the thoughts I experience – these are being initiated by someone else -, I am still ‘the moved mover’. Patients indeed complain that – to use Freud’s expression – they are no longer ‘master in their own house’, but all of this still happens in their ‘own house’. In terms of the phenomenological concepts we introduced: although the sense of agency is disrupted, the sense of ownership is preserved.

In order to explain this particular phenomenology of VHs, Frith appeals to an extrapolation of the neurocognitive mechanism that are thought to be involved in the coordination of motor behavior. In the same way as my normal ‘stream of consciousness’ is characterized by a sense of agency, motor behavior is usually accompanied by the experience of being the cause or author of the movements that are being initiated. For example, I raise my hand because I have the intention to grab something. In this action, I am the one who raises my hand and I have the idea that I could have just as well refrained from doing so. However, in case someone else grabs my hand, my sense of agency is lost. Although it is still my hand that is being moved – that is, my sense of ownership endures -, in this situation, the two phenomenological variables come apart. Obviously, the difference between the movement condition and VHs is of course that in the latter situation, no one is actually present in reality: it only *seems* that way for the hallucinating person. Nonetheless, the potential explanatory power of Frith’s model should be clear: if we would be able to identify the mechanisms that are responsible for the generation of our sense of agency in the case of movement, then *eo ipso* we have identified possible candidates for explaining the characteristic misidentification in the case of VHs. This is one of the main reasons why, for example, the philosopher John Campbell (1999) considers Frith’s model to be one of the most promising and elegant proposals for the explanation of VHs in contemporary psychiatric research.

Yet, although Frith’s theory indeed seems to contain all the features of a good scientific explanation of VHs – i.e. empirical evidence<sup>33</sup>, conceptual parsimony<sup>34</sup>, consistency with other scientific explanations of human behavior<sup>35</sup>, ... -, our goal in the present article is to offer a critical reading of this model of self-monitoring. We will proceed in the following way: in section 1 we will offer an extensive account of the strengths and weaknesses of Frith’s model for the explanation of VHs. We will try to show that the latter are due to both (2) an inadequate idea about subjectivity and (3) an inadequate

---

<sup>33</sup> For a series of experimental studies which seem to confirm Frith’s self-monitoring model (the list is not meant to be exhaustive), see Frith & Done (1989); Malenka, Angel, Hampton & Berger (1988); Frith & Done (1988); Jeannerod (1999).

<sup>34</sup> Some authors, however, do not concur with this point and argue that a meta-representational mechanism to generate a sense of agency in the case of self-consciousness of the intentionality of thoughts is in fact superfluous, since, by itself, the idea of consciousness already implies self-consciousness (see Gallagher, 2007). For a similar critique as to the redundancy of meta-representation in order to generate a sense of agency, this time offered from a neurocognitive perspective, see Jeannerod (1994) and Proust (2000).

<sup>35</sup> Frith’s self-monitoring mechanism is consistent with fundamental neurocognitive explanations of, amongst others, motoric, perceptual and linguistic behavior. See Sperry (1950) and Von Holst & Mittelstaedt (1950).

description of hallucinatory experience. In the last part (4), we will delineate, on the basis of (2) and (3), some requirements which should be met by a neurocognitive explanation of VHs.

## 2. Thinking as movement

As mentioned above, Frith's neurocognitive explanation of some of the most distinctive symptoms of schizophrenia – i.e. VHs, delusions of control, thought insertion, ... - departs from models which proved to be successful in explaining the control and coordination of motor behavior. The starting point for these theories - which were developed in the 1950s by von Holst & Mittelstaedt (1950) and Sperry (1950) - is the rather trivial observation that, as von Helmholtz (1866) already remarked, the world doesn't seem to move when we move our eyes. In order to understand this elementary property of our perceptual phenomenology, these authors introduced the so-called 'comparator-model'. In contrast to classical reflex theories, the comparator-model allows for an explanation and prediction of the fact that a different behavioral response will be elicited by one and the same stimulus, depending on whether that stimulus is generated *externally* (through a movement in the external world) or *internally* (through a movement of the organism). A fly, for example, which has been placed in a vertically striped cylinder, will turn if the cylinder turns around it. The movement of the fly is such that a change in the visual information (the alternation of the vertical stripes) will be neutralized by its concordant positioning. In that way, the fly can maintain and stabilize its visual field. If, however, the fly starts to move itself while the cylinder is hold stable, the reflexive response will not occur, despite the fact that the change in visual information is the same as in the first situation. Therefore, the fact that optical-motoric reflex behavior is dependent on the cause and localization of the change in stimuli (external vs. internal) cannot be explained by a simple reflex-model that doesn't bring the attribution of causality of change in stimuli into account.

The so-called *Reafferenzprinzip* of von Holst and Mittelstaedt (1950) holds that a behavioral response, as in the example of the fly, is not merely determined by changes in afferent input (e.g. the change in the retinal image) as would be the case in a classical reflex-model, but due to a comparison between efferent and afferent signals made by a hypothetical 'comparator-mechanism'. That is, when an intentional movement is performed, not only will there be a transmission of efferent signals to the muscles in order to initiate the movement, but also an 'efference copy' – a copy of the motor intention – will be stored in the sensory-motor system. The function of this efference copy is to neutralize and exclude the so-called 'reafferent signals' – i.e. the efference caused by our own movement – from subsequent perceptual and higher cognitive processing. In that way, a distinction can be made between the changes in stimuli which are due to our own movement and those which are due to an external cause. So, returning to the remark by von Helmholtz, the reason why the world does not seem to move when we move our eyes is because the efference copy of our intentional eye-movement neutralizes the reafferent change in visual perception. However, if, for example, we would press our eye with a finger,

then it would seem that the world really makes a jump, for the efference copy of the finger movement would not be used to correct the afferent visual input.

So, more generally, the comparator model works on account of the rather straightforward idea that an efference copy is used to make predictions about the sensory feedback that will occur as a consequence of our own movement. This allows for a distinction between the perceptual consequences which are the effect of our own movement from those which are externally produced, i.e. the attribution of causality mentioned above. Moreover, the model can be easily expanded in order to explain all forms of intentional movement. As such, it will comprise the following steps:

1. An intention to move is formed;
2. An efferent instruction to move will be send of which a copy will be stored in forward models;
3. These forward models will make a prediction about the sensory feedback that will be due to the movement, *before* the actual movement has been performed;
4. The prediction will be compared with what actually occurred, i.e. the actual sensory feedback.

Currently, different versions of this comparator-model have been proposed (Proust, 2006; Pacherie, 2007; Synofzik & Vosgerau, 2008; Frith, 2012) because of the growing awareness that it needs some refinement and complementary steps in order to provide a sufficient explanation for motor control. Frith and colleagues (2000) therefore introduced a second comparator-component which is also part of the forward pre-motor system and provides predictive information for the coordination and control of motor movement, this however *before* the actual performance of the movement and *before* the actual sensory feedback. One of the problems with the previously discussed, simpler model is that I would need to wait until the actual performance of my movement and the accompanying sensory feedback have occurred in order to adjust my movement in case of possible deviations. The second comparator-component ensures that the movement can be adjusted *before* the actual feedback by comparing the goal of the movement, i.e. the desired outcome, on the one hand, and the prediction of the sensory feedback related to the motor intention of the efference copy, on the other hand. In that way, if a discrepancy would occur between the movement goal and the motor intention, this could be corrected and adjusted in a quicker and more automatic way, that is, before the sensory feedback.

Yet, what is the relevance of this excursion into the neurophysiology of motor control and perception for the effort to understand schizophrenic symptomatology, and more in particular, for the distinction between the sense of agency and the sense of ownership we set out in our introduction? In several experimental studies, Frith et al. (Frith & Done, 1988; Blakemore, Wolpert & Frith, 2000) showed that patients diagnosed with schizophrenia experienced difficulties with the forward monitoring of movement, but not with the motor control based on sensory feedback. Non-schizophrenic control subjects who were instructed to follow a target on a computer screen by means of a manually controlled joystick, were able to rely on either the sensory feedback related to the movement of their hand, or the

more automatic forward pre-motor monitoring of goals and motor intentions. When the visual feedback of the movement was unavailable, non-schizophrenic subjects were able to correct possible deviations with regard to the target in a quicker and more fluent way. Schizophrenic patients, by contrast, were unable to control their own motor intentions on this level of forward monitoring. Like the control subjects, they corrected deviations when visual feedback was available, but not when the latter was absent.

It is important to see how this distinction between pre-motor intentional self-monitoring and monitoring on the basis of sensory feedback is related to the distinction between the subjective experience of ownership and agency and the specific way in which this is disrupted in, for example, delusions of control. In the latter pathology, the intentional agency for movement is being denied and subsequently ascribed to something or someone else. As, for example, in Frith's following clinical illustration of a patient who claims that it is not he, but something else which makes him speak: "The force moved my lips. I began to speak. The words were made for me" (1991, p. 66). Frith explains this loss of agency in terms of a disruption of the efference copy for the motor-verbal intention in the self-monitoring of speech. Given our previous sketch of Frith's self-monitoring model, this is readily understood: if no motor intention has been registered in the efference copy, then the only information which is available for the system is the sensory feedback – in this case 'there is speech'. The sensory feedback still provides for a sense of ownership – it's my lips which move -, but since no intention to speak is available in the efference-copy, the sense of agency is lost – 'speaking is going on, but I am not speaking'. The similarity to our previous example of raising one's hand is clear: in that case too, no motor intention has been registered which would be necessary for the monitoring of the movement and the experience of agency. Only the sensory feedback is available, which stands for a sense of ownership.

In sum, Frith's neurocognitive model for explaining the loss of agency in various experiences of passivity in schizophrenia holds that a comparator-mechanism, which is responsible for the comparison between movement intentions and sensory feedback, is disrupted. However, the important question is whether it is possible to extend this movement model so as to include the cognitive domain? Differently put: is thinking a form of movement which is controlled by the same mechanisms as in the case of motor control? Should we therefore rephrase the Cartesian formula in terms of a 'I think therefore I move therefore I am'? Frith and colleagues believe it should.<sup>36</sup> The same model of self-monitoring is employed in order to understand the intentionality of thinking and, by extension, to explain schizophrenic symptoms like verbal hallucinations and thought insertion:

---

<sup>36</sup> In this, they follow the illustrious footsteps of, for example, Freud, who often made use of sensory-motor scheme's in drawing up his topography of the psychic system, and Hughlings Jackson. The latter formulated his position as follows: "The nervous states concomitant (correlative) with psychical states are, according to the doctrine of evolution, sensori-motor. The highest centres (popularly the 'organ of mind', 'mental centres', etc.) are, according to this doctrine, only the most complex, etc. and latest developed of a series of centres, every one of which represents impressions or movements, or both" (Jackson, 1958, p. 367), and elsewhere: "It being quite certain that the lower centres are sensori-motor, it is surely a legitimate hypothesis that the highest are so too" (Ibid.).

Thinking, like all our actions, is normally accompanied by a sense of effort and deliberate choice as we move from one thought to the next. If we found ourselves thinking without any awareness of the sense of effort that reflects central monitoring, we might well experience these thoughts as alien and, thus, being inserted into our minds. Similarly, actions would appear to be determined by external forces if there was no awareness of the intention to act. Likewise, if we could not distinguish between events caused by our own actions and those of external origin, then we might attribute events caused by our own actions to external events (or vice versa). One manifestation of this effect would be auditory hallucinations. The patient hears a voice and does not recognize it as their own (Frith, 1991, pp. 81-82).

Let's unpack Frith's basic idea in terms of the different steps we identified earlier in the case of intentional movement:

1. An intention to think is formed;
2. An efferent instruction to think will be sent of which a copy will be stored in forward models;
3. These forward models will make a prediction about the sensory feedback which will be the result of thinking, that is, 'what will be thought', before the thinking is effectively performed;
4. The intention to think will be compared with what has effectively taken place, 'what has been thought'.

As in the case of movement, the normal thinking process is, according to Frith, characterized by a sense of agency and a sense of ownership. In this case, it is the 'intention to think' which will normally result in a sense of agency for our thinking, since an efference-copy of that intention will be transmitted to a comparator-component which, in its own turn, will compare that intention with the sensory feedback, that is, with 'what is thought'. The experience of agency is, in sum, the experience result of a correspondence between my intention to think and what has been thought. If something goes wrong with the efference copy and if, as a consequence, the latter is unavailable as a comparative term in the comparator-component, then the result is a thinking which does not seem to be generated by the subject. The efferent instruction will still take place and therefore results in the generation of thoughts, but these will not be accompanied by that which, according to Kant, *should* be able to accompany all my representations: 'I think *p*'. The mechanism of sensory feedback will still register *that* there is thought and *what* has been thought – the sense of ownership is retained – but this appears as unexpected and strange for the subject: 'He or it (the thing) thinks *p*'.<sup>37</sup>

---

<sup>37</sup> There is a lot of discussion in the philosophical literature about whether or not VHS and other related schizophrenic experiences offer an empirical falsification of the famous transcendental principle according to which we should always be able to recognize thoughts as *our* thoughts (better known as the "immunity to error though misidentification – principle", see Shoemaker (1968)). Campbell, for example, expresses the philosophical relevance of schizophrenic pathology as follows: "In making a judgement like, "I see a comet," there are not two steps, finding out who is seeing the thing and finding out what it is that is being seen, so that you could go wrong

Concluding, we can say that Frith's model for the explanation of VHs offers a number of advantages relative to classical psychiatric approaches in which hallucinations are usually conceived in terms of a 'perception without an object'.<sup>38</sup> Following Merleau-Ponty (1945, pp. 385-397; see also Lacan, 1966, pp. 531-583), these can be roughly divided into *empiristic* and *rationalistic* approaches. Empiristic approaches of VHs assume that the problem is situated in a disruption of the perceptual system, reason why certain things are being 'heard' which are not present in external reality. Rationalistic approaches, on the other hand, stress the contribution of the patient's judgment who wrongly concludes that a perceptual object is available when none is. Although both approaches differ with respect to the underlying cause of VHs (perceptual disruption vs. inadequate judgment), they nonetheless share a number of important assumptions: more specifically, (1) that the hallucination is essentially reducible to a perception; and (2) that the reality value of the hallucinatory experience for the patient is entirely reducible to an empirical claim about the presence or absence of the hallucinatory

---

at either step. The only place to go wrong is in your description of what is being seen. [...] A patient who supposes that thoughts have been inserted into his mind by someone else is right about *which* thoughts they are, but wrong about *whose* thoughts they are. So thought insertion seems to be a counterexample to the thesis that present-tense introspectively based reports of psychological states cannot involve errors of identification" (Campbell, 1999, p. 90, our italics). Descartes, as one remembers, fell back on the immediate certainty of his thinking I in order to escape the (implausible, but nonetheless imaginable) possibility of general deception. Despite the fact that, according to Descartes, I might be mistaken about quite a lot of things that I believe, I cannot be mistaken about the fact that it is I who is believing/thinking. Yet, the question is: did our previous analysis of VHs rebuke Descartes? And furthermore, did we offer, by means of Frith's neurocognitive modeling of the process of self-monitoring, a naturalistic explanation of a transcendental argument?

Gallagher, for example, in his general effort to draw together naturalistic neuro-cognitivist accounts and (Husserlian) phenomenology more closely, believes this to be the case: "If we can identify which mechanisms fail at the cognitive or neurological level when the schizophrenic patient suffers from thought insertion, then we also have a good indication of the mechanisms responsible for (or at least involved in) the normal immunity to error found in self-reference, and the immediate sense of self. This insight moves us from the conceptual and often abstract arguments of philosophy to the more empirical inquiries of neuropsychology and neurophysiology" (Gallagher, 2000, p. 17). Yet, we believe there is something suspicious about these sorts of claims, since the applauded shift from a priori to empirical, a posteriori arguments about self-consciousness always seems to be effectuated in the same way, thereby revealing a one-sided view on the relation between phenomenology and neurocognitive explanations. Although one of the most ardent advocates of this marriage, Francisco Varela, originally spoke about a "methodology of reciprocal constraints" (Varela, 1996, p. 343), it seems that in Gallagher's case, Varela's reciprocity is rather hard to find – as becomes clear in Gallagher's following *précis* of this relation: "If the phenomenology indicates that we are not conscious of X, or that our consciousness is not X-like, but the neurological explanation inserts X into the picture, so to speak, then the neurological explanation is not explaining consciousness as it is" (Gallagher, 2007, p. 303). So in Gallagher's view, the applauded shift is therefore only to be 'applauded' when neurocognitive science adopts *what was already ascertained* in phenomenology. If, on the other hand, it chooses to deviate from these phenomenological insights, then apparently neurocognitive science is explaining something else. It could be that we are missing out on something here, but it seems that in such a relation Varela's reciprocity is nowhere to be found, and that the happy marriage between transcendental phenomenology and naturalistic cognitive science shows all the signs of a sort of 'forced marriage'.

<sup>38</sup> This definition of VHs as 'perception without an object' has a long history in classical psychiatry and goes back to the following description by the French psychiatrist Esquirol: "L'hallucination est un homme qui a la conviction intime d'une sensation actuellement perçue alors que nul objet extérieur n'est à portée des sens [...]" (Esquirol, 1838, p. 80). That, in this regard, not too much has changed in contemporary psychiatric thought about VHs is clear from the following, more recent definition: "[...] a sensory experience which occurs in the absence of corresponding external stimulation of the relevant sensory organ, which has a sufficient sense of reality to resemble a veridical perception, over which the subject does not feel s/he has direct and voluntary control, and which occurs in the awake state" (David, 2004, p. 110).



object in the external world. A fundamental problem for these assumptions is that schizophrenic patients who experience VHS in general tend to make a distinction between reality bound perceptions and the essentially private character of VHS (Moritz & Laroi, 2008). Yet, despite that distinction, they nonetheless testify to the reality of the hallucinatory experience as such. We should therefore distinguish – to put it in Husserlian terms – between, on the one hand, the noetic claim that the hallucinatory experience is real, and, on the other hand, the noematic claim that the object of hallucination is real. In perception, these two dimension coincide, in VHS this is clearly not the case.

Frith, on the other hand, is able to avoid this confusion between perception and hallucination by no longer considering VHS as a ‘perception without an object’ but rather as an ‘action without an agent’. Because of this, the classical emphasis on the presence or absence of the perceptual object as determinative for the hallucinatory experience is left behind. That is, in Frith’s model, it is the passive relation of a subject with regard to VHS which is central: patients do not *have* hallucinations, but are being *hallucinated*. Because of a disruption in the comparator-mechanism, patients are no longer able to recognize their thoughts *as* their own thoughts. As we made clear, this idea is grounded in the extrapolation of the neurocognitive mechanism which are held responsible for the subjective realization of agency and ownership in the case of movement towards thinking. In the next two sections, we will critically discuss this extrapolation by means of (1) an immanent reading of the comparator-model in the case of thinking; and (2) a comparison with the nature of hallucinatory experience.

### 3. Comparator-supposed-to-know

A first question we want to address is whether Frith’s model is able to offer a plausible explanation for the sense of agency in the case of non-schizophrenic thought, that is, regardless of the question whether it also constitutes a sufficient explanation for VHS. This first question is based on another, more fundamental question: that is, whether thinking is characterized by a sense of agency in the first place? It is important to stress that the latter question is phenomenological: we are *not* asking whether we are the author or intentional cause of our thoughts *in reality* – but whether our thinking is accompanied by the subjective experience of agency?<sup>39</sup> It should be clear that the answer to this question has some

---

<sup>39</sup> That it might be important to distinguish between these two variables – reality/appearance – in the case of agency is evidenced in the work of Daniel Wegner. In a series of empirical studies (Wegner & Wheatley (1999); Wegner (2002)), Wegner developed the theory of ‘apparent mental causation’ according to which “people experience conscious will when they interpret their thought as the cause of their action. This means that people experience conscious will quite independently of any actual causal connection between their thoughts and their actions” (Wegner, 2002, p. 64). Continuing Hume’s original idea about causality, Wegner identified three factors which seem to be sufficient to provoke a sense of agency while in reality there was no causal influence of the subject on a given experimental event – namely, (1) *priority*: the time interval between the conscious intention/thought and the effect must be sufficiently short; (2) *consistency*: the intention/thought must be consistent with the supposed effect; and (3) *exclusiveness*: the thought/intention must be the only apparent cause of the effect. Yet, the question is whether, in Wegner’s study, it is indeed the experience of agency which is at stake, or rather, that the experimental subjects only *concluded*, on the basis of the temporal contiguity between intention and effect,

important implications for assessing the relevance of Frith's model as a neurocognitive explanation of a certain aspect of our subjectivity: if the explanandum is not characterized by a sense of agency, then it appears to be senseless to include the explanandum under that description as something which is to be explained. Furthermore, it should urge us to reconsider the way in which VHS are described in the research literature: is it correct to reduce the pathological essence of hallucinatory experience to a sense of ownership without a sense of agency if our normal thinking is characterized as well by the absence of a sense of agency?

### 3.1 Ghostly doubles

Let us first assume that we have obtained an adequate phenomenological description of our thinking – i.e. as follows: “Thinking, like all our actions, is normally accompanied by a sense of effort and deliberate choice as we move from one thought to the next” (Frith, 1991, p. 66). We saw that it is the intention to think which, in Frith's model, is determinative for the “sense of effort and deliberate choice” we experience in the course of running through our thoughts. Yet, how should we understand such an intention to think? As Gallagher (2004, p. 11) remarks, what possible role does something like an intention to think play in the case of the conscious experience of thinking? It seems to be difficult to even conceive something of this nature unless, for example, we consciously decide to start thinking about the problem of infinite regress which looms large in the very idea of thinking thoughts. But in that case, the intention to think this problem is itself also a thought and the regress indeed becomes unavoidable: do I need an intention to think for the intention to think for ...?<sup>40</sup> A simple solution – suggested by Stephens and Graham (2000), as well as Gallagher (2004) and Campbell (1999) – seems to save us from this embarrassment: if we move the efference copy (and hence, the intention to think which it represents) to the neurocognitive unconscious, then we won't have to assume that a conscious intention to think should precede our thinking. In that case, the only thing we are actually conscious of is the outcome our thought process, i.e. our thoughts. The intention to think, like the feedback mechanism described earlier, would be part of the sub-personal mechanism responsible for the

---

that they were the cause of the effect (this being the most plausible explanation given the experimental set-up), that is, without their being any phenomenology of agency involved.

<sup>40</sup> This objection was already raised by Akins & Dennett (1986) in response to Hoffman's explanation of VHS. Although Hoffman's model is not entirely similar to Frith's proposal, the objection is also valid in Frith's case. Here's how Akins & Dennett describe the problem: “Hoffman's account is threatened by (inter alia) an infinite regress: If we identify “slips of the tongue” as misexecutions of ... intentions, relative to what could there be “slips of the thought”? Wittgenstein thought that “slips of thought” were impossible because a mistake presupposes an intention. Although we can intend to reach a particular (cognitive) conclusion — say, discovering a new way to drive home from the office — we cannot intend each of the thoughts that constitute our attempted derivation or discovery, on pain of generating a never-beginning regress of intentions to form thoughts. Some thoughts must just “come to mind,” however apt, well-ordered, and useful they prove to be in the larger project” (Akins & Dennett, 1986, p. 517).

experience of agency in generating our thoughts. As Campbell (1999) puts it with regard to the preceding 'conscious intention'-model and the suggested solution:

This would, I think, lead to a quite implausible account of ordinary knowledge of your own thoughts on which you had, as it were, a "double awareness" of their contents, once as the contents known by introspection, and once as the contents of the efferent copies. But the model I set out (...) suggests a different way of developing the picture. You have knowledge of the content of the thought only through introspection. The content of the efferent copy is not itself conscious. But it is a match at the monitor between the thought of which you have introspective knowledge and the efferent copy that is responsible for the sense of being the agent of that thought (p. 34).

Yet is this suggested solution really sufficient in order to avoid the problem of infinite regress? According to both Campbell's and Frith's proposals, there should be two different mental representations with the content  $p$ : the first with the propositional attitude 'intention' towards  $p$  which would be unconscious – the second, the proposition  $p$  which would be conscious. In both cases, the content should indeed be the same for the comparison by the comparator-mechanism to become possible. So it seems that the only difference between both thoughts is related to the fact that one is unconscious, while the other is conscious.<sup>41</sup> However, if (1) the intentional thought is itself an (unconscious) thought and (2) all thoughts are considered to be the result of a motor process initiated by an intention, then again (3) it seems necessary to presuppose another (unconscious) intention to produce this (unconscious) intention, ... In other words, this still leads to an infinite regress.

---

<sup>41</sup> What we seem to have here, in this endless proliferation of thoughts with the same propositional content, is a regressive situation which can be understood in terms of Ryle's description of the use of "ghostly doubles" in what he calls "the para-mechanical hypothesis" of mental life. He describes this situation as follows: "[...] for a person to use a significant word, phrase or sentence, there must antecedently or concomitantly occur inside him a momentary something, sometimes called 'the thought that corresponds with the word, phrase or sentence', leads us to expect that this supposed internal occurrence will be described to us. But when descriptions are proffered, they seem to be descriptions of ghostly doubles of the words, phrases or sentences themselves. The 'thought' is described as if it were just another more shadowy naming, asserting or arguing. The thought that is supposed to bear-lead the overt announcement 'tomorrow cannot be Sunday, unless today is Saturday' turns out to be just the announcement to oneself that tomorrow cannot be Sunday without today being Saturday, i.e. just a soliloquized or muttered rehearsal of the overt statement itself. We certainly can, and often do, rehearse in our heads or sotto voce what we are then going to tell the audience, or write down on foolscap. But this makes no theoretical difference, as the same supposed question again arises, 'In what does the significance of this soliloquized or muttered expression consist? In yet another 'thought to correspond' going on in yet another still more twilight studio? And would this be just another rehearsed announcement?" (G. Ryle, *The Concept of Mind*, Chicago: University of Chicago Press, 1949, p. 270). Regardless of Ryle's further discussion of the use of mental idioms to describe intelligent behavior in terms of logical categories, we believe that the suggested distinction between 'thinking' and 'thoughts' could be important for research on VHs. Thoughts are the results or outcomes of activities or processes which we often call 'thinking'. Thinking, on the other hand, can be understood as those activities and processes which result in thought. Important however is that the relation between 'thinking' and 'thoughts', or more generally, between 'processes' and 'results', does not have to be marked by resemblances in terms of content – in the same way as we don't assume there being a substantive similarity between, for example, the result 'being on time for an appointment' and the various processes which can lead up to this outcome (watching the time, taking the bike instead of a car, etc.).

Given the preceding difficulty, we seem to encounter a second – more Darwinian – problem. Indeed, one could wonder what the adaptive function of a such a thought-comparator could be besides generating a sense of agency.<sup>42</sup> Why would it be important, besides knowing *what* I think, to also know that it is me who is effectively thinking? Say, because otherwise, we would all end up in the same schizophrenic predicament? Or in order to render Descartes and Kant and evolutionary service? Again, Campbell (2004) comes up with a better explanation:

You can see the comparator as a cognitive mechanism linking your long-standing beliefs and memories and so on, on the one hand, and your current occurrent thoughts, on the other. And part of the role of the comparator here is to help to keep your thinking on track. It is easy to suppose that thinking is somehow an unregulated process; but that is certainly untrue. To see that you need only ask how it is that your thinking manages not to be exclusively the jumble of disconnected thoughts characterized of the formal thought disorder of some schizophrenic patients (p. 485).

Yet let us abstract for the moment from Campbell's doubtful diagnostic depiction of schizophrenic thinking as some kind of uncontrolled jumble of thoughts and let us focus instead on the proposed function of the comparator-mechanism. According to Campbell's proposal, the comparator-mechanism helps us in keeping our thinking 'on track'. What exactly this is supposed to mean is not really explained by Campbell, but we can assume that it should have something to do with the organization of our thinking in function of a desired goal. Suppose, for example, that we need to think about how we could be able to reach location B in the fastest way starting from location A. In that case, my thinking would be 'on track' if the series of resulting thoughts consisted in, for example, the possible options qua means of transportation, an accurate representation of the route I would have to follow, and so on. On the other hand, my thinking would be 'off track' if it would be constantly crossed by thoughts about the other sex or the price of bread. However, in this case, what is it exactly that needs to be 'tracked' by the comparator mechanism? It is clear that the coherency of my thinking in function of a certain goal is not the same as comparing an intended thought with the one which effectively arises. In the first case, the criterion of comparison is pragmatic and it already presupposes a 'space of reasons' within which my thinking can be meaningfully evaluated. In Frith's comparator model, however, the criterion is purely formal and hence – qua comparator-mechanism – cannot have anything to do with the question whether, given the goal of my thinking, the ensuing thoughts are really the desired ones. The only thing the comparator-mechanism is able to check is whether what is thought effectively matches the predicted thought, which results – depending on success – in the experience of agency. Differently put: the content of the intention and the content of the thought *should* be the same for a comparison to become possible, which means that the comparator can only keep one and the same thought 'on track'. Therefore, Campbell's tracking

---

<sup>42</sup> A possible solution to this question could be that not every cognitive function should need to possess a direct, adaptive function.

function cannot be fulfilled by the comparator mechanism under these conditions. Furthermore, in the preceding example with regard to the most efficient way to get from A to B, the regress problem also returns: why is this thought not experienced as a VH? Because the intention to think is preceded by another intention, ...?

Hence, the extrapolation of the comparator model from movement to thinking presents us with a dilemma: either we describe intentions as thoughts and we get confronted by an infinite regress, or we describe them in a different way but in that case the function of the comparator mechanism (generation of agency and/or thought coherency) can no longer be fulfilled.

### 3.2 I think ... I think?

This leads us to our second question: that is, whether the comparator model *should* fulfill this function in the first place? We have specified this question above as the phenomenological question whether our thinking – schizophrenic or not – is characterized by a sense of agency? In the case of Frith's own thinking, this is apparently the case since it is accompanied by "a sense of effort and deliberate choice as we move from one thought to the next" (Frith, 1991, p. 66). Furthermore, it is generally assumed that with this 'we', Frith finds himself in good company. That is, amongst others, Descartes, Kant, Wittgenstein, Husserl and Shoemaker are supposed to have supported this idea of thought-agency and to have considered our question as nonsensical. This nonsensicality is often philosophically expressed in terms of the principle of an 'immunity to error through misidentification': the idea that we cannot be mistaken about *who* is thinking. That is, while, presumably, I would still be able to count on a certain amount of comprehension if I were to question the content of my interlocutor's thoughts, this would no longer be the case if I would go on to ask if he is also certain about whether he is indeed the one who is effectively thinking. The 'mineness' of my thoughts therefore seems to be supported by our everyday intuitions and a long and illustrious philosophical tradition.

Yet, in the preceding, we tried to make clear that this notorious 'mineness' of our thoughts can be understood in at least two different ways: that is, in terms of ownership and in terms of agency. In the case of Descartes, for example, it is not immediately clear which one of these ideas is supposed to have survived the Cartesian epoché: whether I necessarily exist because I have thoughts or because I think thoughts? In the same way, with regard to Kant's famous principle according to which an 'I-think' should be able to accompany all my representations, this idea still seems to be compatible with a distinction between an idea of ownership and agency in the case of thinking. Here the idea is that I have only one consciousness and that this consciousness is indeed *my* consciousness. In that sense, every thought of which I am conscious is part of the same unified stream of consciousness and hence in principle susceptible to being preceded by an 'I-think'. However, this does not mean that I am therefore also the author of these thoughts. The only thing Kant claims is that whenever I am conscious of my thoughts, they are indeed my thoughts in the sense that they arise in my stream of consciousness: that

is, a transcendental defense of ownership rather than agency.<sup>43</sup> Parallel to this is the phenomenological emphasis on the subjectivity of our experience in terms of a transcendental *ipseity* according to which experiential givenness is always and necessarily a givenness for-me. The I here is never part of what is given on the level of the content of experience, but rather of the form of the ‘givenness of what is given’. That is, in the phenomenological tradition, the I refers to a formal condition of possibility of experience – what Heidegger (1967, p. 43) designated as *Jemeinigkeit* – and therefore comprises the tautological idea of ownership: everything I experience, is experienced by me (see Husserl, 1976, p. 168).<sup>44</sup>

Yet, do such arguments equally hold in the case of agency? Is agency something which is given in my experience of thinking? Is agency, for example, a transcendental condition of possibility without which I would no longer be able to recognize my thinking *as* thinking? We don’t think so. In order to introduce the defense of this claim, let us begin by considering a famous quote from Nietzsche’s *Beyond Good and Evil*:

With regard to the superstitions of logicians, I shall never tire of emphasizing a small terse fact, which these superstitious minds hate to concede – namely, that a thought comes when “it” wishes, and not when “I” wish, so that it is a falsification of the facts of the case to say that the subject “I” is the condition of the predicate “think”. *It* thinks; but that this “it” is precisely the famous old “ego” is, to put it mildly, only a supposition, an assertion, and assuredly not an “immediate certainty”. After all, one has even gone too far with this “it thinks” – even the “it” contains an *interpretation* of the process, and does not belong to the process itself. One infers here according to the grammatical habit: “Thinking is an activity; every activity requires an agent; consequently – (2000, p. 214).

Different elements from this quote are relevant for our discussion. First of all, Nietzsche’s remark “that a thought comes when “it” wishes, a not when “I” wish”. According to Nietzsche, our thinking is fundamentally characterized by a sense of a non-agency or passivity which renders us rather the inert witness of thoughts which impose themselves rather unsolicited and more or less unexpected. ‘It thinks’ ... whatever it is I might think about this. This lack of intentional agency is also found in various mental idioms by means of which we often describe our mental life: ‘I suddenly remembered I had to do *x*’ – ‘I realized that *p*’ – ‘finally, I understood that *q*’ – ‘I was thinking that *r*, but that is not how I actually think about it’ – ‘I suddenly saw the solution’ – ‘it slowly began to dawn on me that *s*’ – ..., even the ever spontaneous Kant confessed to have been *woken* up from his dogmatic slumber after his

---

<sup>43</sup> For a similar analysis, see Sellars (1974).

<sup>44</sup> As Zahavi puts it: “Whether a certain experience is experienced as mine or not does not depend on something apart from the experience, but on the givenness of experience. If the experience is given to me in a first-personal mode of presentation, it is experienced as my experience, otherwise not. To be conscious of oneself, is consequently not to capture a pure self that exists in separation from the stream of consciousness, rather it just entails being conscious of an experience in its first-personal mode of givenness. In short, the self referred to is not something beyond or opposed to the stream of experiences, rather it is a feature or function of their givenness. It is the invariant dimension of first-personal givenness in the multitude of changing experiences” (2005, p. 9).

confrontation with Hume's *Treatise*. Let us consider, for example, what happens when I try to think about, say, the difference between movement and thinking in terms of agency. I try to bring the word 'agency' before my mind; an image of a hand movement imposes itself despite not being very useful; some murmuring on the background; an unpleasant feeling that 'it' won't come; the names of Frith and Blakemore; I ask the question again *sotte voce* and wait what happens; a half-articulated thought about sensory re-afference emerges, ... Is what we presently described marked by "a sense of effort and deliberate choice as we move from one thought to the next"? Undoubtedly as far as the 'sense of effort' is concerned: I try to bring myself to think the thoughts I want to think. But all of this is – to use Strawson's expression (2003, p. 231) – merely "catalytic" or preparatory activity.<sup>45</sup> Differently put: all that I do, even in the most ideal case of trying to think about some subject in a conscious and focused way, consists in activity which is merely indirect with respect to the effective appearance of thought itself. Trying to bring certain words before one's mind, to focus one's attention on oneself, rewinding the train of thoughts, ... all such activity consists in facilitating conditions for thinking, but not the immediate production of thoughts. Thinking therefore lacks Frith's "sense of deliberate choice as we move from one thought to the next". Perhaps it is activity, but in the sense that there is also volcanic activity. It might even be – as Kant put it – spontaneous, but in the sense that one cannot try to be spontaneously spontaneous. In the less ideal case, this lack of agency should even be more obvious: unwanted thoughts about adulterous partners, sexual fantasies, doubts with regard to one's own competence, ... come when 'it' wants, and not when 'I' want – what on earth does 'it' want from me?

A second important element from Nietzsche's quote is the claim that the intentional activity of the I should be regarded as merely "a supposition, an assertion" and "assuredly not an immediate certainty" (Nietzsche, 2000, p. 214). If what we have argued is correct, then a sense of agency is precisely *not* given at the level of our immediate experience of thinking. It is not a feature of thinking, and therefore cannot arise on the basis of this thinking activity. However, is this to say that we have therefore eliminated the diagnostic distinction between everyday thoughts and VHs? This seems to be too high a price to pay, especially because schizophrenic patients seem to distinguish themselves between normal thoughts and VHs. Such a distinction is moreover necessary for the patient in order to be able to recognize VHs *as* VHs: the complaint that someone else is thinking can only be formulated on the basis of a distinction between thinking and VHs, since that complaint is itself not experienced as a VH. VHs are precisely *dystonic* in relation to the patient's normal world of experience. Hence, this leads us the following situation: if we want to keep the distinction between thinking and VHs, there are two different possibilities, namely

- (1) Thinking is, contrary to what we have argued above, characterized by a merging of ownership and agency and this is precisely what sets it apart diagnostically from VHs;

---

<sup>45</sup> For a similar analysis in the case of 'believing' and 'judging', see Setiya (2013).

- (2) The distinction between, for example, unbidden thoughts and VHs cannot be made on the basis of the presence or absence of agency.

The second option implies that, in Frith's model, hallucinatory experience is not adequately described and therefore also fails as an explanation. We will discuss this second option in the next section.

#### **4. Who (He or It (the Thing)) thinks what?**

Let us begin by summarizing what in Frith's model, on the basis of the neurocognitive mechanisms which are advanced in order to explain the subjective experience of movement and thinking, is precisely considered as the pathological essence of hallucinatory experience. In Frith's model, VHs are above all experiences of *passivity*: the hallucinated subject is overwhelmed by various mental contents which the comparator-mechanism was not able to adequately predict. As a consequence, the subject is surprised by capricious and autonomous thoughts which no longer seem to correspond to its intentions to think. Hence, the hallucinatory condition is structurally similar to the condition in our earlier example of someone else grabbing my hand unexpectedly and unsolicited: I move/think without having the feeling of being its initiator. The question is therefore: is such a description sufficient as a phenomenological characterization of VHs?

A first problem that arises for the comparator-model is that passivity or a sense of non-agency with respect to our thinking is not simply to be identified with a causal attribution of this thinking to a specific other or strange entity.<sup>46</sup> In terms of our Kantian idiom: 'there is thought' is not the same as 'he or it (the thing) thinks'. Or better yet, egocentric passivity does not equal allocentric activity. Yet this is nevertheless precisely what patients with VHs seem to uphold: not merely that there is thought, but that a specific other (God, the psychiatrist, a friend, ...) or entity (cosmic powers, the universe, a formless soul, ...) thinks in their place. The mere discordance between efferent intentions to think and re-afferent thoughts results at most, under the conditions of Frith's model, in an absence of agency, but not in the supposed presence of intentional activity of someone else.<sup>47</sup> Of course, it is not too difficult to introduce a second element by means of which the absence of agency at the level of experience could give rise to an interpretation *ex post facto* about the presence of another, in terms of the idea: 'if it's not me, then it must be he or it (the thing)'. This rationalistic maneuver can be found in Frith's following suggestion:

---

<sup>46</sup> For this point, I rely on Gallagher (2009). It is clear that this problem is not only confined to the specific case of VHs, but also has broader implications for the comparator-model as a general theory for the distinction between self/world. For an account of this broader problem, see Synozik, Vosgerau & Newen (2008).

<sup>47</sup> The alternative 'simulationist-model' developed by Jeannerod et al. might be able to account for this alterity (for an overview, see Pacherie, Green & Bayne (2006)). However, as we have argued, it is doubtful whether such models can be expanded in order to account for thinking.



We suggest that, in delusions of control, the prediction mechanism is faulty and as a consequence self-generated movements are not attenuated and are wrongly classified as externally generated. The patient is not aware of the predicted consequences of a movement and is therefore not aware of initiating a movement. *In parallel, the patient's belief system is faulty so that he interprets this abnormal sensation in an irrational way* (Blakemore, Wolpert & Frith, 2002, p. 240, our italics).

So the schizophrenic patient not only appears to be the victim of a malfunctioning self-monitoring system, reason why he is no longer conscious of his own intentions, he furthermore succumbs to an irrational interpretation of such an unusual experience. That is, the experience of non-agency is followed by an incorrect and irrational judgment about the intentional presence of another, while in this case only the claim 'there is thinking' would constitute the right interpretation of the phenomenological facts. Such rationalistic solutions have one important advantage: namely, that they are able to explain the alterity of VHS rather easily by referring to the contribution of an interpretation. Yet, whatever rationalistic approaches gain on this level, they lose with regard to the explanation of this interpretation itself. Indeed, now the question becomes the following one: why do schizophrenic patients opt for precisely this (pathological) explanation and not for the more plausible explanation that, for example, there might be something wrong with their self-monitoring system? With respect to this point, one can hardly award patients Ockhamian virtues, despite the fact that alternative explanations are often actively suggested by professional helpers and family. Furthermore, there seems to be something strange about the way in which these 'incorrect' interpretations are being upheld: on the one hand, they often lack the provisional and tentative character of ordinary interpretations (i.e. Karl Jasper's "Unbeeinflussbarkeit" (1973, p. 78)), on the other hand, such 'interpretations' often do not seem to lead to behavior which would be consistent with these interpretations (i.e. Eugen Bleuler's "double bookkeeping" (1950, pp. 127-130)). In short, a rationalistic interpretation of VHS is always confronted with difficulties concerning the explanation of this irrational interpretation itself.

A second problem, closely related to the former, is the fact that 'he or it (the thing)' doesn't seem to think no matter what. Differently put, in so far as Frith's comparator model does seem to hold a certain amount of plausibility as long as we focus exclusively on the phenomenological *form* of hallucinatory experience in terms of the variables ownership and agency, from the moment we begin to consider the specific *content* of VHS this becomes even more problematic. As we have seen, the comparator-model, *qua* predictive agency mechanism, does not take into account the content of the thoughts that are being used as comparative material in the efference copy. On this level, therefore, we should not expect any difference between the thoughts someone attributes to himself and VHS which are attributed to someone else. However, as evidenced in clinical practice<sup>48</sup>, there seems to be an

---

<sup>48</sup> See, for example, the study by Jackson & Fullford (1997). In our own interview study, conducted with 20 patients diagnosed with a schizophrenic disorder (Feyaerts & Vanheule, unpublished), it was clear that the content of VHS was anything but neutral with regard to the self-experience of patients, but often contained an explicit sexual, metaphysical or identity-related message. One patient described the 'messages' he received as

important difference with respect to the content of self-attributed thoughts and VHS. ‘He or it (the thing)’ usually doesn’t think trivial things like “don’t forget to take out the garbage”, often the content is related to issues which are meaningfully connected to the identity of patient. VHS are regularly offensive, outrageous or blasphemous, often contain imperative messages, address a subject in a most intimate way or question its existential position– as for example man/woman, father/mother or son/daughter – in relation to others (Laroi et al., 2010; Vanheule, 2011). In other words, VHS are not only experienced as ‘strange’ because they lack a formal sense of agency, it is also the content of these messages which lead to a situation in which the subject is confronted with a fundamental perplexity or, to use Blankenburg’s eloquent term, *Ratlösigkeit* (1971, p. 54).

To conclude, we can say that Frith’s phenomenological description of VHS as an experience of passivity does not offer a sufficient characterization of hallucinatory experience and this on two levels. On the one hand, because both the essential alterity as well as the content of VHS cannot be understood merely in terms of passivity. On the other hand, because it doesn’t offer a sufficient differential-diagnostic criterion to distinguish VHS from, for example, unbidden or unexpected thoughts. The questions this raises and the requirements this imposes for efforts to offer a neurocognitive modeling of hallucinatory experience, will be discussed in our conclusion.

## 5. Conclusion

In the preceding, we have critically discussed Frith’s self-monitoring hypothesis as an explanation of hallucinatory experience. In our discussion, we raised two central questions. First, whether the extrapolation of the self-monitoring model from movement to thinking is justified on the conditions which were set forth to explain ownership and agency in the case of movement. Here we argued that in the case of thinking, Frith’s model either leads to an infinite regress, or is unable to fulfil Frith’s sense of agency or Campbell’s proposed function of thought coherency. Secondly, we raised the phenomenological question whether Frith’s model should offer an explanation of the sense of agency in the subjective experience of thinking in the first place. Moreover, this led to the further question whether the differential-diagnostic characterization of VHS as a loss of agency while preserving ownership offers an adequate description of hallucinatory experience when compared to normal thought. We have answered both questions negatively. On the one hand, we pointed out that agency is not an immediate phenomenological given in our experience of thinking. Despite the fact that most people claim to be the authors of their thoughts, in our reading, such claims cannot be grounded in phenomenological appeals to ‘givenness’. Not primarily because such claims that are grounded in intuitive and immediate evidence

---

‘mind-fucking’ and ‘rape’ since the content of the VHS seemed to consistently allude to the sexual identity of the patient. As becomes clear in the following fragment: “The voices I hear, it’s always a matter of “you’re a whore boy, you’re a whore”. Yes, or “you have to show you are a whore”, things like that. And then I think to myself “man, just leave me alone, what the fuck is this!?””.

are philosophically suspect (see Feysaerts & Vanheule, 2015), but because even such a subjective ‘intuition’ is in this case absent. However, does this mean that we should cast the ordinary claim of agency to the domain of fiction? As we made clear in our question as to the correct phenomenological characterization of VHS, this seems problematic for two reasons. On the one hand, it would henceforth become difficult to preserve the distinction between thinking and VHS, and on the other hand, because even possible fictional elements require an adequate explanation.

This leads us to the following, speculative proposal: if our sense of agency is not dictated by the phenomenology of thinking, it follows that the self-attribution in the case of normal thinking, as well as the other-attribution in the case of VHS, are not so much a misrepresentation of an underlying reality in the sense that both accounts ascribe the origin of our thoughts to a wrong instance. Differently put: the misidentification is not misguided on the level of the supposed ‘thinking’ subject – i.e. ‘I’ or ‘he or it (the thing)’ – but rather at the level of this supposition itself – that is, an ‘I’ or ‘he or it (the thing)’ is identified despite the fact that nothing of this sort is given in our experience. It is for this reason that we want to consider the attribution of agency on the level of thinking as a transcendental illusion as in the precise definition given by Zupancic:

[...] transcendental illusion is the name for something that appears where there should be nothing. It is not the illusion of something, it is not a false or distorted representation of a real object. Behind this illusion there is no real object; there is only nothing, the lack of an object. The illusion consists of “something” in the place of “nothing”, it involves deception by the simple fact that it is, that it appears (Zupancic, 2002, p. 69)

Or as she put it elsewhere: “Transcendental illusion has to do not with content of an ‘image’, but with its very existence – it deceives on the level of being” (Zupancic, 2000, p. 67). Yet although it might seem that we have entered the field of wild, metaphysical speculation and therefore are a long way from the earlier discussed neurocognitive models, we can nevertheless try to deduce some implications from this idea which allow us, if not to solve, then to put the discussed problems of Frith’s model in a different light.

First of all with regard to the phenomenological description of VHS as a loss of agency with a conservation of ownership. We argued that a lack of agency is not a sufficient condition to distinguish VHS from, for example, unbidden and intrusive thoughts. Yet, there seems to be a clear distinction between both: patients with VHS ascribe thoughts for which they do not have an experience of agency to an external instance, while a same sense of non-agency in the case of, for example, obsessive patients with intrusive thoughts, still seems to lead to a self-attribution (see Hoffman, 1986; Stephens & Graham, 2000). Above, we have described both attributions as transcendental illusions: an ‘I’ or ‘he or it (the thing)’ appears where nothing should have appeared. The attribution is hence not so much to be understood as an incorrect *representation* or *description* of an underlying experience of agency, but

rather as an *identification* with what essentially doesn't carry any information about identity.

The question by which we want to conclude therefore becomes, secondly, how we should understand the possibility of such an identification? How is it that we identify with thoughts despite the fact that our representational system doesn't comprise information about the identity of a thinker, not at the level of the process of thinking, nor at the level of the content of our thoughts? In other words, our proposal is to effectively *reverse* the usual research logic in the study of VHs: whereas, in most cases, one starts from a normative description of the non-pathological – i.e. normal thinking as 'I think' – in order to compare it subsequently with the pathological – i.e. VHs as 'he or it (the thing) thinks', in our view, it seems more fruitful to start from a similar subjective indeterminacy – 'there is thinking'. Only in a further step does this lead (or not) to an identification as transcendental illusion. The question for a neurocognitive modelling of such a process hence becomes the following: which elements are minimally responsible for, not the correct or incorrect representation of the underlying activity of a neurocognitive system, but rather the identification with what is essentially neutral with respect to the attribution of subjectivity? More specifically: which criteria do, for example, obsessive and schizophrenic patients use in order to arrive at the fictional conclusion that thoughts are respectively accompanied by an 'I think' or an 'he or it (the thing) thinks', despite the fact that in both case a sense of agency is lacking?

## References

- Akins, K. & Dennett, D. (1986). Who may I say is calling? *Behavioral and Brain Sciences* 9 (3), 517.
- Bentall, R.P. (1990). The Illusion of Reality: a review and integration of psychological research on hallucinations'. *Psychological Bulletin* 107 (1), 82-95.
- Blakemore, S.-J., Wolpert, D. & Frith, C.D. (2000). Why can't you tickle yourself? *Neuroreport* 11, 11-16.
- Blakemore, S.J., Wolpert, D.M. & Frith, C.D. (2002). Abnormalities in the awareness of action. *Trends in Cognitive Science* 6, 237-242.
- Blankenburg, W. (1971). *Der Verlust der Natürlichen Selbstverständlichkeit: Ein Beitrag zur Psychopathologie Symptomarmer Schizophrenien*. Stuttgart: Ferdinand Enke Verlag.
- Bleuler, E. (1950). *Dementia praecox or the group of schizophrenias*. New York: International Universities Press
- Campbell, J. (1999). Schizophrenia, The Space of Reasons and Thinking as a Motor Process. *The Monist* 82 (4), 609-625.
- Campbell, J. (1999). Immunity to error through misidentification and the meaning of a referring term. *Philosophical Topics* 26 (1-2), 89-104.
- Campbell, J. (2004). The first person, embodiment, and the certainty that one exists. *The Monist* 87, 475-488.
- David, A.S. (2004). The cognitive neuropsychiatry of auditory verbal hallucinations: An overview. *Cognitive Neuropsychiatry* 9 (1-2), 107-23.
- Esquirol, J-E. (1838). *Des maladies mentales considérées sous le rapport médical, hygiénique et médico-légal. Tome 2*. Paris: J.B. Baillière.
- Frith, C.D. (1992). *The Cognitive Neuropsychology of Schizophrenia*. Hillsdale: Erlbaum.
- Frith, C.D. (2012). Explaining delusions of control: the comparator model 20 years on. *Consciousness and Cognition* 21, 52-54.
- Frith, C.D., Blakemore, S.J. & Wolpert, D.M. (2000). Abnormalities in the awareness and control of action. *Philosophical Transactions of the Royal Society of Biological Sciences* 355, 1771-1788.
- Frith C.D. & Done, D.J. (1988). Towards a neuropsychology of schizophrenia. *British Journal of Psychiatry* 153, 437-443.
- Frith, C.D. & Done, D.J. (1989). Experiences of alien control in schizophrenia reflect a disorder in the central monitoring of action. *Psychological Medicine* 19, 359-363.
- Gallagher, S. (2000). Philosophical conceptions of the self: implications for cognitive science. *Trends in Cognitive Science* 4, 14-21.
- Gallagher, S. (2004). Neurocognitive Models of Schizophrenia: A Neurophenomenological Critique. *Psychopathology* 37, 2004, 8-19.
- Gallagher, S. (2007). "Neurophilosophy and neurophenomenology", in L. Embree & T. Nenon (Eds.), *Phenomenology 2005: Vol. 5, selected essays from North America*. Bucharest: Zeta, 293-316.
- Gallagher, S. (2007). Sense of agency and higher-order cognition: Levels of explanation for schizophrenia. *Cognitive Semiotics* 0, 2007, 32-48.
- Gallagher, S. (2009). "Delusional realities", in M. Broome & L. Bortolotti (Eds.), *Psychiatry as cognitive neuroscience: Philosophical perspectives*, Oxford: Oxford University Press, 245-266.
- Heidegger, M. (1967). *Sein und Zeit*. Tübingen: Max Niemeyer Verlag.
- Helmholtz, H. (1866). *Handbuch der Physiologischen Optik*. Leipzig: Voss.
- Hoffman, R. (1986). Verbal hallucinations and language production processes in schizophrenia. *Behavioral and Brain Sciences* 9, 503-517.
- Husserl, E. (1976). *Die Krisis der europäischen Wissenschaften und die transzendente Phänomenologie. Eine Einleitung in die phänomenologische Philosophie. Husserliana VI*. Den Haag: Martinus Nijhoff.
- Jackson, H.J. (1958). *Selected Writings. Volume 1*. New York: Basic Books.
- Jackson, M.C. & Fullford, K.W.M. (1997). Spiritual experience and psychopathology. *Philosophy, Psychiatry and Psychology* 4, 41-65.
- Jaspers, K. (1973). *Allgemeine Psychopathologie*, Berlin: Springer-Verlag.
- Jeannerod, M. (1994). The representing brain, neural correlates of motor intention and imagery.

- Behavioral and Brain Sciences*, 17, 187-245.
- Jeannerod, M. (1999). To act or not to act: perspectives on the representations of actions. *Quarterly Journal of Experimental Psychology* 52a, 1-29.
- Lacan, J. (1966). "D'une question préliminaire à tout traitement possible de la psychose", in J. Lacan, *Ecrits*. Paris: Editions du Seuil, 531-583.
- Larøi, F., de Haan, S., Jones, S. & Raballo, A. (2010). Auditory verbal hallucinations: Dialoguing between the cognitive sciences and phenomenology. *Phenomenology and the Cognitive Sciences* 9, 225-240.
- Malenka A.J., Angel, R.W., Hampton, A.J. & Berger, P.A. (1988). Impaired central error correcting behaviour in schizophrenia. *Archive of General Psychiatry* 153, 437-443.
- Mccarthy-Jones, S. (2012). *Hearing Voices. The Histories, Causes and Meanings of Auditory Verbal Hallucinations*. Cambridge University Press.
- Merleau-Ponty, M. (1945). *Phénoménologie de la perception*. Galimard.
- Moritz, M. & Larøi, F. (2008). Differences and similarities in the sensory and cognitive signatures of voice-hearing, intrusions and thoughts. *Schizophrenia Research* 102, 96-107.
- Pacherie, E. (2007). The anarchic hand syndrome and utilization behavior: a window onto agentic self-awareness. *Functional Neurology* 22, 211-217.
- Pacherie, E., Green, M. & Bayne, T. (2006). Phenomenology and Delusions: Who Put the 'Alien' in Alien Control? *Consciousness and Cognition*, 15, 566-577.
- Proust, J. (2000). "Awareness of Agency: Three levels of Analysis", in T. Metzinger (Ed.), *The Neural Correlates of Consciousness*. Cambridge: MIT Press, 307-324.
- Proust, J. (2006). "Agency in Schizophrenia from a Control Theory Viewpoint", in N. Sebanz & W. Prinz (Eds.), *Disorders of Volition*. Bradford Books, 87-118.
- Ryle, G. (1949). *The Concept of Mind*. Chicago: University of Chicago Press.
- Sellars, W. (1974). "... this I or he or it (the thing) which thinks ...", in W. Sellars, *Essays in Philosophy and its History*, Dordrecht: D. Reidel Publishing Company, 62-92.
- Setiya, K. (2013). Epistemic Agency. *Noûs (Philosophical Issues)* 23, 179-198.
- Shoemaker, S.S. (1968). Self-reference and self-awareness. *Journal of Philosophy* 65, 555-567.
- Sperry, R.W. (1950). Neural basis of the spontaneous optokinetic response produced by visual inversion. *Journal of Comparative Physiological Psychology* 43, 482-289.
- Stephens, G.L. & Graham, G. (2000). *When Self-Consciousness Breaks. Alien Voices and Inserted Thoughts*. Cambridge: MIT Press.
- Strawson, G. (2003). Mental Ballistics or the Involuntariness of Spontaneity. *Proceedings of the Aristotelian Society: New Series* 103, 227-257.
- Synofzik, M., Vosgerau, G. & Newen, A. (2008). Beyond the comparator model: a multifactorial two-step account of agency. *Consciousness and Cognition* 17, 219-239.
- Vanheule, S. (2011). A Lacanian perspective on psychotic hallucinations. *Theory and Psychology* 21(1), 86-106.
- Varela, F. (1995). Neurophenomenology: A methodological remedy for the hard problem. *Journal of Consciousness Studies* 3 (4), 330-349.
- von Holst, E. & Mittelstaedt, H. (1950). Das Reafferenzprinzip (Wechselwirkungen zwischen Zentralnervensystem und Peripherie). *Naturwissenschaften* 37, 1950, 464-476.
- Wegner, D.M. (2002). *The illusion of free will*. Cambridge: MIT Press.
- Wegner, D.M. & Wheatley, T. (1999). Apparent mental causation: sources of the experience of will. *American Psychologist* 54, 480-492.
- Zahavi, D. (2005). Being Someone. *Psyche: An Interdisciplinary Journal of Research on Consciousness*, 11 (5), 1-20.
- Zupancic, A. (2000). *Ethics of the Real: Kant, Lacan*. London: Verso Books.
- Zupancic, A. (2002). On Love as Comedy. *Lacanian Ink* 20 (Spring).



## THE LOGIC OF APPEARANCE: DENNETT, PHENOMENOLOGY AND PSYCHOANALYSIS<sup>49</sup>

---

### 1. Introduction

In his *Philosophy and the Scientific Image of Man* from 1963, Wilfrid Sellars offered a condensed, but compelling diagnosis that seems to retain a certain validity with respect to our current philosophical, scientific and social predicament. Sellars famously talked about a discord fueled by the conflict between two apparently opposing “images” of man-in-the-world (1963, p. 5): on the one hand, a *manifest image* which represents the tentative and provisional collection of ideas and assumptions we spontaneously adopt to characterize ourselves and the world we live in. Chief among these are ideas closely related to our immediate self-understanding as subjects endowed with (self-)consciousness, a certain extent of freedom of choice, gifted with reason but also desires, and so on. On the other hand, a relatively recent, but continually expanding *scientific image* of man as a “complex physical system” (1963, p. 25) — that is, an image conspicuously *unlike* its manifest double, but one which can be progressively distilled from various burgeoning scientific disciplines, such as neurology, physics, evolutionary biology and cognitive science. According to Sellars, the aforementioned discord results from the observation that modern science, although it has immeasurably enriched and challenged our understanding of phenomena by way of techniques and resources quite foreign to common sense, has begun to deploy these same techniques to phenomena that are deemed much closer to home. That is, those phenomena of which it was thought (or perhaps even hoped) they would forever lay beyond the pursuit of science and which are variably grouped under the headings of ‘meaning’, ‘first-person experience’ and ‘subjectivity’. In short, phenomena which for many philosophers, past and present, in some sense harbor the very key of what makes that image ‘human’ in the first place.

---

49



In a more recent past, Sellars' distinction has found a renowned application in the work of the well-known philosopher of cognitive neuroscience, Daniel Dennett. For the past few decades, Dennett has been vehemently advocating against "the fantasy of first-person science" (Dennett, 2001), arguing, amongst other things, that there is no such thing as a 'Cartesian theater' where private things happen that are spontaneously revealed before a "mental I/eye", one that, moreover, would form an unsurpassable source for self-knowledge from a first-person perspective. Accordingly, Cartesian intuitive certainty with regard to ourselves should be supplanted by the basic operation of healthy skepticism that characterizes all of modern science. Why indeed, Dennett asks, should we give up our scientific abstinence when, compared to some other notorious fictive entities such as the phlogiston or *élan vital*, it comes to seemingly less exotic matters such as the 'I', 'experience' and 'consciousness'? Quite clearly, no more as we compromise our scientific neutrality in anthropological fieldwork by treating local testimonies as an indication for divine presence, should we give up this abstinence when confronted with statements uttered from the first-person perspective. According to Dennett's well-known proposal, we should treat these manifest testimonies in the same way we treat testimonies in court, that is, as provisional *fictions* (Dennett, 1991, p. 79), where after we can safely let cognitive neuroscience do its job in deciding if, and to what extent, there might be "truth in fiction".

Given the central importance of the first-person perspective – in one way or another, and to be specified throughout our article – in both contemporary phenomenology and psychoanalytic theory, the questions we want to pose in this article are therefore the following ones: where do these disciplines respectively stand with regard to such proposals? And in what way would these two disciplines respond to Dennett's proposal to consider subjective beliefs about ourselves as provisional fictions and, to be sure, in all likelihood, future illusions?

In the first part, we will give a detailed reading of what is understood by such terms as 'first-person experience' and 'subjective appearance' in both phenomenology and Dennett's account and, on the basis of that reading, press some arguments that can be advanced against Dennett's fictionalization of subjectivity in favor of phenomenology. In the second part, we will offer a similar exercise in the case of psychoanalysis. We will argue that Sellars' distinction, exemplified by the discussion between Dennett and phenomenology, touches upon a classic ambiguity in psychoanalytic theory. Finally, a possible way out of this ambiguity will be sought through a close examination of the psychoanalytic concept of 'belief'.

## **2. Dennett and phenomenology: the seeming of seeming**

Let us take as a point of entry in the already quite extensive debate between phenomenology and Dennett<sup>50</sup> one of the latter's most cherished examples which he uses in his *Consciousness Explained* in order to fortify his Multiple Drafts Model of consciousness, a model which, Dennett warns us, "is initially deeply counterintuitive" (1991, p. 45). Dennett refers to a certain perceptual effect, better known in the literature as the *phi phenomenon* or *phi movement*, first so called by the Gestalt psychologist Max Wertheimer (1912), which was the object of repeated study in a series of experiments conducted by Kolers & von Grünau (1976). In the simplest case of the experiment, subjects were shown a single flash of a red spot, followed quickly by a displaced green spot, followed by another flash of the original red (each flash lasting about 150 milliseconds, with a gap of about 50 milliseconds in between). Perhaps not that surprising given our everyday smooth television-experience, the subjects reported seeing not two discrete spots, but one spot rapidly moving over and back again. However, one might ask, what about the change of color from red to green? Interestingly, the subjects reported the spot changing color about halfway through its trajectory from the initial flash to the second. As Dennett (1991) recounts the result: "the first spot *seemed* to begin moving and then change color abruptly in the middle of its illusory passage toward the second location" (p. 114, our emphasis). Given this surprising result, one of the collaborators to the experiments, the philosopher Nelson Goodman, asked the following question: "how are we able to fill in the spot at the intervening place-times along a path running from the first to the second flash *before that second flash occurs*" (Goodman, 1978, p. 73)? A pressing question indeed in so far as we want to avoid positing strange ontological scenarios wherein effects precede their causes and time starts running backwards.

So how then are we to proceed in answering Goodman's question? We will first begin by teasing out some basic principles in the way phenomenology would handle this issue, in order to articulate more clearly the dividing line which separates Dennett's account from the former.

## 2.1 The principle of appearance qua appearance

From a phenomenological perspective, when confronted with such an ambiguous situation, it would be wise to make use of Husserl's notorious maxim of phenomenological inquiry and "to go back to the things themselves" (Husserl, 1900-19001/2001, p. 168). Indeed, one of the guiding motifs and enabling assumptions of phenomenology is that *in principle*, and through availing ourselves of phenomenology's methodological procedures of 'epoché' and 'reduction', we would be able to put these explanatory questions aside for a moment and to focus instead on the experienced object (in this case the moving spot) precisely in the way it is experienced from the subjective or first person point of view. Rather than the hasty conjuring up of various theoretical explanations (e.g. Goodman's 'how are we able to fill in

---

<sup>50</sup> For a representative sample of the different discussions running through this debate, see the collection of essays in *Phenomenology and the Cognitive Sciences* (2007, issue 1-2).

the gaps?’), phenomenology’s traditional first move is to get a clear descriptive picture of what is to be explained in the first place - the *quod erat explicatum* for a science of consciousness -, precisely by letting, to use Heidegger’s inimitable phrasing, “that which shows itself be seen from itself in the very way in which it shows itself from itself” (Heidegger, 1962, p. 34). Despite the different ways in which contemporary phenomenology aims to fulfill its role in the debate around the nature of consciousness, there is indeed at least one unifying principle amongst the various phenomenological commentators. Let us call this the phenomenological principle of *appearance qua appearance*. In order to get a clear view of the sort of claims at stake in such a principle, let us begin by considering a few remarks by Husserl. In the *Crisis of the European Sciences and Transcendental Phenomenology*, Husserl (1970) writes:

The first thing we must do, and first of all in immediate reflective self-experience, is to take conscious life, completely without prejudice, just as what it quite immediately gives itself, as itself, to be (p. 233).

Here we have the first point: if anything, phenomenology, as a transcendental investigation of the first-person perspective, takes its point of departure “in immediate self-experience”, that is, it tries to offer a descriptive analysis of the way the world in general (Wertheimian spotlights included), but also other subjects and we ourselves, appear to us in ordinary experience. Crucially, in offering such a description of our everyday intentional involvement with the world, this appearance should be accepted “just as what it quite immediately gives itself, as itself, to be” (Ibid.). In short: on a phenomenological account, appearance should be taken as appearance *as such*, rather than as something else. Let us take one of Husserl’s examples to clarify and develop the ramifications of such an idea a little bit further.

Suppose we walk through a garden and become struck by a blossoming tree. How do we go about analyzing our immediate experience of the appearing of this blossoming tree? It is here that Husserl’s methodological tool of “epoché” comes into play; a notion, however, particularly fraught with all kinds of misunderstandings we need to clear up before furthering our analysis (see Overgaard 2015). According to a common misunderstanding, for example, the method of epoché or ‘bracketing’ would lead us to a consideration of the immanent contents of perception as they show up ‘in’ consciousness, severed and disconnected from any possible relation to the transcendent, real world which, conversely, is considered to be ‘outside’ consciousness (e.g. Burge, 2010; Carman, 1999). Accordingly, phenomenological bracketing would be first of all the bracketing of the realist presumption inherent in the natural attitude of our everyday experience. That is, the elimination of our pre-philosophical belief in the real world in favor of the appearance of Husserl’s blossoming tree as some kind of second blossoming tree, this time considered as an immaterial entity (i.e. image, datum, perceptum ... name your choice) somehow hovering before or within consciousness as in a Dennettian Cartesian theater. Yet closer inspection of some of Husserl’s own remarks on what the epoché is supposed to accomplish makes it clear that such an introspective reading is simply mistaken. Remarking on how that blossoming tree appears to consciousness after the epoché has been performed, Husserl (1983) writes

[...] everything remains as of old. Even the phenomenologically reduced perceptual mental process is a perceiving of “this blossoming apple tree, in this garden,” etc., and, likewise, the reduced liking is a liking of this same thing. The tree has not lost the least nuance of all these moments, qualities, characteristics *with which it was appearing in this perception*, <with which> it <was appearing as> “lovely,” “attractive,” and so forth “in” *this liking* [p. 216].

Hence, the practice of phenomenological bracketing as originally intended by Husserl does not consist in a reflexive retreat from reality the better to focus on contents or ‘qualia’ that are found to be within the self-enclosed terrain of pure consciousness. Nevertheless, although everything after the epoché “remains as of old” (Ibid.), there is of course a sense in which the epoché, in Husserl’s understanding, does reveal something that might have gone unnoticed before we put it to use. Again, Husserl (1970) remarks:

[...] through the epoché a new way of experiencing, of thinking, of theorizing, is opened to the philosopher; here, situated above his own natural being and above the natural world, he loses nothing of their being and objective truths and likewise nothing at all of the spiritual acquisitions of his world-life (...); he simply forbids himself – as a philosopher (...) – to continue the whole natural performance of his world-life (p. 152).

Two points are important in this rendition of the epoché.

The first thing to emphasize is that this new way of experiencing is not a matter of acquiring *additional* empirical information through further forms of experience, for example, by turning my introspective gaze from the world to my consciousness. This means that, with regard to the earlier example of Husserl’s blossoming tree, the epoché is intended to enable a faithful attending and description of the original experience of the blossoming tree precisely as it appeared in that very experience, that is, without changing anything about all the moments and elements that were already implicit in that experience. So if in our straightforward experience of the blossoming tree, it appeared as ‘actually existing’ or ‘real’, mind independent in whatever sense, but also ‘lovely’, ‘attractive’, ‘in this garden’ and so forth, then these presuppositions, and basically all positions with regard to that tree, should be retained in our description.

Yet, secondly, in this reflexive elucidation of conscious experience through the epoché, not only does the focus shift from world to world-as-experienced, but so does the criterion for *truth* and *error* with regard to phenomenological statements that try to express descriptively what was going on according to experience. That is, by abstaining from taking any positions towards the conscious experience of, for example, spotlights changing color at a certain moment, we seem to be moving to what is at least a different sort of epistemological terrain when compared to the epistemic situation of

empirical claims about the world (see Rosenberg, 2002). Yet, we should be careful as to the precise nature of the epistemic shift that, according to phenomenology, is at stake here.

Does such a claim, for example, imply some sort of “papal infallibility” (Dennett, 2007, p. 6) or incorrigibility with regard to first-person phenomenological beliefs relative to the first-person perspective? Here is how Koch (1997) formulates such an idea:

We shall take the *cogito* only as a means of suspending objectivity claims and of thereby inducing infallibility in what remains of the objectivity claim after suspension. This last point is important. For every objective truth claim, in which I am invariably fallible, there is a corresponding trivial truth claim, in which I am infallible, a truth claim which is fulfilled by the sheer fact that I seriously and honestly claim so. For every objective, thick truth claim, that p, there is a corresponding trivial, thin truth claim, that I think that p (or that it seems to me that p) (p. 73).

Hence, on Koch’s reading (but also Dennett’s, cf. *infra*), the epoché entails a movement from objectivity claims in which I am invariably fallible, e.g. the explicit judgment ‘the spot changes color about halfway its trajectory’, to a *suspension* of the same objectivity claim for which in that case I become infallible, ‘it seems to me that the spot changes color about halfway its trajectory’. Moreover, the reason why such a transition from judgment to suspension implies infallibility is because in this case, according to Koch, the cogito “falls within its own scope” (1999, p. 72). What this means is that the cogito, represented by the expression ‘I think that p’ or ‘it seems to me that p’, cannot be iterated without becoming trivial since an additional deployment of the epoché would add no further relevant information beyond what can already be found after the first suspension: ‘I think that p’ is logically equivalent to ‘I think that I think that p’ or ‘it seems to me that it seems to me that p’.

However, it is highly questionable whether Husserl and other major figures in the phenomenological tradition would apprehend the outcome of the epoché and the suggested first-person incorrigibility in such an intellectualistic and Cartesian way. Here it is important to refer back to what we said earlier about the epoché and, importantly, *how* its putative results are achieved. The question is: should we understand the phenomenological transition from world to world-as-experienced as a suspension from objectivity claims to subjectivity claims in terms of a regression to the ‘I think’ of the Cartesian cogito? Husserl (1970), for one, does not seem to endorse this reading:

It is naturally a ludicrous, though unfortunately common misunderstanding, to seek to attack transcendental phenomenology as “Cartesianism”, as if its *ego cogito* were a premise or set of premises from which the rest of knowledge (...) was to be deduced, absolutely “secured”. The point is not to secure objectivity but to understand it (p. 193).

What is behind Husserl's quite explicit rejection of understanding phenomenology as another form of "Cartesianism" (Ibid.)? Although much can (and has been) said about this (see e.g. Macdonald, 2000; Staiti, 2014), in the present context, the most important point to stress is that Husserl's and Descartes' projects are driven by fundamentally different questions. The relevant difference for our purposes is to be found in the opposition Husserl draws in the passage above between the Cartesian project of "securing" and that of phenomenological "understanding". For Descartes, as is well known, given the rather unlikely but still imaginable possibility of my being deceived by a vicious demon, the crucial question becomes that of securing some 'fundamentum inconcussum' (involving the two figures of the cogito and a benevolent God) in order to ward off the skeptical threat of perpetual illusion. On a more formal level, Cartesian hyperbolic skepticism represents a passage from uncertainty regarding the truth of most of our convictions regarding the world and ourselves, to certainty with regard to the truth of believing itself (i.e. Koch's thesis above). How then is this different from Husserl's description of the task of phenomenology not to secure, but to "understand" objectivity? And how is this related to the issue of first-person incorrigibility and truth? Here is how Husserl (1989) specifies this task:

Its sole task and accomplishment is to clarify the sense of this world, precisely the sense in which everyone accepts it – and rightly so – as actually existing. That the world exists – that it is given as an existing universe in uninterrupted experience which is constantly fusing into universal concordance – is entirely beyond doubt. But it is quite another matter to understand this indubitability, which sustains life and positive science, and to clarify the ground of its legitimacy (p. 420).

So, on the one hand, Husserl would basically subscribe to Peirce's sobering advice not to "pretend to doubt in philosophy what we do not doubt in our hearts" (1992, pp. 28-29), thereby denying that such a radical pretense is genuinely possible to begin with. Doubting, according to Husserl, is not a matter of simple decision or sheer will. In order to doubt something and to go beyond a philosophical play of mere make-believe, I need a relevant motivation to doubt and such motivation seems entirely lacking in the case of my experience of reality. The world, Husserl writes, stands before our eyes as existing "without question" (1995, p. 17); or, in Heidegger's words, it is there "*before* all belief" and never "experienced as something which is believed any more than it is guaranteed by knowledge" (Heidegger, 1985, p. 216). Basically all phenomenological authors would concur on this basic point: rather than being the effect of what I believe it to be, my experience of reality, in its own characteristic way of appearing, determines and makes possible any beliefs I might form about it in turn (see also Merleau-Ponty, 2002, p. 44 *et passim*).

Yet this, according to Husserl, still somehow constitutes a problem. Again, however, one should avoid precipitously identifying this 'problem' with the Cartesian effort of trying to secure the world or any of its objects through some kind of miraculous deduction starting from the immanent sphere of consciousness. The main reason for Husserl's misgivings with regard to such an effort is not so much

that it is a hopeless endeavor (although it is), but that it already betrays one's lack of a proper understanding of consciousness and world-experience. That is, within this Cartesian setup, conscious experience is already pre-understood as a matter of having 'subjective representations' or 'images' that may or may not correspond to the object as it is in-itself standing behind the veil of appearances, thereby opening up the skeptical challenge of knowing whether and how knowledge can extend beyond the prison-house of subjective-relative appearances. This conception of consciousness as some kind of "bag" or "container" where representational contents, to use Dennett's expression, "swim along" (Dennett, 1991, p. 44) is, in Husserl's view, a "fatal mistake" (1999, p. 35).

To give just one example of the 'mistake' Husserl is aiming at: when I visually perceive, say, a chair, almost no one (except, perhaps, some philosophers) makes the mistake of claiming that what we are actually conscious of is some sort of image, sense-datum or another mental representative of the chair. According to Husserl, being conscious of the chair through the presentational consciousness of some other chair-like simulacrum precisely does not count as a perception of a chair, even if this representative image somehow resembles the original chair. This is (but one limited way) how our ordinary experience of everyday objects immediately strikes us, as somehow presenting us with the chair itself. Yet most of us, Husserl included, would agree that objects don't appear before consciousness magically, as - in Sellars' idiom - "a seal on melted wax" (Sellars, 1981, p. 12): at least some sort of process is involved that makes the presence of the object - in the way that it is present - possible. It is precisely this possibility which Husserl, under the heading of his transcendental-constitutive analysis, wants to *understand*: not *if* conscious experience is involved with the world itself rather than its ghostly double, but *how* it is so. Now the 'fatal mistake' - of which Descartes was in Husserl's eyes a paradigmatic offender - occurs when we substitute some part of this process, i.e. that in virtue of which we are conscious of objects in the way that we do, for our consciousness of the object itself. In contemporary terms: when we confuse properties of *what is represented* with properties of the *representing*. To stick with the chair-example: as I visually perceive the chair, my experiences change in various detectable ways. What I perceive of the chair will, amongst other things, depend on my position relative to it. Yet, the chair itself does not appear to get bigger or smaller whenever I get nearer or back away from it, nor does it appear to turn black when I switch off the light. I perceive the chair as having one uniform color, size and shape, while the sensations in virtue of which I perceive those qualities can, and often must, vary with changes in my environment and condition. However, in straightforward perception I do not attend to these sensations, and they are not among the perceptual experience's object. As Husserl (1970) writes: "I do not see color-sensations, but colored things, I do not hear tone-sensations but the singer's song" (p. 559).

Although we will refrain from going further in Husserl's detailed analyses of perceptual experience, Husserl's notorious quarrel with Descartes' theory of ideas on what it is like to perceive a world already reveals that there is nothing specifically 'incurrable' about the phenomenological epoché and its turn to world-as-experienced. So unlike Koch's reading of Descartes, the phenomenological



version of the epoché does not entail any sort of insularity thesis according to which first-person beliefs are immune to error as long as they are restricted to the principle of appearance qua appearance. Put yet another way: not how I think the world appears, but how it actually appears is the principal focus of phenomenological inquiry. And this is precisely what the phenomenological epoché, on Husserl's understanding, should enable us to do: no more and no less than to clarify the implicit sense of what is always already familiar and taken for granted in straightforward experience.

## 2.2 Dennett's heterophenomenology: Cartesianism without a theater

We seemed to have wandered needlessly into the historical vagaries of philosophical dispute between Husserl and Descartes in order to shed a light on an apparently trivial example like that of the phi-phenomenon discussed by Dennett. But, as pointed out endlessly by Dennett himself, seeming is not always being, and the triviality of the example stands in sharp contrast with the wider philosophical ramifications Dennett develops on a certain reading of the phenomenon for the very idea of consciousness and the first-person perspective. So what about this reading?

There is indeed a much simpler and parsimonious approach to the topic of subjective experience compared to the one we have been offering on our reading of Husserlian phenomenology. Dennett is a firm defender of the methodological principle that scientific theories, especially when it comes to such a contested field as consciousness, should avoid multiplying entities beyond necessity (1991, p. 134). By this he means primarily that if a certain gap should arise between how things stand in reality and how they 'seem' from the first-person perspective, one should avoid postulating what he famously calls 'the Cartesian theater': i.e., some kind of inner screen where shadowy representatives of an absent reality are henceforth *really* present before the inner eye of an equally shadowy observer. Yet, as Dennett points out, the idea dies hard and in order to dispel its seductive lure, it is not enough to avoid explicit reference to outdated forms of dualistic metaphysics by, for example, quickly replacing mind-talk with brain-talk. That is, even within the now dominant terms of neurocognitive materialism, the Cartesian specter tends to make its unholy reappearance – under the bastard idea of 'Cartesian materialism' – whenever a central point is presupposed where everything "comes together" and various brain processes present their final product to consciousness (Dennett & Kinsbourne 1992, p. 185). Such an idea is pernicious in a variety of ways: on the one hand, it usually makes for a sloppy functional analysis of how the brain is able to perform its various capacities because such an abstract center tends to clean up all our functional loose ends. On the other hand, wherever there is talk of something being finally presented to consciousness, we can be sure that the idea of a central observer appreciating and observing this product is never too far away. As Dennett (2007) points out with regard to the example visual consciousness:

We are making steady progress on this subpersonal story [...] and we can be confident that there *will be* a subpersonal story that gets all the way from eyeballs to reports and in that story there *will not be* a second



presentation process with an inner witness observing an inner screen and then composing a report. As I never tire of saying, the work done by the imagined homunculus has to be broken up and distributed around (in space and time) to lesser agencies in the brain (p. 11).

This is the crux of Dennett's Multiple Drafts Model, specifically devised as an antidote to the crippling but tenacious legacy of the Cartesian Theater. However, what exactly does it mean to deny the existence of this central 'Oval Office' and to replace it with "various events of content fixation occurring in various places at various times in the brain" (Dennett, 1991, p. 365) for our overall discussion on first-person appearance? Did we, for example, not already see how Husserlian phenomenology left behind such a theater-conception of consciousness quite some time ago for an externalist understanding of consciousness as essentially world-involving? To appreciate in what way Dennett's proposal specifically diverges from the phenomenological tradition, let us see how this model works "in vivo" with the empirical example of the phi-phenomenon.

In response to Goodman's question (cf. *supra*) for a possible explanation of the avowed experience of seeing the moving spot change color before it actually occurred, Dennett considers two possible hypotheses: an Orwellian and Stalinesque version, respectively. That is, either we concoct false memories of *having seen* the spot change color, though in fact we had no prior immediate experience of the change (this is the post-experiential 'Orwellian' revision, suggesting a kind of *ex post facto* rewriting of history); or else we really do have a genuine, albeit delayed, consciousness of the light moving continuously and changing color, due to a kind of unconscious rapid editing process that 'fills in' the missing data and presents the finished product to the eye of the beholder (this is the pre-experiential 'Stalinesque' revision, reminiscent of the cooked-up evidence presented in communist show trials, which does in fact make a genuine appearance in court, despite its fraudulence). Common to both versions is the subject's *belief* in the conscious experience of the moving spot, although they seem to differ substantially in how that belief came about: due to false memories of a veridical experience in the Orwellian case, and accurate memories of a non-veridical experience in the case of the Stalinesque show trial.

Yet, in glossing over these two separate accounts of the phi-phenomenon, Dennett's 'counterintuitive' proposal finally reveals its true color: according to Dennett, the intuition that one of the explanations must be correct to the exclusion of the other is, once again, a clear indication of our persistent captivity in the Cartesian Theater. To be precise: Dennett's claim is *not* that it is impossible to decide between these two hypotheses because we lack sufficient data to determine which one of them offers the most accurate description of the experience, but more radically, that there is *no* empirical difference between these two versions. Postulating a difference between *a real experience* above and beyond or, more properly, before or after what subjects *believe they experienced* is, in Dennett's functionalist idiom, "a difference that makes no difference" (Ibid., p. 125). As Dennett writes:

Consider how natural is the phrase “I judged it to be so, because that’s the way it seemed to me”. Here we are encouraged to think of two distinct states or events: the seeming-a-certain-way and a subsequent (and consequent) judging-that-it-is-that-way. [...] Some people presume that this intuition is supported by phenomenology. They are under the impression that they actually observe themselves judging things to be such as a result of those things seeming to them such. No one has ever observed any such thing “in their phenomenology” because such a fact about causation would be unobservable [...] (Ibid., p. 133).

Hence, in clear opposition to the phenomenological account, the conclusion Dennett wants us to draw from the numerous experiments discussed in his *Consciousness Explained* is that there is no difference between how things *seem* to us and how we *think* they seem. Bringing this all together, Dennett consequently dubs his Multiple Drafts Model as a *first-person operationalism* “for it brusquely denies the possibility in principle of consciousness of a stimulus in the absence of the subject’s belief in that consciousness” (Ibid., p. 132).

Needless to say, such a conception of the first-person perspective as entirely constituted by subjective beliefs immediately carries some remarkable consequences for the scientific effort to devise a proper naturalistic account for this dimension. These become most salient in Dennett’s own methodology of choice for approaching the scientific study of human consciousness, his *heterophenomenology*, advertised as an explicitly *third*-person approach to the issue of subjectivity (Dennett, 2003). On the face of it, this seems like a remarkable thing to do: subjectivity is traditionally regarded as occupying something rather like another dimension – the private, ‘first-person’, more or less self-consciousness sort of being – in comparison with the privileged objects of “third-person” methodologies like meteors, sound waves or bone density. This ostensible difference between meteors and minds has led some phenomenological authors, like e.g. Evan Thompson, to claim that phenomenology, as an explicitly first-person investigation, is in fact an inescapable prerogative for any naturalistic account because “any attempt to gain a comprehensive understanding of the human mind must at some point confront consciousness and subjectivity – how thinking, perceiving, acting and feeling are *experienced in one’s own case*” (Thompson, 2007, p. 16, our emphasis; see also Gallagher, 2007; Feysaerts & Vanheule, 2015). The reasoning behind such a claim is relatively straightforward and is meant particularly as a warning sign for Dennett’s – but also others’ – overambitious naturalistic inclinations: any explanatory scheme which departs so completely from first-person experience that it calls into doubt the very existence of such experiences can no longer claim to be elucidating that explanandum, since it has effectively ‘eliminated’ the very object it purports to be an account of. Yet Dennett is not impressed. In an essay addressing some misunderstandings concerning his apology for his heterophenomenology with the little concealing title “Who’s On First?”, Dennett asks:

Can the standard methods be extended in such a way as to do justice to the phenomena of human consciousness? Or do we have to find some quite radical or revolutionary alternative science? I have

defended the hypothesis that there is a straightforward, conservative extension of objective science that handsomely covers the ground – *all* the ground – of human consciousness, doing justice to all the data without ever having to abandon the rules and constraints of the experimental method that have worked so well in the rest of science (2003, p. 19).

Heterophenomenology is the method of choice whenever one wishes to take the first person point of view serious – that is, as serious as it *can* be taken – while keeping the methodologically scrupulous requirements undergirding the same objective perspective that reigns in those sciences that aspire to explain different, but from a methodological point of view, no more daunting phenomena such as continental drift, biodiversity or the digestive tract. So what then, exactly, does Dennett’s heterophenomenology commit us to? What precisely does it mean to take subjects serious, but not too serious?

### 2.3 Return of the Cogito: serious, all too serious

Given that we have to adopt a strict third-person perspective, since as Dennett is keen to say, “*all* science is constructed from that perspective” (1991, p. 68), the only access to the phenomenological realm will be via the observation and interpretation of publicly observable data, including utterances and other behavioral manifestations. Furthermore, we will have to find a neutral way of describing these data – that is, a way that doesn’t prejudge the issue – by adopting a special form of heterophenomenological abstinence. Such an abstinence – likened by Dennett to a third-person version of Husserl’s epoché (2003, p. 22) - brackets the issue about whether the subject is in fact a computer, a mere zombie, a dressed-up parrot or a real conscious being. The only thing such a third-person epoché *does* commit us to is adopting what Dennett calls *the intentional stance*: we should interpret the vocal sounds emanating from the subjects’ mouths as speech acts that express the subject’s beliefs in order to compose a text of what that subject wants to say about his or her own conscious experiences. As Dennett (1991) explains:

People undoubtedly do believe they have mental images, pain, perceptual experiences and all the rest, and these facts – the facts about what people believe, and report when they express their beliefs – are phenomena any scientific theory of mind must account for. We organize our data regarding these phenomena into theorists’ fictions, “intentional objects” in heterophenomenological worlds. Then the question of whether items thus portrayed exist as real objects, events, and states in the brain – or in the soul, for that matter – is an empirical matter to investigate (p. 98).

Again, this merely seems like a fairly trivial rehearsal of the principle of scientific neutrality: from the fact that people believe all kinds of things, even when these things happen to refer to their own

experience, we shouldn't immediately conclude that these things actually exist. As a case in point, take the example of visual experience. People often seem to believe that their visual experience represents the world in sharp focus, uniform detail and high resolution from the center out to the periphery, as a kind of snap-shot like picture of the world (Noë, 2004). Yet as Dennett points out, convincingly, such a view of visual experience as gap-free and wholly continuous is mistaken. Take the example of looking at a wall: while staring at it, it looks to you as if there is unbroken expanse of wall. However, this is not to say that it seems to you as if, in a single saccadic fixation, the whole of the wall's surface is presented in equally rich detail from the center out to the periphery, as if every part of the wall's surface is immediately presented to consciousness at once. Rather, as Husserl would say, the wall is 'transcendent' to my experience insofar as it has more to it than what is revealed to me in any single phase of experience and, moreover, is precisely *given* or *experienced* (as opposed to inferred or deduced) *as* having more to it. So, as the example illustrates, we would do well to follow Dennett's advice by not taking everything a subject believes about his or her own experience at face value. On this account, taking subjects' reports and beliefs serious means that we take into account that they can be mistaken; if not, that is, if doubt with regard to conscious experience would be strictly speaking senseless, then – as Wittgenstein famously said - so too would any assertion of knowledge (1958, p. 211).

However, despite Dennett's repeated insistence on our pervasive fallibility with regard to first-person experience as a central argument for the adoption of his heterophenomenological stance, the question is how this methodological modesty squares with his aforementioned first-person operationalism. How to reconcile the by all parties acknowledged fact that we can be wrong about first-order experience with Dennett's thoroughgoing fictionalization of subjectivity up to the point where we have "total, dictatorial authority over the account of how it seems to you, about what it is like to be you" (Dennett, 1991, p 96)? To put it in the language of Dennett's own cherished functionalism: is fallibility a difference that can make a difference for the defense of heterophenomenology when there is no possibility of consciousness in the absence of a subject's belief in that consciousness? It must be clear by now that it cannot, and furthermore, that this inability to make room for a distinction between appearance and reality within the field of conscious appearance is itself a remarkable and, as per the defense of heterophenomenology, undesirable consequence of what we would call Dennett's own doxastic Cartesianism. For if my subjective experience is entirely constituted by what I believe to be the case, or what I believe that seems to be the case, then there is of course – and as we made clear in our reading of Descartes - no 'case' about which I can be wrong either. Consequently, Dennett is confronted with the following paradox:

1. either make room for the distinction between experience and judgment in which case the plea for heterophenomenology can be saved but at the cost of first-person operationalism;
2. or retain the latter, but at the cost of being unable to countenance fallibility and hence of introducing an unavoidable return to the philosophical core of Cartesianism, although this time without a theater.

Yet secondly, apart from this general philosophical tension, a further problem for Dennett's doxastic account arises when we raise the further question about how it construes the relation between belief and experience. As we have seen, the central strategy of Dennett's heterophenomenology is to attribute a belief to the subject whenever he or she reports an experience, rather than the other way around. And indeed, it has been a traditional argument ever since Descartes and Kant that an 'I think' or 'I believe' accompanies, or at least in principle *must be able* to accompany, any experience I might claim to be having, for otherwise, as the famous quote from Kant continues, "something would be represented in me that could not be thought at all, which is as much as to say that the representation would either be impossible or else at least would be nothing to me" (1996, B131-32). Importantly, Dennett trades on this ineliminable occurrence of the 'I think' within reflective consciousness to advance his first-person operationalism and thus to deny the possibility that there could be any difference between beliefs and experience. Since, as Dennett's argument goes, if you have conscious experiences you don't believe you have – those experiences would be just as inaccessible to you as to the heterophenomenological observer. Conversely, if you believe you have conscious experiences you don't in fact have, it is your beliefs that need to be explained, not the non-existent experiences (Dennett, 2003, p. 21). Hence, the crucial question becomes: is it really necessary for a subject while enjoying or undergoing an experience, to immediately believe that he is enjoying or undergoing that experience?

It is safe to say that one of the grounding motives in the phenomenological works of most notably Sartre, Merleau-Ponty and Heidegger is a forceful critique of this intellectualistic idea. The claim has been that there are indeed conscious intentional experiences that a subject could have without immediately believing that he is having them. Even more, those experiences precisely depend on the subject's *not* having a belief about them: if the subject were to believe he was having the experience, instead of merely living it through, the experience would become impossible (see also Dreyfus & Kelly 2007). Differently put: a necessary condition for those types of experience is precisely the *absence* of the Cartesian spectator. A famous example are of course the following ones by Sartre and Merleau-Ponty:

When I run after a streetcar, when I look at the time, when I am absorbed in contemplating a portrait, there is no I. There is consciousness *of the streetcar-needing-to-be-overtaken*, etc., and non-positional consciousness of consciousness. I am then plunged into the world of objects; they constitute the unity of my consciousness, they present themselves with values, with attractive and repellant qualities – but me, I have disappeared (Sartre, 2004, p. 32).

We must recognize first of all that thought, in the speaking subject, is not a representation, that is, that it does not expressly posit objects or relations. The orator does not think before speaking, nor while speaking,

his speech is his thought. [...] The end of the speech or text will be the lifting of a spell. It is at this stage that thoughts on the speech or text will be able to arise (Merleau-Ponty, 2002, p. 209).

Central to such examples is that the subject, to put it in Lacanian terms, is ‘internally excluded’ from the experience while living them through: in keeping an appropriate distance while talking to people, in reaching automatically for the proffered handshake, laughing while listening to some amusing story, rapidly walking down the stairs or backing away from a painting to get a better view, there is a certain solicitation from the world to engage in various activities without us having to decide to do so. Moreover, reflexively deciding, contemplating or believing in such activities is the best way to lose all grips on what the situation calls for. Such an experience of spontaneous immersion or absorption “into the world of objects” only arises precisely when we are not looking for them; conversely, they would vanish immediately insofar I would try to think or believe them. Yet, *pace* Dennett, this does not mean that in those instances I momentarily become a zombie or unconscious machine; there still is something ‘it is like’ to be absorbed in these activities which set them apart from, say, a comatose condition, and the absence of belief is one example of its constitutive features. It is precisely this “internally excluded” character of belief and experience in such phenomena which Dennett’s doxastic Cartesianism cannot countenance.

### **3. Dennett and psychoanalysis: the curious case of the ‘objectively subjective’**

Before we can begin to specify in what way psychoanalytic theory could bear on this debate between Dennett and phenomenology, some precursory remarks are in order. Because, first of all, it might not be entirely clear where, if at all, we should situate psychoanalysis, whether in its original Freudian formulation or in its Lacanian version, along Sellars’ somewhat contrived, but perhaps still hermeneutically useful dichotomy. For in contrast to phenomenology’s straightforward and programmatic insistence on first-person experience as an indispensable datum for any sort of naturalism, the claims and subject matter of psychoanalytic theory are in this regard, to put it mildly, somewhat more ambiguous. Of course, it might be argued that psychoanalysis is not really that interested in, nor in itself capable of, pursuing or resisting the naturalistic agenda of bridging the gap between neurocognitive and phenomenological accounts of blind sight or the rubber-hand illusion. And indeed, one could say, what would be further removed from psychoanalysis’ basic insight into man’s constitutive *disadapted* relation to its ‘natural’ environment than Dennett’s talk of the design stance? Yet such a first and perhaps heartening impression is easily discarded when we shift the terrain to those phenomena, however ill designed, that are considered to be at least more amenable to psychoanalytic elucidation, such as the nature of psychopathology, sexuality, language or desire. With regard to those, the question emerges whether or not psychoanalysis would simply join in with Dennett’s deflationary tactics of reducing the several issues that may arise in the effort of articulating the intertwinement of

subjectivity and language - for example, the relation a subject entertains with its own speech – or subjectivity and desire – for example, what it means to desire the desire of the other - to the austere categories of ‘seeming’ or ‘fiction’? Presumably, the pre-reflexive answer would consist in a quick and resounding ‘no’.

Yet, on the other hand, given Freud’s inaugurative departure from Brentano’s views on intentionality and self-consciousness (see Seron, 2015, pp. 15-30; Bernet, 2002) or Lacan’s unreserved critique of both Sartre’s existential psychoanalysis (Lacan, 2006, p. 80) and Merleau-Ponty’s bodily unconscious (Lacan, 1982-1983, p. 75), it is equally clear that the intellectual relation between phenomenology and psychoanalysis is complicated enough to avoid any easy wedding of psychoanalysis with manifest experience. Given this ambiguous position, a clarification of how psychoanalysis relates to the ongoing ‘scientific disenchantment’ along the terms set by Dennett’s ingenious codification of the claims and methods of cognitive science, and conversely, to the phenomenological arguments that are being advanced against Dennett’s proposal, indeed becomes a reasonable task.

### 3.1 Freud and Lacan between phenomenology and cognitive science

To begin with some sweeping examples: what exactly is left of the once so seductive Freudian image of the three successive humiliations of man-in-the-world (Copernicus-Darwin-Freud) now that we can safely add a fourth one, -Dennett? As noted by several commentators (Gardner, 2001; Zizek, 2006), it seems that Freudian psychoanalysis, steeped as it is in the intentional language of repression, desires and negation – now finds itself amongst those traditional ‘manifest’ philosophies threatened by the more recent neurocognitive humiliations. And the reason for this clear: though, in the felicitous wordings of Freud, “psyche is extended; knows nothing about it” (Freud, 1938, p. 300), there has never been a lack of critics pointing out that what it does not know is still suspiciously close to the philosophical problematic of intentional consciousness. And it is indeed striking and ironic, given psychoanalysis’ original declaration of independence from any philosophy issuing from the cogito (see Freud, 1923, p. 13; Lacan, 1949/2006, p. 75), that one of the standard ways of defending the unconscious today is precisely by referring to the properties of consciousness, of which intentionality is held to be pre-eminent. As Freud (1915) himself indeed pointed out:

All the categories which we employ to describe conscious mental acts, such as ideas, purposes, resolutions, and so on, can be applied [to unconscious mental acts]. Indeed, we are obliged to say of some of these latent states that the only respect in which they differ from conscious ones is precisely in the absence of consciousness (p. 168).



Turning to Lacan, we can discern the same manifest motives despite the latter's proto-cognitivist, structuralistic inclinations: even on the most cursory reading of the *Ecrits*, it would be difficult to avoid the perennial 'philosophemes' of subject, alienation, desire, Other and so on. But Lacan, of course, would try to avoid any quick philosophical (mis)recognition by reminding his readers to this subject is split, barred and divided by the signifier, therefore not the phenomenological subject of *Jemeinigkeit*, always already conscious of itself in its intentional involvement with the world. Nevertheless, however split or deferred this subject might be, this failed or misfired cogito is still close enough to its Cartesian predecessor so as not to completely break with it. On the contrary, Lacan only ever wanted to think the unconscious, not as antithetical to consciousness, not as "a species defining the circle of that part of psychical reality which does not have the attribute (or the virtue) of consciousness" (Lacan, 1964/2006, p. 703), but precisely as the implicit truth already contained in the Cartesian *cogito me cogitare* (see Dolar, 1998). But this also means that this truth, whatever else it may be, is the truth of that same old Cogito, gained through a certain radicalization of that traditional aporia, confronted differently by Husserl, Fichte or Sartre, of conceiving subjectivity from and within its own terms. Therefore, in so far as phenomenology is susceptible to Dennett's heterophenomenological recasting of subjectivity in terms of a 'theorist's fiction', the same holds for Lacan's particular defense of the subject. And is it indeed not quite imaginable, for example, that Dennett's response to Lacan's notorious anti-humanistic formula in *Science and Truth* according to which "science's man does not exist, only it's subject does" (1966/2006, p. 730) would be the sobering 'nor does it's subject, although it might seem to'?

On the other hand, however, it could be pointed out that we might have proceeded a little bit too quickly and carelessly in lumping together psychoanalysis with the manifest philosophies of intentional experience. And in fact, as is well known, both Freud and Lacan have been taken to task for neglecting and underestimating the accomplishments of phenomenology in a way that is, although not strictly equivalent, remarkably close to the contemporary phenomenological arguments raised against Dennett (see Schafer, 1976; Chalmers, 1996). Suffice to recall, in Freud's case, the succinct phenomenological critique of the psychoanalytic unconscious by Eugen Fink in the appendix to Husserl's *Crisis*:

This objection, which is raised in many variations against the so-called 'consciousness-idealism of phenomenology,' is based on a fundamental philosophical *naïveté*. [...] The *naïveté* we refer to consists, before all theory-construction about the unconscious, in an *omission*. One thinks one is already acquainted with what the 'conscious', or consciousness', is and dismisses the task of first making into a prior subject matter the concept against which any science of the unconscious must demarcate its subject matter, i.e., precisely that of consciousness. But because one does not know what consciousness is, one misses in principle the point of departure of a science of the 'unconscious'. (Husserl, 1970, p. 386)



Despite the shifting terminology, it's not too difficult to recognize in this depiction of psychoanalysis' 'naïveté' the same point of contention we already encountered in Dennett's case. That is, Freud is held to have tried to develop a naturalistic third-personal theory of mental functioning, couched in broadly functionalistic-energetic terms (the dynamic trinity of 'ego', 'id' and 'superego'; the energetic metaphors of cathexis, displacement, condensation, and so on), destined to explain various pathological phenomena, without however offering any explicit and sustained consideration of "consciousness". As noted by Tugendhat (1986, pp. 131-132) and Ricoeur (1970, p. 420 *et passim.*), Freud indeed never seems to talk about phenomenal consciousness and self-consciousness directly, but only and rather loosely in terms of the 'ego' which is portrayed as yet another homuncular-functional entity amongst others. A similar story can be told with regard to the notion of 'repression' and other 'mechanisms' of defense of which Freud gives an extended account in terms of word- and thing-presentations, cathexes and so on, again strongly suggesting that these concepts designate sub-personal operations of some kind. Accordingly, repression, displacement, condensation and other similar notions, should not be conceived as first-personal actions, whether conscious or unconscious, operating within a normative framework of beliefs and desires, but as functionally specifiable processes requiring no agents, experiencers or observers whatsoever. In short: on this reading, psychoanalytic theory turns out to share roughly the same agenda as cognitive neuroscience in trying to offer a homuncular-functionalist top-down analysis of mental functioning in terms of various subsystems until we reach the level of non-intelligent functionaries that can be replaced by a machine (see Freud 1938, p. 144).

As for Lacan, although one of the remarkable features of his self-proclaimed return to Freudian orthodoxy is precisely the explicit avoidance of all those orthodox-naturalistic elements, this didn't necessarily seem to imply a return to a consideration of *l'expérience vécue* at the expense of Freud's naturalism. On the contrary, Lacan's structuralist-materialist reworking of Freudian metapsychology in terms of his "logic of the signifier" appears to have been driven by an even more fierce ambition to finally establish psychoanalysis as a rigorous, because aspiring to the Koyréan ideal of mathematical formalization, scientific endeavor (see Milner, 1995; Johnston, 2012; Tomsic, 2012). And indeed, after some years of flirting with phenomenological accounts, the famous 1953 *Discourse of Rome* can be seen as a veritable trumpet blast fiercely announcing the new footings on which to resurrect the psychoanalytic edifice amongst the "sciences of subjectivity" (1953/2006, p. 235):

The form of mathematicization in which the discovery of the *phoneme* is inscribed, as a function of pairs of oppositions formed by the smallest graspable discriminative semantic elements, leads us to the very foundations that Freud's final doctrine designates as the subjective sources of the symbolic function in a vocalic connotation of presence and absence. And the reduction of any language (*langue*) to a group comprised of a very small number of such phonemic oppositions, initiating an equally rigorous formalization of its highest-level morphemes, puts within our reach a strict approach to our own field (Ibid., pp. 235-236)

This is indeed a far cry from Lacan's earlier phenomenological declaration that "language, prior to signifying something, signifies to someone" (1936/2006, p. 66). Here we are incited to the view that language, loosely defined in terms of the symbolic order as somehow pertaining to both the regulatory principles of social exchange and the material-phonetic level of 'pure signifiers' (on this equivocation, see Descombes, 1979), is "a form in which the subject is inserted at the level of his being" (Lacan, 1955-1956/1993, p. 179). In the *Seminar on "The Purloined Letter"* and *The Instance of the Letter*, gloomy images are being invoked to convey the radical nature of this 'insertion'. Lacan (1956/2006, p. 21) talks about "machines-that-think-like-men", subjects "more docile than sheep", lined up before the gripping effect of the signifying chain as "ostriches", determined by the signifier's displacement in their "acts, destiny, refusals, blindnesses, success, and fate" (Ibid., pp. 29-30). How all of this is supposed to be an effect of 'symbolic determination' is never really convincingly spelled out (Dennett's anti-mysterian concept of "wonder tissue" might be put to good use here; see Dennett, 2013, pp. 198-205), nor is it clear if and in what way Lacan's structuralistic remolding of Freudian concepts (e.g. metaphor/metonymy for condensation/displacement, repetition of the signifying chain for *Wiederholungszwang*, ...) is really an advance compared to the latter. Yet, for our present purposes, at least it is clear that on this reading, Lacan's intent was to purify the Freudian unconscious even further from any remaining residues of metaphysical humanism and its associated anthropomorphic terminology. Hence, as we have noted in the case of Freud, Lacan's structuralistic rearticulation of the unconscious could therefore equally well be seen as a forerunner to the sort of materialist program better known today under the heading of cognitive science (see also Dupuy, 2000).

### 3.2 Psychoanalysis' 'abominable mess'

Now this Janus-faced character of psychoanalysis has of course never slipped the attention of its various commentators, provoking blame and rejection from its critics; praise, confusion or silent endorsement from its defenders. On Sartre's diagnosis, for example, psychoanalysis is precisely founded on a systematic confusion of the personal and sub-personal levels and this constitutes its unavoidable yet inadmissible paradox: "I both *am* and *am not* the psychic facts hypothesized by psychoanalysis" (Sartre, 1958, pp. 50-51). In roughly the same way, Wittgenstein famously spoke of psychoanalysis' "abominable mess" (Wittgenstein, 1978, p. 316), premised as it is on a confusion between first-personal reasons and third-personal causes, imposing the former even if the patient doesn't accept them, resorting to the latter and being content with the patient's acceptance for their justification (Cioffi, 1998).

Similar accusations apply to Lacan's 'materialism of the signifier', representing yet another confused effort to straddle both sides of this fence at once: the signifier being regarded as both amenable to a 'materialist' description in terms of its meaningless phonemic properties and an 'idealist' description in terms of its meaningful semantic properties. By cutting the cord between signifier and signified, and in

the same move declaring the primacy of the former over the latter, Lacan thought to be able to transmute these ‘pure signifiers’ into causal forces operating in a mechanical way, yet responsible for the different meaning effects encountered in psychoanalytic practice. This is, for some, the very formula of magic (Descombes, 1999, p. 95).

So what then are we to make of this confusion? There are of course some well-trying reconciliatory attempts which basically consist in either translating psychoanalysis into phenomenological terms<sup>51</sup>, or, embracing it precisely as a form of sub-personal psychology<sup>52</sup> (see Gardner 2000, for a detailed discussion), yet perhaps the more interesting thing to notice is the recurrent image from which such solutions spontaneously spring. By this we mean, as regards the present discussion, the very form in which the difference between Dennett and phenomenology is formulated: that is, in terms of ‘fictional beliefs’ (Dennett) and ‘manifest experience’ (phenomenology). In the following section, we will therefore try to explicate in what way psychoanalytic theory could distinguish itself from both these readings.

### 3.3 Decentering appearance: belief in belief

Let us begin this section by recapitulating in a condensed form the core disagreement which stands behind the discussion between Dennett and phenomenology. Whether it is the topic of perceptual consciousness, imagination, sense of agency, the painfulness of pain or the problem of other minds, is not really that essential here: the crucial divergence already appears at the level of “the organization of the data” (Dennett, 2003, p. 9), that is, the way in which researchers should proceed whenever they set out to explain those phenomena of consciousness.

In his *Who’s On First?* (2003), Dennett offers a clear summary of his position that can put the problem in sharper focus. “What”, Dennett asks, are the “primary data” from which we ought to depart in order to arrive at “a science of consciousness” (p. 21), i.e.: (a) conscious experiences themselves; (b) beliefs about these experiences; (c) verbal judgments expressing those beliefs; or (d) utterances of one sort or another? Whereas, as we have seen, phenomenology vigorously opts for (a); Dennett’s heterophenomenology is premised on the elimination of (a) in favor of (b) and (c). From this disagreement, it necessarily follows that the philosophical concepts of ‘appearance’ and ‘seeming’ function as a kind of *shibboleth* in the ensuing debate: designating ‘immediate experience’ in the case of phenomenology, ‘beliefs about experience’ in Dennett’s account. Furthermore, it becomes obvious why the way in which arguments are being pressed for one position or the other, take on the form in which we have argued for some of the claims found in phenomenology: i.e. those cases that point

---

<sup>51</sup> This is indeed the proposal we can read in the works of Sartre and – even earlier – Politzer: to release psychoanalysis from its “mechanistic cramp” (Sartre 1979, p. 32) or “abstract” and “realist” exigencies (Politzer 1994 – for an excellent overview of Politzer’s somewhat forgotten critique, see Bianco 2016)

<sup>52</sup> For a defense of this reading, see, e.g., Moore (1988), Kitcher (1992) and Livingstone-Smith (1999).

towards a *disjunction* between belief and experience – that is, on the one hand, ‘beliefs without experience’ as in the discussion about the level of detail found in visual consciousness; on the other hand, ‘experience without belief’ as in Sartre’s example of ‘pre-reflexive immersion’.

From a psychoanalytic perspective, however, what should immediately strike us is not so much this manifest level of disagreement in terms of the difference between ‘belief’ and ‘experience’, rather, it is the implicit Cartesian agreement between Dennett and phenomenology on the concept of *belief* itself that should interest us here. To put it in Spinozistic terms, there is still, however slightly, something sticking out as ‘a kingdom within a kingdom’ throughout this whole discussion, something which becomes most explicit in Dennett’s response to some of his critics – Schwitzgebel (2007) and Noë (2007) - in his *Heterophenomenology Reconsidered* (2007). As before, the critique involves Dennett’s seemingly incompatible insistence on both the essential fallibility of first-person accounts and some kind of Rortian incorrigibility with regard to “how it seems to them” (p. 263). In the same volume, Noë (2007) evokes the familiar Wittgensteinian theme on the logical limits of justification of first-person accounts which comes to an end rather sooner than we might think:

It [Wittgenstein’s view on justification] applies to the special case where what is in question is how I take my own experience to be, what I take it to be like. I *can* be mistaken about the nature of my experience – about how I, in experience, take things to be. [...] But it would be a different kind of mistake for me also to be mistaken about how I *take* my experience to be. I can be wrong then about how things seem but not wrong about how I take things to seem (p. 244).

Dennett (2007) wholeheartedly agrees:

If somebody says her visual field *seems* detailed all the way out to the periphery, which lacks a perceptible boundary, there is no gainsaying her claim, but if she goes on to theorize about “the background” [...], she becomes an entirely fallible theorist, no longer to be taken at her word (p. 263).

So, at least when it comes to this notion of ‘taking’, of what I believe my experience ‘is like’, phenomenology and Dennett prove themselves to be something like an “epistemological pair” (Bachelard 1949, p. 5). In short: whereas we can be wrong about ‘seeming’ understood in terms of experience, we cannot however be wrong about ‘seeming’ in terms of belief. Now the proper psychoanalytic questions that should arise in this context are the following ones: whence this sudden confidence? Which arguments are being advanced here to uphold this last inviolable sanctum of minimal infallibility with regard to this Cartesian theme of ‘belief in belief’? Why is it so obvious that ‘immunity to error’ immediately ensues whenever we shift the terrain to the topic of belief itself? Is it because, in this case, the well-known phenomenological figure of the ‘unbeteiligte Zuschauer’ suddenly reappears

on the scene? If so, why would that be? And why precisely at this point? Surprisingly, precious little can be found on these questions when we consult these various commentators.

Nevertheless, here, as everywhere else, it could be wise to stick to Althusser's famous "golden rule of materialism" to "not judge a being by its self-consciousness (Althusser, 1996, p. 115). Particularly interesting in this context are the analyses we can find in the works of Žižek (1998; 2006; 2014) and Pfaller (2014; 2015) on the psychoanalytic concept of 'belief'. Elaborating on and extending Octave Mannoni's perspicuous theory of fetishism (see Mannoni, 2003, pp. 68-92), both these authors have drawn attention to a series of entirely common phenomena which nevertheless share some striking features that are directly relevant for our overall discussion. To start with some of these examples:

Imagine you're sitting in a bar reading a newspaper, waiting for a friend. The friend arrives. He says hello, and then continues: 'Excuse me, can I have a quick look at your newspaper? I know it's silly, but I just have to know the score from yesterday's game' (Pfaller, 2014, p. 1).

For reasons that are no doubt suspect, but hidden, I sometimes read the rather rudimentary horoscopes published in certain papers. It seems to me that I do not take much of an interest in them. I wonder how people can believe in them. [...] Once, last year, my horoscope said that "tomorrow will be an extremely favorable day for tidying up the house". This was not a spectacular prediction, except that I had long been planning to move on the day in question. I burst out laughing at so funny a coincidence. [...] I could say that I am not superstitious because I pay no mind to such things. But, to be precise, I should rather say: I know quite well that coincidences of this kind are meaningless, but I take a certain amount of pleasure in them all the same (Mannoni, 2003, p. 78).

What is common to these examples – and numerous others (see Žižek, 1998; Pfaller, 2014, for an extensive collection) – is the paradoxical relation a subject seems to entertain with its belief (i.e. the importance of sports; the significance of horoscope predictions). In a way that is difficult to render conceptually transparent, these sorts of beliefs are never really believed in ("I know it's silly" – "I know quite well that these coincidences are meaningless"), yet in one way or another, they nonetheless exert a particular influence on their subjects (the compulsion to look at the paper – the pleasurable laughter). Moreover, as Pfaller argues, it is not so much their *content* that is primarily of interest here, but rather the *form* in which people refer to these beliefs. That is, as was already apparent from Mannoni's formula for these types of beliefs ("I know quite well ... but still"), its form is characterized by a complex coexistence of 'better knowledge' and 'belief'. The fidget sports fanatic knows quite well that yesterday's results are not important, but still he *has* to see them. Despite the better knowledge, and despite the gap this places between him and his silly practice, he nevertheless acts *as if* sports are of utmost importance. Similarly, horoscope predictions seem rather ridiculous to Mannoni, nonetheless, as he admits: if the horoscope had said that "tomorrow will be an extremely unfavorable day for moving", "it would have made me laugh differently" (2003, p. 78.).

So in an initial approximation of the formal structure of this phenomenon, one should say that, on the one hand, these are beliefs that are never really one's own beliefs in the sense that one would claim ownership over them (à la "I believe in horoscopes, I really do"). The 'better knowledge' is immediately invoked to signal such a subjective position of transcendence towards the belief. Yet, different from, for example, the Sartrean formula for transcendence – i.e. "To believe is to know that one believes, and to know that one believes is no longer to believe" (Sartre, 1958, p. 114) – such a better knowledge seems to be of little avail with respect to the efficacy of the belief in question. On the contrary, as Mannoni points out, one should even wonder if this position of knowledgeable transcendence does not somehow *contribute* to the maintenance of the disavowed belief? Such a claim would immediately contradict the seemingly incontestable principle according to which the dialectical relation between 'knowledge' and 'belief' is grounded in *exclusion*: on such an account, we (erroneously) believe (in the importance of sports, in the predictive validity of horoscopes, ...) precisely because we do not know (that sports are not important, that horoscopes do not possess predictive value, ...); conversely, better knowledge cancels out what we previously believed. Here, on the contrary, the dialectical relation between knowledge and belief would be based on the principle of *necessary conjunction*: only on the condition of better knowledge are we susceptible to those beliefs; conversely, in the absence of better knowledge, believing becomes impossible – as Mannoni succinctly puts it: "Evidently, the sole reason for the "but all the same" is the "I know well"" (2003, p. 71). If this principle is somehow sustainable (cf. *infra* for further argumentation), we can now proceed to advance a more precise formulation of this dialectic in terms of our general discussion: on a psychoanalytic account, the way things appear to me, that is, of what I "objectively believe" (cf. Žižek, 1989, p. 35) - as evidenced, despite myself, in my nonsensical behavior, spontaneous reactions and utterances (e.g. encouraging my car by means of little compliments when it won't start, wearing my 'lucky shirt' for an important interview which I did prepare for in a 'professional' way,, ... ) - as opposed to what I believe to believe, is structurally dependent on the self-reflexive movement of intellectual distancing through 'knowing that one knows'.

### 3.4 Belief in the uncanny or uncanny belief?

Now, in order to substantiate this rather strong claim, what we seem to need is a methodological procedure closely resembling the phenomenological tool of "eidetic variation" (see Husserl 1977, sec. 9a) in and through which, formally speaking, the necessary conditions of a certain phenomenon are established by 'varying' those conditions up to the point where the phenomenon itself would become (in)conceivable. Now it happens that a procedure of this kind is to be found implicitly in Freud's 1919 treatise on *The Uncanny*, a phenomenon which, moreover, bears some striking resemblances to the examples discussed above (see Dolar, 1991; Župancic, 2005; Pfaller, 2006). Regardless of the wealth of literary examples and psychoanalytic concepts Freud employs throughout the article, this dense and, to a certain extent, inconclusive text, essentially revolves around one central effort: to circumscribe the

“common core” which would allow us “to distinguish as ‘uncanny’ certain things which lie within the field of what is frightening” (1919, p. 219). Freud’s question is: how should one proceed in defining the common denominator in such a variety of examples as the mechanical doll Olympia, the obscene figure of the Sand-Man, the recurrent theme of the “double”, the “repetition factor” or the Rat Man’s “omnipotence of thoughts” – all taken as typical instances of the uncanny (Ibid.)? Why are they not simple examples of garden variety phobias such as the fear of heights or crowded places? What is it that sets them apart from the more epistemologically innocent logic of cognitive surprise or unfulfilled anticipation? And, finally, why should they even produce this peculiar affective response rather than none at all? As an instructive introduction for a further demarcation of the problem, Freud offers the following, general formula:

[...] an uncanny effect is often and easily produced when the distinction between imagination and reality is effaced, as when something that we have hitherto regarded as imaginary appears before us in reality [...] (Ibid., p. 242).

Accordingly, we seem to have three relevant conditions – (1) the uncanny affect; (2) imagination; (3) appearance. First, the feeling of uncanniness should be distinguished from more common affective responses such as surprise, fear or mere indifference. Nor is it merely the univocal opposite of what is familiar, in the sense of ‘homely’, ‘cozy’ or ‘intimate’, since not everything which entails a negation of the familiar in this sense will provoke the feeling of the uncanny. As Freud argues, “something has to be added to what is novel and unfamiliar in order to make it uncanny” (Ibid., p. 221). Secondly, what it is that should be added seems to be contained in the two other conditions: appearance and imagination. Freud illustrates this point with an example taken from his case study of the Rat Man:

In the case history of an obsessional neurotic, I have described how the patient once stayed in a hydropathic establishment and benefited greatly by it. He had the good sense, however, to attribute his improvement not to the therapeutic properties of the water, but to the situation of his room, which immediately adjoined that of a very accommodating nurse. So on his second visit to the establishment he asked for the same room, but was told that it was already occupied by an old gentleman, whereupon he gave vent to his annoyance in the words: ‘I wish he may be struck dead for it’. A fortnight later the old gentleman really did have a stroke. My patient thought this an ‘uncanny’ experience (Ibid., p. 239).

Here we have all the conditions combined: (1) the uncanny fearful experience, (2) the imagination pertaining to the obsessive wish and (3) reality appearing as if it immediately responded to what was “hitherto regarded as imaginary” (Ibid., p. 242). At this point we can also introduce our previous dialectic between ‘better knowledge’ and ‘belief’. Freud’s first decisive move is to exclude Jentsch’s hypothesis of intellectual uncertainty as an essential factor for the production of the Rat Man’s uncanny



experience and seemingly absurd guilt feelings. On such a cognitive account, the Rat Man would be struck by the uncanny precisely because he was, for a brief moment, a spontaneous theoretician of the ‘omnipotence of thoughts’, intellectually unable to rid himself from the illusion that ‘wishes can truly kill’. And the reason for this is precisely because reality appears to correspond with what was held to be impossible: better knowledge is suspended because of the qualities of the appearance. Conversely, lifting the veil of appearances would mean to break the spell of the uncanny. However, not only would this idea square rather badly with another of Freud’s remarks on the peculiar epistemic position of the Rat Man as being different from “the superstition of uneducated people who feel themselves at one with their belief” (Ibid., p. 229), it would also mean that, in the case of Hoffman’s story, our uncanny experience would be due as well to our intellectual uncertainty as regards the true status of the Sand-Man. Yet, as Freud remarks, the story makes it quite clear that the Sand-Man is not really the Sand-man, nevertheless, “this knowledge does not lessen the impression of uncanniness in the least degree” (1919, p. 230). Furthermore, it can even be said that the logic of the uncanny is directly opposed to the logic of intellectual uncertainty. As Dolar (1991) points out, the most uncanny thing is not so much that we are kept uncertain as to the true nature of events, whether or not what happens is really due some unfathomable scenario. On the contrary, what is uncanny is that one knows in advance precisely what is bound to happen, even how and when it will happen ... and then, effectively, it really does happen. So here we seem to have attained a first step towards establishing our thesis: the uncanny is an experience that is not altered by a transcendent position of ‘better knowledge’. The Rat Man knows quite well that ‘wishes can’t kill’, that these absurd coincidences mean nothing at all, nonetheless, he is plagued by uncanny feelings and a mythical guilt strangely proportional to a disavowed belief in ‘the omnipotence of thoughts’.

Now the second part of Mannoni’s formula arises: could it be that the uncanny is not only impervious to ‘better knowledge’, but more strongly, that the latter should be seen as its enabling condition? This would mean that in the absence of better knowledge, whenever the distance between myself and the suspended belief disappears, the phenomenon of the uncanny would disappear as well. Again, Freud has noted this point in his subsequent effort, later on in the text, to shift the emphasis from his earlier material definition of the uncanny in terms of the content of the appearance, to a formal definition in terms of the subjective position of those afflicted by the uncanny towards that very same content. Why, for instance, would the prompt fulfilment of Polycrates’ wishes induce an uncanny effect, while the very same appearance of immediate wish-fulfilment in the story of *The Three Wishes* produces none at all? Or whence the uncanniness surrounding the living doll Olympia in Hoffman’s story, while again the same theme of coming-to-life of inanimate objects in most fairy tale stories remains without effect? If it is not the content of the appearance as such that seems to be responsible for the uncanny feeling when reading those stories, then, as Freud argues, we should consider the subjective position from which such stories are read (1919, pp. 245-246). What sets apart, for example, the aesthetic of Hoffman’s story from that of fairy tales is that in the former, at the outset, we are invited to adopt the



natural attitude of everyday life. Within this attitude, we know that such figures as the eye-robbing Sand-Man do not exist. In fairy tales, however

[...] the world of reality is left behind from the very start, and the animistic system of beliefs is frankly adopted. Wish-fulfilments, secret powers, omnipotence of thoughts, animation of inanimate objects, all the elements so common in fairy stories, can exert no uncanny influence here [...] (Ibid., p. 250).

Here, precisely, in the enchanted world of fairy tales where the existence of ghosts, resurrection of the dead, miracles and other gruesome plots is assumed, we have nothing to fear. And we have nothing to fear precisely because the difference between better knowledge and belief is temporarily suspended. On the other hand, only if we are invited to the enlightened position of better knowledge, only if we know that wishes are not causally responsible for certain outcomes in the world, do we spontaneously react to those outcomes as if we believed that things could have been different. Similarly, precisely because the Rat Man was “a highly educated and enlightened man of considerable acumen” and therefore “did not believe a word of all this rubbish” (1909, p. 229), did he react with a strange and compulsory form of credulity in the face of the most trivial coincidences. Hence, as Freud aptly notes, we must have “surmounted” certain beliefs in order to be susceptible to the effects of the uncanny (1919, p. 247).

#### **4. Conclusion**

To conclude our article, we will note three points that are particularly relevant for our discussion. *First*, as was the case in the examples discussed earlier, the logic of the uncanny should be understood in terms of a complex dialectic between better knowledge and belief. Within this dialectic, better knowledge does not form the disjunctive counterpart to belief, but should itself be considered as but one necessary element situated on the same “plane of immanence” (Deleuze 2005) in conjunction with belief (see also Pfaller 2014, pp. 41-43). Starting from this general principle, psychoanalytic theory therefore argues for at least two critical points: *first*, pace Dennett and phenomenology, we do not always know what we believe or how things seem to us; *second*, we misrecognize the true function of the alleged distance towards the belief that is expressed through ‘knowing better’ (see Zizek 1989, pp. 32-33; Pfaller 2005, p. 118).

*Second*, this idea also enables us to put forward a further, finer distinction between psychoanalytic theory and the aforementioned phenomenological principle of ‘appearance qua appearance’ (cf. supra). According to this principle, beliefs are explained by the properties of the *apparent object*. Because reality appears in such a way as if, for example, it corresponds to the immediate fulfilment of my wishes, I (wrongly) believe that it does. Psychoanalytic theory, however, significantly departs from this picture: if, for instance, the uncanny appears uncanny and I strangely react, despite better knowledge, as if I believed in ‘the omnipotence of thoughts’, then this suspended

belief is not primarily due to the properties of the presented object, but first of all to the *act of presentation*. More precisely: on a psychoanalytic account, the uncanny is uncanny not because I confuse appearance with reality, not because I succumb to the illusion and naively follow the appearance, but, on the contrary, precisely because I do *not* follow the appearance, because I am able to recognize the illusion *as* an illusion (or at least believe I can) and therefore picture myself as occupying a transcendent position of better knowledge. In short: exactly when – to put it in Husserlian terms – I do not confuse properties of *what is represented* with the *act of representing*. The illusion, therefore, is not situated at the level of the object, nor is it situated at the level of judgment towards that object, but in how this very act of judgment, and the reflexive self-consciousness of that judgment, contributes to what I ‘objectively believe’.

*Third* and finally, this notion of ‘objective belief’ – i.e. beliefs that are suspended by ‘better knowledge’ - further allows us to confront in a distinctive way what may be considered as the defining gesture of Dennett’s heterophenomenology: the rejection of “the bizarre category of the objectively subjective” (Dennett, 1991, p. 132). According to Dennett, this obscure philosophical notion of the “objectively subjective” should be disposed of because it is once again a clear expression of our obstinate belief in the Cartesian theater. It would correspond to that mythical moment of pure phenomenological ‘givenness’, where the ‘given’ is ‘given’ before being ‘taken’ in one way or another, in a theoretically unencumbered place where consciousness *really* happens. As Dennett notes, such a place might be a comforting image for some philosophers because it preserves the traditional distinction between reality and appearance at the heart of human subjectivity, “the seeming-a-certain-way over and above a believing-that-it-is-that-way” (Ibid., p. 133). Yet, the examples we discussed point out that this is perhaps not the only way in which “the objectively subjective” can be read. Indeed, instead of believing in ‘appearances’ and not knowing whether or not they correspond to reality, ‘objective beliefs’ display exactly the opposite structure: we know that they are not true, perhaps we even have a thoroughly adequate picture of reality, but the fact that we are nonetheless defined by them, that we act on them, remains undetected (Pfaller, 2014). Hence, from a psychoanalytic view, it is not so much the illusion of the ‘Cartesian theater’, but the Dennettian theme of ‘belief in belief’ considered as an exhaustive account of subjectivity that should be questioned.

## References

- Althusser, L. (1996). *Writings on Psychoanalysis*. New York: Columbia University Press.
- Bachelard, G. (1949). *Le Rationalisme appliqué*. Paris: PUF.
- Bernet, R. (2002). Unconsciousness consciousness in Husserl and Freud. *Phenomenology and the Cognitive Sciences* 1(3), 327-351.
- Bianco, G. (2016). *La signification du concret. Psychologie, philosophie et politique chez Georges Politzer*. Paris: Hermann (Hermann Philosophie).
- Burge, T. (2010). *Origins of Objectivity*. Oxford: Clarendon Press.
- Chalmers, D. (1995). *The Conscious Mind*. New York, Oxford: Oxford University Press.
- Cioffi, F. (1998). *Wittgenstein on Freud and Frazer*. Cambridge: Cambridge University Press.
- Deleuze, G. (2005). *Pure Immanence: Essays on Life*. New York: Zone Books.
- Dennett, D. (1991). *Consciousness Explained*. New York: Back Bay Books/Little, Brown and Company.
- Dennett, D. (2001). The Fantasy of First Person Science, online: <http://ase.tufts.edu/cogstud/papers/chalmersdeb3dft.htm> (accessed. 1/04/15)
- Dennett, D. (2003). Who's on First? Heterophenomenology Explained. *Journal of Consciousness Studies* 10(9), 19-30.
- Dennett, D. (2007). Heterophenomenology Reconsidered. *Phenomenology and the Cognitive Sciences* 6(1-2), 247-270.
- Dennett, D. (2013). *Intuition Pumps and other Tools for Thinking*. New York: W. W. Norton & Company.
- Dennett, D., & Kinsbourne, M. (1992). Time and the observer: The where and when of consciousness in the brain. *Behavioral and Brain Sciences* 15(2), 183-201.
- Descombes, V. (1999). *The Mind's Provisions: A Critique of Cognitivism*. Princeton and Oxford: Princeton University Press.
- Descombes, V. (2009). L'équivoque dy symbolique. *Revue du Mauss* 34(2), 438-466.
- Dreyfus, H. & Kelly, S.D. (2007). Heterophenomenology: Heavy-handed Sleight-of-hand. *Phenomenology and the Cognitive Sciences* 6(1-2):45-55.
- Dolar, M. (1991). "I Shall Be with You on Your Wedding-Night": Lacan and the Uncanny. *October* 58, 5-23.
- Dolar, M. (1998). "Cogito as the Subject of the Unconscious", in S. Zizek (Ed.), *SIC 2: Cogito and the unconscious*. Duke University Press, 11-40.
- Dupuy, J. M. (2000). *On the Origins of Cognitive Science: The Mechanization of the Mind*. Princeton and Oxford: Princeton University Press.
- Freud, S. (1909). Notes upon a case of obsessional neurosis. *Standard Edition* 10, 153-318.
- Freud, S. (1915). The unconscious. *Standard Edition* 14, 166-215.
- Freud, S. (1919). The "uncanny". *Standard Edition* 19, 219-256.
- Freud, S. (1923). The ego and the id. *Standard Edition* 19, 12-66.
- Freud, S. (1938). An outline of psycho-analysis. *Standard Edition* 23, 141-207
- Freud, S. (1938). Findings, ideas, problems. *Standard Edition* 23, 299-300.
- Gallagher, S. (2007). "Neurophilosophy and neurophenomenology", in L. Embree & T. Nenon (Eds.), *Phenomenology 2005: vol. 5, selected essays from North America*. Bucharest: Zeta, 293-316.
- Gardner, S. (2000). Psychoanalysis and the personal/sub-personal distinction. *Philosophical Explorations* 1, 96-119.
- Goodman, N. (1978). *Ways of Worldmaking*. Indianapolis: Hackett.
- Heidegger, M. (1962). *Being and Time*, Oxford: Blackwell Publishers Ltd.
- Heidegger, M. (1985). *History of the Concept of Time*. Indiana University Press.
- Husserl, E. (1970). *The Crisis of European sciences and transcendental phenomenology*. Evanston, IL: Northwestern University Press.

- Husserl, E. (1977). *Phenomenological Psychology: Lectures, Summer Semester, 1925*. The Hague: Nijhoff.
- Husserl, E. (1983). *Ideas pertaining to a pure phenomenology and to a phenomenological philosophy. First book: General introduction to a pure phenomenology*. The Hague: Martinus Nijhoff.
- Husserl, E. (1989). *Ideas pertaining to a pure phenomenology and to a phenomenological philosophy. Second book: Studies in the phenomenology of constitution*. Dordrecht: Kluwer.
- Husserl, E. (1995). *Cartesian meditations: An introduction to phenomenology*. Dordrecht: Kluwer Academic Publishers.
- Husserl, E. (1999). *The Idea of Phenomenology*. Dordrecht: Springer.
- Husserl, E. (2001). *Logical Investigations*. 2<sup>nd</sup> ed. 2 vols. London: Routledge.
- Johnston, A. (2012). "Turning the Sciences Inside Out: Revisiting Lacan's "Science and Truth," in *Concept and Form: Volume Two*, eds P. Hallward and K. Peden. London: Verso, 105-122.
- Kant, I. (1996). *Critique of Pure Reason*. Hackett.
- Kitcher, P. (1992). *Freud's Dream: A Complete Interdisciplinary Science of Mind*. Cambridge, Mass.: MIT Press.
- Koch, A. (1997). The Limit of My World. *Parabel* 1, 70-82.
- Kolers, P.A., & von Grünau, M. (1976). Shape and Color in Apparent Motion. *Vision Research* 16: 329-335.
- Lacan, J. (1982-1983). Merleau-Ponty: In Memoriam. *Review of Existential Psychology and Psychiatry* 18(1/2/3), 73-81.
- Lacan, J. (1993). *The Psychoses. The Seminar of Jacques Lacan: Book III 1955-1956*. London: Routledge.
- Lacan, J. (2006 [1936]). "Beyond the Reality Principle", in J. Lacan & J-A. Miller (Eds.), *Ecrits*. New York and London: W.W. Norton, 58-74.
- Lacan, J. (2006 [1953]). "The Function and Field of Speech and Language in Psychoanalysis", in J. Lacan & J-A. Miller (Eds.), *Ecrits*. New York and London: W.W. Norton, 197-268.
- Lacan, J. (2006 [1956]). "Seminar on "The Purlined Letter", in J. Lacan & J-A. Miller (Eds.), *Ecrits*. New York and London: W.W. Norton, 6-50.
- Laplanche, J. (1999). *Essays on Otherness*. London: Routledge.
- Livingstone-Smith, D. (1999). *Freud's Philosophy of the Unconscious*. Kluwer Academic Publishers.
- MacDonald, P.S. (2000). *Descartes and Husserl. The Philosophical Project of Radical Beginnings*. New York: SUNY Press.
- Mannoni, O. (2003). "I Know well, But All the Same ...", in M.A. Rothenberg, D.A. Forster & S. Zizek (Eds.), *Perversion and the Social Relation*. Durham, NC: Duke University Press, 68-92.
- Merleau-Ponty, M. (2002). *Phenomenology of Perception*. London: Routledge.
- Milner, J.C. (1995). *L'Oeuvre claire: Lacan, la science et la philosophie*. Paris: Le Seuil.
- Moore, M. (1988). "Mind, brain and unconscious," in P. Clark & C. Wright (Eds.), *Mind, Psychoanalysis and Science*. Oxford: Blackwell, 141-166.
- Noë, A. (2004). *Action in Perception*. Cambridge, London: The MIT Press.
- Noë, A. (2007). The Critique of Pure Phenomenology. *Phenomenology and the Cognitive Sciences* 6(1-2), 231-245.
- Overgaard, S. (2015). How to do things with brackets: the epoché explained. *Continental Philosophy Review* 2, 179-195.
- Peirce, C.S. (1992). *The Essential Peirce*, vols. 1 and 2. Bloomington, IN: Indiana University Press.
- Pfaller, R. (2005). "Where is Your Hamster? The Concept of Ideology in Zizek's Cultural Theory", in J. Glynos & M. Sharpe (Eds.), *Traversing the Fantasy: Critical Responses to Slavoj Zizek*. Ashgate Publishing Company, 198-216.
- Pfaller, R. (2006). "The Familiar Unknown, the Uncanny, the Comic," in S. Zizek (Ed.), *Lacan: The Silent Partners*. London, New York: Verso, 198-216.
- Pfaller, R. (2014). *On the Pleasure Principle in Culture: Illusions Without Owners*. London, New York: Verso.
- Pfaller, R. (2015). "The Althusserian Battlegrounds," in A. Hamza & F. Ruda (Eds.), *Slavoj Zizek and Dialectical Materialism*. London: Palgrave Macmillan, 23-42.
- Politzer, G. (1994). *Critique of the Foundations of Psychology*. Pittsburg: Duquesne University Press.

- Ricoeur, P. (1977). *Freud and Philosophy: An Essay on Interpretation*. New Haven: Yale University Press.
- Rosenberg, J. (2002). *Thinking about Knowing*. Oxford and New York: Oxford University Press.
- Sartre, J.P. (1958). *Being and Nothingness: An Essay on Phenomenological Ontology*. London: Methuen.
- Sartre, J.P. (1979). *Between Existentialism and Marxism* [Situations VIII and IX]. New York: Morrow Quill.
- Sartre, J.P. (2004). *The Transcendence of the Ego*. London: Routledge.
- Schafer, R. (1976). *A New Language for Psychoanalysis*. New Haven: Yale University Press.
- Schwitzgebel, E. (2007). No Unchallengeable Epistemic Authority, of Any Sort, Regarding Our Own Conscious Experience – Contra Dennett? *Phenomenology and the Cognitive Sciences* 6(1-2), 107-112.
- Sellars, W. (1981). Foundations for a Metaphysics of Pure Process. *The Monist* 64, 3-90.
- Sellars, W. (1991). "Philosophy and the Scientific Image of Man," in *Science, Perception and Reality*, Atascadero: Ridgeview Publishing Company, 1-40.
- Seron, D. (2015). "L'équivalence entre "mental" et "conscient" chez Brentano," in M. Gyemant & D. Popa (Eds.), *Approches phénoménologiques de l'inconscient*. Hildesheim : Georg Olms Verlag, 15-30.
- Staiti, A. (2015). On Husserl's Alleged Cartesianism and Conjunctivism: A Critical Reply to Claude Romano. *Husserl Studies* 31(2), 123-141.
- Thompson, E. (2007). *Mind in Life*. Cambridge, Massachusetts, London: Harvard University Press.
- Tomšič, S. (2012). Three Notes on Science and Psychoanalysis. *Filozofski vestnik* 33(2), 127-144.
- Tugendhat, E. (1986). *Self-Consciousness and Self-Determination*. Massachusetts: The MIT Press.
- Wittgenstein, L. (1958). *Philosophical Investigations*, 3<sup>rd</sup>ed. New York: Macmillan.
- Wittgenstein, L. (1978). *Lectures and Conversations on Aesthetics, Psychology and Religious Belief*. Oxford: Blackwell.
- Zizek, S. (1989). *The Sublime Object of Ideology*. London, New York: Verso.
- Zizek, S. (1998). The Interpassive Subject. Centre Georges Pompidou, Traverses, [lacan.com/zizek-pompidou.htm](http://lacan.com/zizek-pompidou.htm)
- Zizek, S. (2006). *The Parallax View*. Cambridge, London: MIT Press.
- Zizek, S. (2014). *Less Than Nothing*. London, New York: Verso.
- Zupancic, A. (2005). Reversals of Nothing: The Case of the Sneezing Corpse. *Filosofski vestnik* 26(2), 173-186.



# 6

## Expression and the unconscious

---

### 1. Introduction

A recurrent effort that has spurred much discussion in post-Cartesian<sup>53</sup> philosophy of mind has been to defend Descartes' basic insights on first-person authority while avoiding the tenacious image of what Dennett derisively called the "Cartesian Theater" (Dennett, 1991). That is, most philosophers would agree that I am often able to say what is going on 'in' my mind, what I want, believe, desire, intend and so on; whether I like to watch television, feel pain or find your joke funny, without defending the claim that I am able to do so by some sort of mysterious process of 'inner perception' or 'inner sense' involving ghostly spectators and shadowy images. A second and closely related claim is that this presumed first-person authority is in some important sense different from our third-person entitlements when speaking on behalf of others, or, similarly, when speaking in such a third-personal mode on behalf of ourselves. While it is not immediately clear whether this difference is to be cashed out in epistemic terms – i.e. the claim that self-knowledge always surpasses the knowledge I might have of the other's mental happenings – at least it seems to be a common assumption that there is a basic *first/third-person asymmetry* in the very way I ascribe such happenings to myself or to others which ought to be respected. Indeed, the success and validity of philosophical theories dealing with the perennial issues of subjectivity and first-person avowals is often measured in precisely those terms: they should (i) be able to preserve and explain the first/third-person distinction without (ii) succumbing to the Cartesian picture. Moreover, one of the main reasons why the latter is thought to be uninviting is precisely because talk about inner observation or *introspectio*, however intimate, clear or distinct such a process is held to be,

---

<sup>53</sup> To be sure, it is not so much the historical philosopher Descartes that is at stake here, but a certain general picture of the human being that has emerged – for better or worse - from the Cartesian account, regardless of the question whether this picture accurately reflects Descartes's philosophy. Central to this picture are the distinction between extended bodies and minds, issues of "privileged" access to my own mind as opposed to others', the certainty of the first-person vs the uncertainty of the third-person, a theory of ideas and related spectator-conception of consciousness, and so on. For some doubts about whether this 'picture' accurately reflects Descartes' own theory, see M. C. Eshleman (2007).

is a prime example of such an illegitimate extension of third-person observation language into the realm of the first-person. While it might enable us to explain first-person authority (only I am able to observe my inner thoughts, desires and so on), it does not sufficiently account for the difference from third-person ascriptions: that is, in both cases the same *process* of observation is involved, only the *object* of observation happens to be different (myself – the other). Furthermore, this emphasis on the asymmetry of first/third-person psychological ascriptions, rather than perhaps the Cartesian picture of ‘inner observation’ per se, as the critical touchstone in evaluating different philosophical proposals, is why for example behaviorist accounts are likewise considered to be unsatisfactory. For it might be argued, on stronger or weaker forms of behaviorist philosophy, that inner searchlights, or indeed, the inner as such, should be supplanted by external criteria. And that ascriptions of mental phenomena should be treated as statements referring to actual or potential behavior.<sup>54</sup> Again however, although in such accounts the Cartesian specter is clearly left behind, we are still left unable to explain in what way a self-report that is based on the observation of my own restless behavior and subsequent inferences of mental states, would be different from a similar report made with regard to others. So at least with respect to this inability to account for the first/third-person asymmetry, Cartesianism and behaviorism turn out to be unlikely bedfellows.<sup>55</sup>

Now there seems to be a broad family of philosophical positions that try to accommodate these requirements by turning to the concept of *expression*. Thus, for example, Wittgenstein’s notorious “plan for the treatment of psychological concepts” (1981, §472) invites us to consider first-personal avowals such as ‘I’m expecting an explosion’ or ‘I’m in pain’ not as derived from the observation of expectations and pains and their subsequent description, but as verbal expressions more on a par with how groans and smiles would express headaches and joys. In a similar way, Ryle claimed in *The Concept of Mind* that we should see first-personal utterances not as reports about certain frames of mind, but more as immediate disclosures or manifestations of those frames of mind: “If the lorry-driver asks urgently, ‘Which is the road to London?’ he discloses his anxiety to find out, but he does not make an autobiographical or psychological pronouncement about it” (Ryle, 2000, p. 164). Anscombe, finally, likewise urged us not to confuse verbal expressions of intentions with conjectural estimations of future actions (Anscombe, 1963, p. 6).<sup>56</sup> This emphasis on the expressive dimension of first-personal

---

<sup>54</sup> To quote two canonical formulations of this idea: “The sort of things that I can find out about myself are the same sort of things that I can find out about other people, and the methods of finding them out are much the same” (Ryle, 2000, p. 149); “The behaviorist maintains that our knowledge of ourselves is no different in kind from our knowledge of other people. We may see more, because our own body is easier to observe than that of other people, but we do not see anything radically unlike what we see of others” (Russel, 1949, p. 29).

<sup>55</sup> See Overgaard (2004) for an excellent elaboration of this point.

<sup>56</sup> This idea regarding the *expressive* form of first-person avowals as different from the *descriptive* form of third-personal attributions is of course not only retrievable amongst ordinary language philosophers. For a similar treatment of this issue within the phenomenological tradition, see Husserl’s section on ‘Expression and Meaning’ in his *Logical Investigations* (1970) (see Renaudie (2009) for a fine reading of this section). For a critical appraisal of the concept of ‘expression’ in relation to the philosophical and psychoanalytic problematic of ‘representation’, see Deleuze (1990) and Deleuze & Guattari (1977).



statements, as opposed to considering them as the linguistic outcomes of a sort of auto-descriptive voyeuristic activity, certainly comes a long way in meeting the philosophical requirements we spelled out above. Both first-person authority and the first/third person asymmetry can now be explained by the obvious fact that I can hardly express the intentions, wants or headaches of others (which I, nevertheless, might observe in *their* expressions of them). Yet, furthermore, such an account can also be of considerable importance in light of some of the philosophical discussions that have traditionally surrounded Freud's – and in a different, though nonetheless equally decisive way, Lacan's – understanding and justification of 'the unconscious'.<sup>57</sup> Before proceeding to our reading of the expressivist account of first-person authority, let us briefly note in what sense this might be so.

One way of delineating Freud's discovery that seems to be beyond reasonable suspicion is that, as Laplanche & Pontalis put it, "psychoanalytic theory emerged from a refusal to define the psychical field in terms of consciousness" (1988, p. 84). Very early on some philosophers thought this as paradoxical, but then it seemed it was above all Freud's recourse to the notion of 'unconscious representation' that was considered paradoxical.<sup>58</sup> 'Conscious representation', on the other hand, appeared quite natural. Yet beyond this superficial dispute as to which psychical quality we want to assign to these representations or *Vorstellungen*, the reference to 'representative ideas' or *cogitationes* was firmly established. Therefore, in contrast to, for example, Lacan's somewhat oblique way of defining the common denominator between Freud and Descartes in terms of a "subject of certainty" (Lacan, 1981, p. 35), there seems to be a much more straightforward reason why indeed "the Freudian field was possible only a certain time after the emergence of the Cartesian subject" (Ibid., p. 47). Obviously, the gesture of defining the 'psychic field' in terms of representations, whether conscious or unconscious, and in its wake, postulating the mechanism of inner perception as their ever-faithful companion, present themselves as fine candidates for explaining this historical co-emergence. For if consciousness is conceived in terms of inner sense or self-perception, then it isn't too difficult to imagine the unconscious creeping in as soon as doubts were being raised as to the scope of this mechanism of mental illumination. As indeed Freud was about to propose in his famous article on *The Unconscious*:

In psychoanalysis there is no choice for us but to assert that mental processes are in themselves unconscious, and to liken the perception of them by means of consciousness to the perception of the external world by means of the sense-organs (...) Just as Kant warned us not to overlook the fact that our perceptions are subjectively conditioned and must not be regarded as identical with what is perceived though unknowable, so psychoanalysis warns us not to equate perceptions by means of consciousness with the unconscious

---

<sup>57</sup> Most notably, David Finkelstein has presented some instructive clues as to how a philosophical reading of first-person authority informed by the later Wittgenstein, can be of particular relevance for psychoanalytic conceptions of unconscious mentality. See, in particular, Finkelstein (1999). We will return to some of Finkelstein's suggestions later in our article.

<sup>58</sup> Paradigmatically for this case are Brentano's 'four types of arguments' against the possibility of an 'unconscious consciousness' in his *Psychology from an Empirical Standpoint* (2009, p. 81 *et passim*).

mental processes which are their object. Like the physical, the psychical is not necessarily in reality what it appears to us to be. We shall be glad to learn, however, that the correction of internal perception will turn out not to offer such great difficulties as the correction of external perception – that internal objects are less unknowable than the external world (1915, p. 171).

Although, of course, Freud's specific deployment of Kant as having advanced some sort of psychologistic reading of our faculty of knowledge is contentious, it is nevertheless clear in what way the justification of unconscious mentality is but a necessary consequence of this representational point of departure. If, indeed, our consciousness of mental contents is not merely the consequence of - as some contemporary phenomenologists have it - their 'intransitive self-givenness'<sup>59</sup>, but rather, as Kant put it, a secondary *achievement*<sup>60</sup>, viz. the outcome of a synthetic act which both attends to and abstracts from immediate sense impressions, then the unconscious of such a transcendental consciousness lets itself be understood in basically two ways: not only as (i.) those latent mental processes that are unaccompanied by that act, but also (ii.) the anterior act itself as required for consciousness, the latter due to the fact that, as Kant put it, "I cannot know as an object that which I must presuppose in order to know any object" (1999, A402).<sup>61</sup> It is then no more surprising that this structural *rencontre manquée* between representational consciousness and its fugitive condition, a condition which *as condition* always both enables and hides itself in its product, was bound to serve as the conceptual figure through which psychoanalysis would take up its proper place as the latest and most faithful heir of the Cartesian cogito. This explains why, for Lacan, "the subject, the Cartesian subject" would function as "the presupposition of the unconscious" (2006, p. 712). This also explains why, for Žižek, it is only in psychoanalysis that "the forgotten obverse, the excessive, unacknowledged kernel of the cogito" is properly exposed (1999, p.2).

Yet, leaving these canonical philosophemes behind for now, what becomes of the unconscious when the reference to reflexive consciousness, inner sense and other related notions is rejected? Obviously, insofar as *these* are rejected in serving as a convincing account for first-person authority, so

---

<sup>59</sup> For a detailed exposition of this notion and other related ones (e.g. 'pre-reflective self-consciousness', 'first-personal-givenness', 'for-me-ness', and so on), see Zahavi (2014).

<sup>60</sup> This idea of the necessary intervention of an active judgment in order to explain consciousness is even more explicit in Schelling's *System of Transcendental Idealism*, in particular in the section on "How the Self intuits itself as Sensing": "The self *senses* when it finds in itself something opposed to it, namely, since the self is mere activity, a real negation of activity, or state of being affected. But to be that which senses, for itself, the (ideal) self must posit *in itself* that passivity which till now has been present only in the real; and this can undoubtedly occur only through *activity*" (1978, p. 61).

<sup>61</sup> The clearest formulation of the idea that the psychoanalytic concept of the unconscious is but the 'immanent truth' of German idealism's concept of consciousness, can be found in Henry's *The Genealogy of Psychoanalysis*: "For the last word of the philosophy of consciousness, its limit and paradox, the zenith at which it turns against itself and self-destructs, is truly *the unconscious of pure consciousness itself*, the unconscious of "transcendental consciousness". The historical moment of this turning occurs when German idealism, unable to found the principle on which it rests and eaten away from inside by this major inability, falls into nature philosophy, which is that idealism's truth, affirming nothing less than the unconscious of pure consciousness itself, namely, the unconscious of pro-duction" (Henry, 1993, p. 289). See also Redding (1999), Gardner (2012) and Feyaerts & Vanheule (2015).

should their inverted offspring when it comes to accounting for unconscious phenomena. That is, if first-person authority is defined in terms of the capacity to *say* something about oneself without having to rely on conjectural inferences or observational evidence, then the limits to this authority should not be thought in terms of the possible *failure* of such third-personal procedures. Barring the latter, when is an intention, fear or belief then rightly said to be unconscious? What is it by virtue of which a state of mind should be characterized as unconscious when, what we have termed, ‘the unconscious of transcendental consciousness’ is no longer suitable as a convincing explanation? In order to delineate a possible answer to these questions, let us now turn to the expressivist proposal.

## 2. Explaining first-person authority: from detection to expression

In our introduction, we tentatively characterized the phenomenon of first-person authority in terms of a certain subjective privilege that seems to be attached to a range of mental states, acts and attitudes. Now it is time to clarify what this idea precisely entails.

Firstly, the basic intuition which is captured by such notions as ‘authority’ and ‘privilege’ is that, all things being equal, I am usually the best person to ask if you want to know what I think, feel, intend, imagine and so on. Further classical examples of such a privileged set of mental states are those that are usually invoked to put an end to philosophical discussions on skepticism. Thus, according to Descartes, I can be mistaken whether I’m really taking a walk or writing a text, but not whether I believe this is so, or whether it seems to me that way, or judge it to be so. These typical *cogitatio*<sup>62</sup> are all members of a larger class which Wittgenstein designated as ‘psychological verbs’. Amongst these psychological verbs, we find those which point towards experiences (‘I see’, ‘I hear’, ‘I have a headache’), cognitive activities (‘I think’, ‘I judge’), but also intentions (‘I am going to ...’, ‘I propose to ...’) or acts of the will (‘I’ve decided to’, ‘I want to’).

Secondly, our ‘subjective privilege’ with regard to these psychological verbs basically consists in being exempt from the demands of verification or evidence that are usually attached to our ordinary judgments, which indeed can be contested, contradicted, denied or, at the very least, questioned as to their justification. This is the feature Wittgenstein had in mind with the formula ‘asymmetry of the first and third persons in the present indicative’. That is, if, for example, I say I have the beginnings of juvenile dementia, or a broken arm, then the reasons I would cite in order to justify these claims would be the same kind of reasons as were I to say someone else had these medical conditions. Roughly, reasons pertaining to observation, testimony and inference. Hence, with respect to the question of justification,

---

<sup>62</sup> As noted by several commentators, Descartes uses the words ‘*cogitatio*’ and ‘*cogitare*’ in both an extended as well as a more restricted sense compared to our ordinary understanding of those notions in the sense of ‘thinking’ or ‘intellectual operations’. Extended because it also comprises, for example, the acts of imagination, sensing and willing. Restricted because Descartes retains only those acts which are immediately and absolutely ‘given’ to consciousness. See Descombes (2004, pp. 176-189) and Anscombe (1981b).

in these cases there is no asymmetry between, say, the first-personal statement ‘I lost my wallet’ and its third-personal inversion ‘he lost his wallet’: both are justified by, for example, checking one’s pocket. However, if I say I want to eat French fries or am contemplating Freud’s *Future of an Illusion* or desiring to go on a holiday, that similarity to the third person vanishes. Clearly, there seems to be something strained, or perhaps even uncanny, about someone asking me to justify these claims (à la ‘So you believe you’re contemplating Freud’s writings. Ok, but how do you know?’). And if I were nonetheless tempted to respond to such queries in order to ward off any remaining misunderstandings, citing behavioral evidence or inferences to the best explanation to the effect that I am really contemplating would seem to make matters worse. In fact, as Wittgenstein points out, an important feature of the apparently incorrigible character of self-ascriptions deploying psychological verbs in the present indicative is that one cannot ask the subject *how* he knows. By contrast, the third-personal claim that someone else is contemplating Freud’s writings is devoid of any such privileges and in principle susceptible to further demands for justification.

## 2.1 Detectivism

So how are we to explain this remarkable capacity to talk in such a seemingly effortless and authoritative way about our concurrent thoughts, hopes and fears? Following Finkelstein’s suggestion<sup>63</sup>, let us call *detectivist* any philosophical account that tries to explain first-person authority by invoking a special mode of *epistemic access* that allows the subject of these mental states to *know* about them. That is, detectivism takes the problem of first-person authority as actually being but one specific application of the more general epistemological question of knowledge – in this case ‘self-knowledge’ -, and the problem is therefore, as befits such epistemological queries, of determining the origin and conditions of possibility of that knowledge. The implicit reasoning behind such a claim seems to be the following: in the same way as I would be unable to say of this particular object that it is a table when I do not know what it is for something to be a table, I would not be able to say, for example, that I want to watch *Better Call Saul* if I did not know what is for a mental state of mine to be one of wanting-to-watch that series.<sup>64</sup> The detectivist account of first-person authority therefore assumes a phenomenological position<sup>65</sup> with

---

<sup>63</sup> See Finkelstein (2003). Finkelstein refers to Wright’s 2001 paper as the principal source of this term.

<sup>64</sup> Here I rely on Descombes’ admirable analyses of what he calls “l’égologie cognitive” (2004, pp. 146-156).

<sup>65</sup> In his recent book *Self and Other* (2014), Zahavi argues, contrary to our suggestion, that a phenomenologically inspired account of consciousness and first-person authority does not amount to detectivism. The main argument being that from a phenomenological perspective, self-consciousness “is not brought about by some kind of reflection or introspection or higher-order monitoring”, but is in fact always already “an intrinsic feature of the primary experience” (p. 35). However, I am skeptical about this argument for the following reasons: first, ‘detectivism’ is not primarily about what constitutes self-consciousness, that is, it doesn’t single out a position with regard to the discussion between ‘one-level’ or ‘two-level’ accounts of self-consciousness. Rather, it is a philosophical account of what explains first-person authority, i.e. our seemingly effortless ability to say what we think, feel or intend to do. To put it in Sartrean terms, it is an account of *connaissance de soi*, rather than *conscience de soi*. Secondly, detectivism amounts to the claim that first-person authority is in some, very general sense, a

regard to the conditions of self-ascription of first-person avowals: in order for a first-personal psychological avowal to have any reference, it is necessary to invoke a phenomenon of ‘experiential evidence’ for, or ‘givenness’ of, my mental states, in this case, the first-personal givenness of wanting-to-watch that series.

Now what it is that provides the subject with this evidence for her own mental states is something about which detectivists have held (and still hold) a variety of views. In order to offer a somewhat accessible overview of the widely divergent epistemic mechanisms on offer in the philosophical literature, let us begin by describing the *explanandum* of the detectivist account in the following way:

If I x, then I know that I x (where ‘x’ stands for the different psychological verbs).

That is, whenever I feel, imagine, will or think, I always know that I do so. How to explain this remarkable lucidity?

A first traditional approach consists in postulating *two acts* in order to explain a subject’s cognitive access to its mental states. The first act being quite simply whatever mental activity happens to occur in my consciousness, the second ‘reflexive’ act being whatever epistemic activity that allows the subject to come to know about the first act by being ‘directed’ upon or towards it. Here we encounter some illustrious notions like ‘introspection’, ‘inward observation’, ‘acquaintance’, ‘inner sense’ or ‘higher order thoughts and representations’, but the precise nature (quasi-perceptual or representational) of these reflexive acts should not detain us. The essential thought is that such higher-order or two-act theories guarantee immediate knowledge of my first-order mental acts through the recourse to a distinct – logical, ontological or temporal – further act, one that, moreover, can only be directed to my own mental states, not those of others. That is, I cannot ‘introspect’ and hence know immediately whether the other is in pain or intending to go for a walk, I only have this privileged spectatorial position with regard to myself.

The objections and difficulties that such higher-order accounts have encountered have been repeatedly discussed elsewhere, so I will not recount them here *in extenso*.<sup>66</sup> Suffice for our present purposes to refer to two of those that have figured most prominent amongst them: on the one hand, ever since Hobbes, Leibniz or Brentano, it has been noted that two-act theories always lead to an infinite regress which can only be blocked by positing anterior unconscious mental acts.<sup>67</sup> On the other hand,

---

matter of ‘finding out’ what we think, hope or feel, regardless of the sort of *mechanism* that is invoked to describe this epistemic detection. In other words, if detectivism is to be rejected, then it is not only ‘introspection’ which serves as an exemplary culprit, but *any* account that distinguishes two steps – one of knowing by detection, the other describing what one knows – in order to explain first-person authority. And this, I believe, is precisely the sort of account Zahavi advances when he writes about the transition between “pre-reflective self-consciousness” and “self-knowledge” in terms of “reflection” (p. 36).

<sup>66</sup> See Smith (1986); Thomasson (2000; 2006); Kriegel (2003); Shoemaker (1996) and Bar-On (2004).

<sup>67</sup> In short: if I know that I x whenever I x, it is because I x that I x whenever I x, and so on. As Leibniz put it: “(I)t is impossible that we should always reflect explicitly on all our thoughts; for if we did, the mind would

the fact that self-knowledge is granted such unusual epistemological features as immediacy, infallibility, immunity to error through misidentification, and so on, makes it hard to understand how any quasi-perceptual or representational mechanism would ever be able to explain these. For it seems to be part of our language games that talk about perception and representation at least implies the possibility of misperception and misrepresentation. Yet insofar as the latter are already excluded by the very definition of the detectivist explanandum, so should their extraordinary counterparts if we want to keep using our concepts of perception and representation in any meaningful way.<sup>68</sup>

A second approach, most commonly associated with phenomenological writers such as Sartre or Husserl, tries to explain the special knowledge I have of my mental acts, while evading the regress objection, by simply stating that such psychological verbs like thinking, perceiving or doubting are ‘acts of consciousness’. In keeping with Brentano’s definition of mental phenomena, such acts of consciousness are said to be *intentional*, i.e. they are acts through which I am immediately conscious of *something*; differently put, they are what allows me to relate to a world that is transcendent to my consciousness. Perception, for example, is that mode of consciousness through which a perceptual object appears; imagination, that mode of consciousness through which something is given as imagined, thinking, that by which something is thought; ... Formally:

If I x, then I am conscious of something in the mode of x

Therefore, on this account, in contrast to two-act theories, one does not say that my intention to go on a holiday is something I perceive or represent by way of a further act, but rather, that my intention is a ‘consciousness’ of a future holiday in the way of intending it. Similarly, one does not say that my desire for a bloody mary is something I apperceive as a disinterested spectator, but that desire is a mode of consciousness that puts me in a relation to a bloody mary in order for the latter to appear desirable. One has to distinguish, as Husserl (1976, pp. 303-4) put it, the cogitatum from its cogito, the intentional object as it is intended (*im Wie seiner Bestimmtheiten*) from the intentional act through which it is intended (*im Wie seiner Gegebenheitsweisen*). In terms of the kind of first-personal knowledge this yields:

---

reflect on each reflection, *ad infinitum*, without ever being able to move on to a new thought” (1981, II, I, §19). As we have noted, this is one *locus classicus* to argue for the unconscious. This idea has been most rigorously conceptualized, with some modifications, by Lacan in his adoption of Bouasse’s ‘optic mirror device’. See his “Remarks on Daniel Lagache’s Presentation” in *Ecrits* (2006, pp. 543-574). For a clear exposition of this idea in relation to Kant’s thesis on the impossibility of an ‘intellectual intuition’ of the ‘I think’, see Zupancic (1996).

<sup>68</sup> Importantly, this is also why the defense of unconscious mentality is actually *more* rational and consistent within the framework of Cartesian higher-order theories than ideas about transparency or infallibility. From the moment one accepts two-act theories in order to explain first-person authority in a non-metaphysical way, that is, without implicitly assigning a different meaning to ordinary concepts like perception, belief or representation, one has to be prepared to endorse the idea that mental phenomena are indeed, as Freud put it, not necessarily what they appear to be (cf. *supra*). This would mean, for example, that I can ‘misperceive’ my desire as a rational belief, can ‘misrepresent’ my groundless anxiety as a reasonable intention, or that such mental phenomena can even go totally ‘unnoticed’.

If I x, then I know *what* I x and I know *that* I relate to this intentional object in the mode of x

It is, in other words, through the very transitive or 'thetic' consciousness of, say, a burning house as perceived or imagined, that I am immediately alerted to the intransitive or 'non-thetic' act of perception or imagination. Importantly, it is only because the act through which I am consciously related to objects is itself also 'experienced', *viz.* that there is something 'it-is-like' to perceive, believe, think and intend, that I am subsequently able to report on that act and its intentional object whenever I'm asked. In Sartre's cigarette-counting example:

(...) at the moment when these cigarettes are revealed to me as a dozen, I have a non-thetic consciousness of my adding activity. If anyone questioned me, indeed if anyone should ask "What are you doing there?" I should reply at once, "I am counting" (2004a, liii)

Hence, on a phenomenological account, the fact that Sartre is able to speak with first-person authority about what he is doing is grounded in the experiential evidence attached to the counting activity. Furthermore, that evidence should not be considered as the secondary result of introspection, remembrance or reflection, but as something that is already part and parcel of the pre-reflexive consciousness (i.e. 'non-thetic consciousness') of Sartre's counting. To put it in the typical phenomenological neologisms for denoting this kind of intransitivity, such pre-reflexive experiences are said to be 'intrinsically self-revealing', 'self-disclosing', 'pre-reflectively self-given', 'self-appearing' or 'self-manifesting' (for these ideas, see Henry (1973)). Coining another neologism appropriate to our discussion, we can say that such acts of consciousness are 'self-detectifying': that is, it is in their very nature to procure their own evidence.

Yet, apart from the strange observation that in order to make sense of such one-level accounts, we have to invoke what appears to be *sui generis* epistemological procedures and associated philosophical vocabulary, it is clear that the proposal's *prima facie* plausibility derives from the decision to extend the class of experience verbs so as to include acts like thinking, judging or remembering. Apparently, all these different psychological verbs now become 'acts of consciousness', and therefore, insofar as my ability to report on them is explained by their 'self-revealing' character, must have a distinctive phenomenology about them. Yet, believing that it is about to rain or remembering to take out the garbage is hardly to have experienced something in the mode of thinking or remembering, and it is highly questionable whether there even is something like an experiential quality of the act of intending to read the newspaper. But even in the case of proper experience verbs like 'feeling ill' or 'sensing a pain in my shoulder', it is difficult to conceive in what way my self-ascriptions of illness or pain could amount to any sort of epistemic achievement. For does it make sense to say that 'I know I'm not feeling

quite well today' because I have a pre-reflexive consciousness of my current misery? Or, indeed, more generally, to say that 'I know' these troublesome sensations at all?<sup>69</sup> What is that supposed to mean beyond the deplorable fact that, quite simply, I don't feel well?<sup>70</sup> Obviously, epistemic operators like 'I know that p' or 'I know whether p' only make sense if they can be used in contrastive conjunction with their negations or modifications like 'I don't know that p', 'I doubt whether p' or 'I wonder if p'. Whereas in the third-person case, I can indeed be ignorant or remain in doubt about someone else's feeling of nausea, similar expressions in the first-person present tense like 'I wonder if I have a headache', 'I have a pain in my back, but I'm not sure', ..., are strictly speaking meaningless. Yet they are not meaningless because, to paraphrase Sartre, we are dealing with "a cogito that retains its rights even with psychopaths" (2004a, p. 148), or with the ineliminable occurrence of a consciousness unable to forget itself, but rather because there are no epistemic rights to retain, nor any business of forgetting or remembering involved in the first person present employment of such psychological verbs.

## 2.2 Wittgenstein's expressivism

Now the latter conclusion might, for some, already count as a *reductio ad absurdum* of the critique of detectivism just presented. To deny psychological self-ascriptions any epistemic purchase seems to amount to denying the phenomenon of first-person authority itself. That is, if it can't be said that I know whether I have a headache or believe that it is raining, why then should I be credited with any kind of authority on the matter? Additionally, anticipating our discussion of unconscious mental states, if talk about knowledge is denied for those beliefs and desires we are seemingly in a position to avow without any hesitation, then it seems to be a direct consequence of such a non-epistemic approach that talk about beliefs and desires that do not share these first-personal asymmetries is equally ruled out in advance.

So the problem we have been developing can therefore be reformulated in the following way: how to save the Cartesian intuition that all of us enjoy a privileged and reliable authority concerning our present tense employment of psychological verbs if we reject the idea that this ability is epistemically grounded in whatever sense? How to explain the ostensible authority that seems to adhere to claims maintaining that I want to play chess, intend to go out or meant to say whatever it was before you interrupted me, without invoking the idea of inner observation or experiential givenness of wants, meanings and intentions?

As pointed out in our introduction, one way of addressing this problem which arises from Wittgenstein's later work, begins with the suggestion that our self-ascriptions owe their status, which is so apparently authoritative, to the simple fact that they are not, after all, *reports* of current psychological states, but rather count as *expressions* of the states they only appear to describe on a superficial

---

<sup>69</sup> See Hacker (2005) for an extensive discussion. See also Bouveresse (1976, pp. 421-23).

<sup>70</sup> As Wittgenstein put it: "It can't be said of me at all (except perhaps as a joke) that I *know* I'm in pain. What is it supposed to mean – except perhaps that I *am* in pain?" (2009, §246).



consideration of their surface grammar. In the case of sensation-language, Wittgenstein puts this point as follows:

It is not, of course, that I identify my sensation by means of criteria; it is, rather, that I use the same expression. But it is not as if the language-game *ends* with this; it begins with it.

But doesn't it begin with the sensation – which I describe? – Perhaps this word “describe” tricks us here. I say “I describe my state of mind” and “I describe my room”. One needs to call to mind the differences between the language-games (2009, §290).

What are ‘the differences between the language-games’ of describing one’s room and describing one’s state of mind Wittgenstein is hinting at in this quote? An important point seems related to - if we can put it that way - the ‘temporal logic’ of the descriptive ascription. Wittgenstein appears to be intent on denying that in the case of, for example, saying that I am in pain, I should begin with the examination of my pain-sensation (through, say, identifying my headache *as* a headache by means of headache-criteria) before proceeding to the self-ascription; whereas, in the case of describing my room, it is of course indispensable that I should have observed things and looked it over before engaging in my description. In the latter case, the description comes at the *end* of the language game – after I have identified whether the room is spacious or contains a *chaise longue*. Moreover, this identification is what in fact *justifies* my descriptive assertions. Yet, if Wittgenstein denies that this is what we ordinarily do when giving one’s thought on some philosophical matter or complaining about a pain in one’s back, then what is it that justifies these first-personal ascriptions? And what does it mean to say that in this kind of language game such ascriptions come *first*, rather than at the end of some justificatory activity? To continue the analogy with the room-description, one might think, for example, that to begin with the ascription would actually amount to describing one’s room without looking at it first, that is to say, to simply engage in some highly speculative endeavor. And since, in fact, Wittgenstein abrogates epistemic justifiers in the case of psychological self-ascriptions, this seems to leave us with the following option: to construe such talk as being merely the result of – to use Dennett’s expression (1991, p. 67) – some “impromptu theorizing”. However, that this is *not* the conclusion Wittgenstein wants to draw from his critique of inner observation or any similar detection mechanism, becomes clear in the following passage:

“When I say ‘I am in pain’ I am at any rate justified *before myself*.” – What does that mean? Does it mean:

“If someone else could know what I am calling ‘pain,’ he would admit that I was using the word correctly”?

To use a word without a justification does not mean to use it wrongfully (Wittgenstein, 2009, §289).

So, according to Wittgenstein, it is not because I am not ‘justified before myself’ in saying that I’m in pain, that I am therefore *unjustified* when resorting to this pain-talk in the sense of making a mistake or simply talking nonsense. Such a confusion will remain as long as we do not reject “the grammar which

tends to force itself on us here” (Ibid., §304), which is precisely the descriptive grammar that is premised on the model of “object and name” (Ibid., §293) as in the case of the description of my room. To rid oneself of the idea that, in order to avoid skeptical worries about our ‘folk-psychological’ practices, authoritative psychological self-ascriptions of beliefs and sensations *must* be grounded in evidence roughly in the way that the observation of my room supports my subsequent descriptive assertions, we have to question the assumption that first-person avowals function as *reports* about psychological facts the speaker has previously learned or ascertained. This is the guiding thought behind Wittgenstein’s following remarks:

The paradox disappears only if we make a radical break with the idea that language always functions in one way, always serves the same purpose: to convey thoughts – which may be about houses, pains, good and evil, or whatever (Ibid., §304).

When someone says “I hope he’ll come”, is this a *report* [*Bericht*] about his state of mind, or a *manifestation* [*Äußerung*] of his hope? – I may, for example, say it to myself. And surely I am not giving myself a report (Ibid., §585).

Here we retrieve the earlier suggestion that psychological utterances are primarily *manifestations* or *expressions* of the very states they self-ascribe, rather than secondary thoughts or beliefs about these states uttered to inform the listener about whatever psychological fact it was I had previously ascertained. My utterance of ‘I believe (hope, desire, intend, ...) that *p*’ hence expresses, rather than reports, my belief (hope, desire, intention, ...) that *p*. Therefore, utterances like ‘I am in pain’ or ‘that really hurts!’ are not pain-reports on a par with weather-reports, but, like my moans or cries, manifestations of my pain; similarly, the utterance ‘I want to go out’ does not inform someone about some planning experience, but voices, depending on the context in which it is said, my desire to go out or my agreement to your proposal; finally, ardent exclamations like ‘I love you’ or ‘I really hate your guts’ do not communicate subjective facts to which you may or may not agree or proceed to inquire some further information, but should be more properly considered as verbal equivalents of bringing flowers and throwing plates in your direction.

Now, to conclude this section, what is the philosophical import of Wittgenstein’s expressivism with regard to our discussion of first-person authority and how precisely does it differ from the detectivist solutions earlier discussed?

First of all, despite some other difficulties it has encountered<sup>71</sup>, expressivism has one striking advantage when compared to other accounts of first-person authority: the reason why, indeed, I am

---

<sup>71</sup> Most notably on the issue of whether or not expressive self-ascriptions can be truth-evaluable. See Rosenthal (1993) and Wright (1998) for typical formulations of this worry. See Alston (1967) and Finkelstein (2004, pp. 94-97) for responses.

usually the best person to ask if you want to know what I believe or desire, is that, on an expressivist account, my self-ascriptions ordinarily count as manifestations of those beliefs and desires, roughly in the same way as driving a car shows my ability to do so, or as blushing reveals my embarrassment. This means that, contrary to detectivist accounts, there is then *no* epistemological question to be answered about how I am able to avow my beliefs, fears and sensations, just as my ‘ability’ to laugh with your joke requires no preliminary introspective investigation nor realizes any epistemic accomplishment. The ‘authority’ or ‘privilege’ that accompanies my first-personal avowals is thus decisively *unlike* the authority of a privileged observer: I require no behavioral or experiential evidence to avow that I am afraid for exactly the same reason why the trembling of my hands or grimace on my face can forego personal efforts at justification. And since my mental state self-ascriptions are typically not reports or descriptions of these mental states, that is, since they do not involve subjective *judgments* about the presence of these mental states, solicitations for further proof or evidence should not be dismissed on the grounds of being overly scrupulous superfluities, but merely as betraying a fundamental misunderstanding of the expressive character of my linguistic performance.

Secondly, starting from this expressivist appraisal of psychological avowals, we can also begin to see in what way expressivism allows for an elegant solution of the first-third person asymmetry in mental state ascriptions that avoids some of the pitfalls detectivist accounts have encountered. That is, expressivism should be understood as claim about what is special to the use of psychological self-ascriptions in contrast with a wide array of other uses. But the contrast to be effected is not between a first-personal intimate, direct epistemic relation, by comparison to which third-personal inferences from behavior, for example, are indirect and less secure; rather, the asymmetry rests entirely on the difference between simple present tense expressive uses and all other employments of psychological terms. As Wittgenstein remarks in the *Blue Book*:

The difference between the propositions “I have pain” and “he has pain” is not that of “L.W. has pain” and “Smith has pain”. Rather, it corresponds to the difference between moaning and saying that someone moans (1969, p. 68).

So what is distinctive about an avowal like ‘I have pain’ in contrast to its third-personal inversion ‘he has pain’ is that in the former case, the avowal is issued by the very person who is said to be in pain, at the very time she is in pain, in the course of expressing her pain. Obviously, this is something I am in a unique position to do: only *I* can express or give voice to my own present states of mind, and it is only states of *my* mind that I can express. By contrast, my ascriptions of mental states to others, e.g. saying that someone else is in pain, will typically count as descriptions or reports of these states of mind as expressed in their (linguistic) behavior. In that case, my assertions can serve to express my beliefs about whether or not ‘he is in pain’, but thereby I do not, of course, express his pain, and furthermore, I ought to be able to provide some reasons for my belief (for example, because I saw he was in pain, or because

he told me so). Furthermore, this also explains why our first-person privilege extends only to *some* aspects of our mental lives and, thus, why the asymmetry between first –and third-person ascriptions is not reducible to the distinction between my psychological states and those of the other. Since first-person authority is restricted to those avowals which express my mental states in self-ascriptions of them, past-tense ascriptions ('I felt really sad while watching that movie'), future-tense ascriptions ('I will feel better after going for a walk') or self-attributions of psychological dispositions ('I'm a hopeless neurotic') will fall on the third-personal side of the asymmetry because my assertions about these matters will not express those states themselves.<sup>72</sup>

Finally, if, as we have argued, this asymmetry is not the consequence of any supposed epistemic asymmetry - in terms of either the superiority of self-knowledge in contrast to the uncertainty plaguing our knowledge of 'other minds', or the sort of mechanism invoked to explain this ostensible epistemic difference -, but rather the effect of a difference in terms of expressive versus descriptive usages of psychological terms, then some of the skeptical worries that have continued to resurface with regard to the phenomenon of first-person privilege permit for a different response. To take but one famous example drawn from the philosophy of cognitive neuroscience: from the deficiency of the Cartesian theater and its associated spectatorial accomplice in serving as a convincing account of first-person authority, Dennett famously argued for a view on first-personal avowals that regards them as "theorist's fictions" (1991, p. 98), that is, as provisional beliefs about mental items whose putative reality will be corroborated (or not) depending on future developments in neuroscientific research. As Dennett puts it:

People undoubtedly do believe they have mental images, pain, perceptual experience and all the rest, and these facts – the facts about what people believe, and report when they express their beliefs – are phenomena any scientific theory of mind must account for (Ibid.).

Here we see clearly how staunch materialists like Dennett, despite rejecting a metaphysically suspect account of what makes the phenomenon of first-person authority distinctive, still retain the essential idea that sustains the detectivist picture of psychological avowals where the latter, as we have seen, always serve to express *beliefs* about mental life, rather than mental life itself.<sup>73</sup> Yet, in thus reducing psychological avowals to the expression of beliefs about mental states, such materialist accounts cannot retain what *is* indeed special about first-personal self-ascriptions in comparison with their third-personal counterparts. And it is here that the explanatory fecundity of Wittgenstein's proposal most clearly shows

---

<sup>72</sup> See Bar-On (2004, p. 200 *et passim*) for some further distinctions with regard to the ascriptive asymmetry.

<sup>73</sup> And since first-person ascriptions are treated alike by detectivists, i.e. as (second-order) reports or descriptions of (first-order) mental states indicated by their psychological terms, the only point of contention between, say, materialists and dualists, will be on the metaphysical status of these reported psychological entities (e.g. as (ir)reducible to brain-states) and the way such reports should be adjudicated (through, e.g., phenomenological or hetero-phenomenological methods). For a representative sample of such discussions, see the collection of essays in *Phenomenology and the Cognitive Sciences* (2007, issue 1-2).

itself: in the very place where materialist detectivists locate a class of psychological entities in order to reduce or eliminate depending on future evidence, Wittgenstein points towards a distinctive employment of psychological vocabulary in relation to which the usual questions of justification and evidence quite simply do not arise. Yet, in making this distinction, it is clear that Wittgenstein, in contrast to philosophers like Dennett, does not mean to draw our attention to any first-person epistemic shortcoming, nor means to relegate psychological avowals to the domain of fiction; the claim is rather that our concept of knowledge (or, for that matter, fiction) cannot handle the relation between psychological avowals and mental life to begin with, and consequently, that the deficiency of self-knowledge should not be invoked to raise skeptical arguments against our ordinary practice of speaking our minds.

### 3. Expression and the ‘as-if’ of the unconscious

As suggested in our introduction, apart from its philosophical potential to offer a non-Cartesian solution for the asymmetry between first- and third-person psychological ascriptions, Wittgenstein’s expressivist reading of the kind of authority involved in first-person avowals also permits for a different take on what is or should be one of the major preoccupations of psychoanalytic theory: i.e. to offer a perspicuous account of ‘unconscious subjectivity’. However, it would of course be a slight overstatement to claim that the discussion on first-person authority just presented has figured prominently in the philosophical literature on psychoanalysis (or, in any case, in this particular form or by way of these philosophical concepts).<sup>74</sup> Nor do the adduced quotation marks mean to emphasize a univocal concept which can be invoked without further ado throughout our following discussion. On the contrary, my goal in the present section is precisely to offer some ideas which can serve towards such clarification by taking our cue from the critique of detectivism and Wittgenstein’s expressivist alternative.

#### 3.1 Unconscious: adverbial versus substantial

However, in setting up the problem in this way, that is, by tacitly assuming that a philosophical reading of what it actually means to talk about oneself in an authoritative or non-authoritative way could have significant implications for psychoanalytic theory, it might be objected we are already betraying one of the fundamental tenets of the Freudian approach. And in a certain way this is correct: what is indeed already excluded from further consideration by taking this angle of approach is the reference to ‘the unconscious’ which results from the grammatical transformation of adverbial qualifications such as ‘he unconsciously believed *p*’ or ‘he had the unconscious desire to *p*’, into a substantive employment

---

<sup>74</sup> But see MacIntyre (2004) and Bouveresse (1995) for two early efforts to connect some of Wittgenstein’s reflections on the grammar of psychological concepts with the psychoanalytic problematic of the unconscious. For more recent accounts, see Finkelstein (1999; 2004), and Doyle (2014) for a critical response.

of the term in phrases like ‘his unconscious believed *p*’ or ‘this might have something to do with my unconscious’. That is, there might be a perfectly intelligible reason to speak of someone as, for example, harboring an unconscious desire to please his analyst, or as manifesting an unconscious belief in the omnipotence of thoughts, and to say that, as Freud brilliantly put it (1905, p. 77), even though in such cases “his lips are silent”, he nonetheless “chatters with his finger-tips”. And one (relatively) straightforward way of justifying such adverbial locutions consists in bringing out *why* the reference to ‘unconscious’ matters in such descriptions, or conversely, whether it would make a considerable *difference* if the possibility of using such qualifications would henceforth be denied (say, for example, if instead of using the term ‘unconscious’, we would be obliged to resort to expressions like ‘unknowingly’ or ‘inadvertently’). Additionally, it is to be expected that all these questions can now be treated in a different and, assuming that our arguments for expressivism are correct, more precise way by taking our previous analysis of detectivism into account and thus reformulating the concept of ‘unconscious’ accordingly, i.e. starting from a non-epistemic perspective on first-person avowals.

Yet, an entirely different philosophical problem seems to arise from the moment we leave this descriptive domain of things we might have done or believed unconsciously, and suddenly decide to talk about beliefs, desires, processes or even ‘subjects’ *of* the unconscious. As Wittgenstein put it:

Imagine a language in which, instead of saying ‘I found nobody in the room,’ , one said ‘I found Mr. Nobody in the room’. Imagine the philosophical problems which would arise out of such a convention. Some philosophers brought up in this language would probably feel that they didn’t like the similarity of the expressions ‘Mr. Nobody’ and ‘Mr. Smith’ (1969, p. 69).

And indeed, the sort of discussions that emerge from this substantialist transformation, including the lingering conceptual *unbehagen* Wittgenstein was hinting at, are sufficiently familiar in psychoanalytic circles: here, of course, we encounter the classical and intractable debates gathered under the heading of ‘the reality of the unconscious’ which have animated a good deal of the philosophical literature. Since the concept is now no longer invoked in order to describe but to *explain* certain phenomena, ‘the unconscious’ is henceforth taken to refer to some kind of intangible entity – depending on one’s humanistic sensitivities and scientific aspirations, a second mind or a more impersonal variety – able to exert a more or less causal influence on my current behavior, secretly operating from what Freud famously called, as a sort of backstage-addendum to the Cartesian theatre, ‘der andere Schauplatz’. Closely following from this conception are the typical queries probing its ontological credentials, i.e. whether or not such a thing exists and if so, to put it in Heideggerian terms, what ‘type of existence’ or ‘manner of being’ we are dealing with (is it a reified thing, a slightly less reified event, a new regulative-transcendental category, ...); what kind of ‘stuff’ it is made of (drive representatives, mnemonic traces, psychical ‘matter’, pure intensities?), by which kind of laws, modes of production or mechanisms (condensation, displacement, metaphor/metonymy, ...) it might be able to produce its effects and what

sort of ‘causality’ (‘psychical’ or the more progressive ‘structural’ sort) is possibly involved here. So, without going into much further detail, it should be clear that insofar as this is indeed the meaning one wishes to assign to ‘unconscious’, the questions that would follow from such a substantial employment would obviously exceed the scope of our analysis, for in that case the justification of the concept coincides, as it should, with the (scientific) demonstration of its explanatory potential.

However, as has been noted by numerous commentators, there is on the other hand ample reason to resist reducing the discussion on the logical status of the unconscious to this substantialist locution, not merely because Freud’s arguments for introducing such a notion are essentially vacuous, but also because that employment has been repeatedly denounced – most notably by Lacan - within psychoanalytic theory itself. Both issues here – i.e. the reasons for both the explanatory vacuity of the unconscious and Lacan’s dissatisfaction with it – can be put in sharper focus by revisiting the well-known arguments by which, very early on, Sartre rejected the Freudian unconscious as a possible explanation for the phenomena of ‘bad faith’ and self-deception. Here’s how Sartre famously described the problem:

By the distinction between the ‘id’ and the ‘ego’, Freud has cut the psychic whole into two. I *am* the ego but I *am* not the id. I hold no privileged relation to my unconscious psyche. [...] Thus psychoanalysis substitutes for the notion of bad faith, the idea of a lie without a liar; it allows me to understand how it is possible for me to be lied to without lying to myself since it places me in the same relation to myself that the Other is in respect to me; it replaces the duality of the deceiver and the deceived, the essential condition of the lie, by that of the ‘id’ and the ‘ego’. It introduces into my subjectivity the deepest intersubjective structure of the Mit-sein. Can this explanation satisfy us? (2003, p. 51)

And the reason why, in short, according to Sartre, the latter rhetorical question must be answered negatively is not so much due to, as it is often portrayed, phenomenological arguments invoking the insulated appeal to ‘lived experience’, but rather thoroughly *conceptual*.<sup>75</sup> More precisely, that whenever the psychoanalytic explanation is indeed fully and explicitly set out, we find that the introduction of the substantial use of ‘the unconscious’ as something that ‘lies’ or ‘deceives’ in my place merely relocates the descriptive facts of a person’s actions and intentions at a putatively sub-personal level. It is then only through the optical-illusory device of partitive sub-personal re-description that the unconscious can be made to *seem* as a genuine explanation; what has been offered in reality is merely a trivial reification of the adverbial meaning of unconscious, one which, moreover, accords quite well with the self-deceiver’s

---

<sup>75</sup> This is why roughly the same line of argument can be found in Wittgenstein who, at the very least, does not share Sartre’s phenomenological outlook: “Can we say that we have stripped bare the essential nature of the mind? ‘Formation of a concept.’ Mightn’t the whole thing have been treated differently?” (Wittgenstein, 1969, p. 45).

own self-misrepresentation (see Gardner, 1993, pp. 41-58).<sup>76</sup> And here, furthermore, we also encounter the main source for Lacan's misgivings with respect to this spontaneous tendency to 'reify' or 'hypostatize' the unconscious. As already evident from Sartre's remarks, there is indeed, apart from this conceptual confusion, a further compelling reason why such a view cannot be endorsed by psychoanalytic theory. For as soon as the concept is understood in this way, a comforting picture imposes itself where the unconscious is situated, so to speak, on the outside, and we ourselves, as 'healthy ego's' or unsuspected spectators, on the inside, thereby able to occupy a transcendent position towards an unconscious which becomes nothing more than, as Sartre put it, an "experimental idea" (2003, p. 51). Yet, as Lacan argues in a Spinozistic way, the unconscious is not what we *qua* conscious egos are set over and against, nor is the Freudian *Spaltung* reducible to a mere shifting from one center of (irrational) agency to another; rather, according to Lacan, this movement of displacement or self-differing is itself precisely what we, *qua* subjects, *are*.<sup>77</sup>

### 3.2 The unconscious structured as a cogito

It will take us a few detours to elucidate such an elusive idea which, as it stands, is neither of immediate relevance to our discussion on first-person authority, nor particularly helpful in answering the question we set out in the beginning. There we asked quite simply what, if anything, we are trying to get hold of by describing someone's belief, pleasure or intention as unconscious. And the outcome of our discussion so far has been mainly negative: nothing in particular is gained through the shift from the adverbial to the substantive use of unconscious that was not already left open in the former, nothing but, perhaps, a momentary illusion of sur-plus explanatory value.

Now it is often suggested that the philosophical relevance of a figure like Lacan stems precisely from the fact that he, in contrast to some of Freud's perhaps rather unfortunate formulations, decisively 'de-substantializes' the unconscious. Although it is seldom made explicit why, in all, this is seen as a conceptual progression in the first place – for example, because of the absurdity of second mind-conceptions, or the daunting prospect of having to furnish a convincing 'de-homunculized' alternative, ... -, it is nonetheless generally agreed that Lacan's putative achievement more specifically consists in the very *way* he proposes to de-substantialize the unconscious, that is, by explicitly engaging

---

<sup>76</sup> To quote the famous lines from Hamlet: "Was't Hamlet wrong'd Laertes? Never Hamlet. / If Hamlet from himself be ta'en away, / And when he's not himself does wrong Laertes, / Then Hamlet does it not. / Who does it then? His madness."

<sup>77</sup> One of Lacan's most clear expressions of this idea can be found in *Seminar 2: The Ego in Freud's Theory and in the Technique of Psychoanalysis*, p. 44: "The core of our being does not coincide with the ego. That is the point of the analytic experience, and it is around this that our experience is organized [...]. But do you think that we should be content with that, and say – the *I* of the unconscious subject is not *me* [*moi*]? That is not good enough, because nothing [...] implies the inverse. And normally you start thinking that this *I* is the real ego. You think that the ego is nothing but an incomplete, erroneous form of this *I*. In this way, you have accomplished the decentring essential to the Freudian discovery, but you have immediately reduced it".



with what, at first blush, seemed directly opposed to it, namely, the modern problematic of the subject as formulated by Descartes.<sup>78</sup> And indeed, compared to Freud's rather indifferent attitude towards whatever it was Descartes hoped to establish by means of his hyperbolic diligence<sup>79</sup>, it is certainly true that Lacan lucidly exploited the ever productive paradoxes which result from the philosophical decision to conceive the apparent certainty of the Cartesian *cogitata* as grounded in a reflexive, transitive relation to oneself. That is to say, precisely that sort of self-reflexive relation on account of which Sartre believed to protect translucent consciousness from encountering any unconscious opacity among one of its self-negating moments, is now deployed in order to expose its 'inner limit' or 'hidden truth'. In sum, instead of an unconscious which simply reduces to an anonymous other whilst leaving the authority of the Cartesian cogito intact, Lacan promotes the latter figure as the most suitable *persona* from which the meaning of 'unconscious' could be deduced; instead of a determining substance as opposed to a reflexive subject, Lacan proposes a 'subversive' unconscious reflexivity<sup>80</sup>, or, to condense the foregoing permutations, a 'subject of the unconscious':

in the term subject ... I am not designating the living substratum needed by this phenomenon of the subject, nor any sort of substance, nor any being possessing knowledge in his pathos, his suffering, whether primal or secondary, nor even some incarnated logos, but the Cartesian subject, who appears at the moment when doubt is recognized as certainty – except that, through my approach, the bases of this subject prove to be wider, but, at the same time much more amenable to the certainty that eludes it. This is what the unconscious is (Lacan, 1981, p. 126).

Now, although it may seem that this particular psychoanalytic version of the cogito constitutes a significant break with a long tradition of exegetical commentary dedicated to determining the precise meaning of Descartes' *fundamentum inconcussum*, it should be apparent however, already from this short quote, that what we are dealing with is actually but one further variation within that same tradition. For when understood as a cogent response to our initial question – i.e. what does it mean for someone's belief (desire, intention, ...) to be unconscious? –, then already one part of Lacan's answer cannot but appear as remarkable. That is, assuming that one horn of that question consists in simply asking 'whose unconscious belief are you talking about?', why is it, after all, that we are suddenly referred to the austere

---

<sup>78</sup> To quote perhaps the most ardent defender of this view - Zizek: "Therein resides Lacan's achievement: the standard psychoanalytic theory conceives the Unconscious as a psychic substance of subjectivity (the notorious hidden part of the iceberg) – all the depth of desires, fantasies, traumas, and so on – while Lacan *de-substantializes* the *Unconscious* (for him, the Cartesian *cogito* is the Freudian subject), thereby bringing psychoanalysis to the level of modern subjectivity" (Zizek, 2016, p. 190).

<sup>79</sup> Cf. Freud (1940a, p. 159): "There is no need to characterize what we call 'consciousness': it is the same as the consciousness of philosophers and of everyday opinion".

<sup>80</sup> Again, in Zizek's words: "[...] the unconscious is not some kind of pre-reflexive, pre-thetic, primitive substrate later elaborated upon by conscious reflexivity; quite the contrary, what is most radically "unconscious" in a subject is his self-consciousness itself, the way he reflexively relates to his conscious attitudes. The Freudian subject is thus identical to the Cartesian *cogito*, or, more precisely, to its later elaboration in Kantian-Hegelian self-consciousness" (Zizek, 2012, p. 554).

figure of the ‘Cartesian subject’? Whence this idea that the subject that truly matters to psychoanalysis is, as Lacan (2006, pp. 870-872) emphasizes, not to be confused with “the suffering subject”, “the subject in relation to his body” or “the subject of love”? Is it, for example, somehow an intrinsic part of psychoanalytic experience that those who participate in that strange *folie-à-deux* turn out to be, on closer inspection, ‘dehumanized’ and ‘disembodied’ subjects devoid of any empirical individuality, strangely akin to those selves of solipsism of whom Wittgenstein (2002, §5.64) aptly remarked that, eventually, they collapse into “points without extension”?

It seems altogether more likely that we have to locate the principal source of this idea *elsewhere*, and more in particular, in those philosophical readings which, as Anscombe pointed out in her classical article on the first person (Anscombe, 1981a), all have tended, in their own distinctive ways, to take the ‘I’ of Descartes famous formula as a referring expression (Ibid., p. 22). In terms of our general discussion, the issue here is closely related to the earlier discussed asymmetry between the first and third persons with respect to psychological verbs, only this time focused on the particular role or function fulfilled by the word ‘I’. As is well known, in the *Blue Book*, Wittgenstein (1969, pp. 66-67) proposes a distinction between two different employments of the first-personal pronoun: on the one hand, those instances where the words ‘I’, ‘me’ and ‘mine’ serve the speaker in talking about himself as an object (e.g. his body) and to attribute some property to this object on the basis of an observation. Examples of this kind of use are: ‘My arm is broken’, ‘I have grown six inches’, ‘The wind blows my hair about’, ... Since such statements are used to convey descriptive information resulting from an observation, they are constructed like every other assertion regarding an individual: they comprise a predicate (whatever it was I did observe) and a subject (the object on account of which I did the observation). Accordingly, there are two ways in which such ‘I-as-object’-statements could be susceptible to errors of identification. Be it on the side of the predicate, because, for instance, my arm turned out not to be broken after all; or on the side of the subject, that is, the arm is indeed broken, but it’s not mine. The latter error is due to the fact that I have confused one person with another, namely myself, and its possibility indicates that what is involved in this type of employment is the reference to and recognition of a particular person. By contrast, the second type of employment, where the speaker appears ‘as subject’, excludes any sort of error with regard to the identity of the person involved: when I say ‘I see someone approaching’ or ‘I believe it will rain’, I might be wrong in the sense that in fact there’s no one there or because my forecast got it wrong, but certainly not because the one who was seeing or believing turned out to be someone else after all. Or, as Anscombe illustrates (1981a, p. 30), when, during a dinner, a bishop lays down his hand on the lady’s knee, he could try to flee the embarrassing moment by claiming that he took the lady’s knee for his own, but, in any case, *not* by conceding that he mistook himself for the lady in laying down his hand.

Now this fundamental distinction with respect to the different ways we speak about ourselves has often been translated in such a way as to make it congenial to a traditional theory of the subject. For it is clear that, thus presented, the opposition between these two different kinds of employment of the

first-personal pronoun not only seems to correspond with the distinction between physical properties (having a broken arm, measuring a certain height, ...) and psychological properties (seeing, believing, ...), but also between two types of 'subjects' or bearers of these properties. That is, between someone who can be identified and referred to in a way that is symmetrical for the first and third persons, i.e. the objective-empirical person, and someone who can only be recognized and referred to in a special first-personal way and who is, *for that very reason*, also a very special person, i.e. a Cartesian subject or ego. This last point is particularly important because it enables us to see not only why, as Lacan rightly emphasizes (cf. *supra*), such a subject defined in terms of an exclusive first-personal reference cannot have anything to do with the concept of a concrete human being, but also why it is precisely on the basis of a certain radicalization of this referential thesis that Lacan wants to establish his idea of the unconscious (cf. *infra*).

To render this conspicuous, we have to bring out why the Cartesian self is not so much the idiosyncratic outcome of Descartes' purifying reductions, but rather of the philosophical attempt to satisfy two basically incompatible conditions in explaining the seemingly guaranteed success of the 'I-as-subject'-use. That is, the Cartesian explanation presupposes, on the one hand, (i) that 'I', in examples of its use as subject (e.g. 'I believe it will rain', 'I doubt whether I exist', ...), *like* uses of other subject-terms in grammatically similar ascriptions (e.g. 'I have a bump on my forehead', 'my eyes are blue', 'my name is René', ...), purport to refer to an object of which something is to be predicated. This in turn presupposes that such self-ascriptions must rely on some form of access to, or recognition of, the referent of 'I'. Yet, on the other hand, (ii) the distinctiveness of this kind of 'I'-ascription relative to, for example, the present tense use of 'I-as-object', other-ascriptions or past-and-future tense ascriptions, is held to be a matter of epistemic security: *unlike* these other referring expressions, the use of 'I' as subject is immune to error through misidentification, which is to say that in this case its referential success is thought to be guaranteed. However, from the effort to *combine* these two conditions, i.e. semantic continuity and epistemic asymmetry, in explaining what is specific to the employment of the 'I-as-subject', it necessarily follows (iii) that the subject who is referred to in such ascriptions cannot be individuated or identified on the basis of *any* identity criteria, since, as a matter of conceptual or logical necessity, talk about reference and identification implies the possibility of their failure or misapplication. Hence the idea, pointed out by Lacan and several other Descartes interpreters, that the ego which gloriously survives the increasingly stringent stages in the skeptical cogito-procedure, has nothing in common whatsoever with the historical-empirical figure of Descartes, nor with any conceivable personal identity: such an ego is said to be 'purified' from all possible determinations, like Pascal's 'ego without qualities', Kant's formal 'I-think' or Husserl's 'reine Ich'.

Yet, leaving the philosophical genealogy of the cogito aside for a moment, this also means that the prospect of getting a satisfactory answer to our initial question suddenly appears to be rather dim. Since if that answer, at least for the part that interest us here, consists in claiming that the subject we are dealing with is an existence without determination, an empty form deprived of all substance and content,

in short: a thinking or doubting in general, then it is not only that subject which appears as 'void', but, it seems, so does our whole questioning. For imagine, for example, what such an answer would amount to for a psychoanalyst who wants to distinguish between two of his patients, or perhaps between himself and his patient, while being forced to proceed on the basis of the Cartesian clue 'a thinker or thinking in general'. It is clear that even a simple task like this would quickly reveal itself as being utterly hopeless. Whatever other theoretical need it is supposed to fulfill, the referential description 'a subject in general' is of course completely useless in distinguishing one subject from another, for all identity criteria which would allow us to answer that question have been expelled, but the theoretician of the subject, whether he wants it to be conscious or unconscious, has failed to give us a proper alternative. Being a 'subject in general', therefore, is perhaps to be a subject in a contrived philosophical sense, but not the analyst, not the patient, nor anyone else. As Descombes put it (2014, p. 68, our translation), it is rather like having an "address in general", that is to say, of living somewhere, but nowhere in particular. And it should be clear that this is not an 'ontological' invitation to henceforth distinguish between two types of addresses, those that are particular and those that are general, but rather to question how we ended up with such non-localizable entities in the first place.

### 3.3 The cogito structured as an unconscious

Already one part of that answer we tried to elucidate above: in trying to account for the difference between the first and third-person uses of the first-personal pronoun, the Cartesian subject emerged as a sort of unstable 'compromise formation' resulting from the attempt to straddle both sides of the following fence at once. On the one hand, aiming to preserve semantic continuity, the expression 'I', like all other referential expressions, is considered to be the term through which a speaker or thinker refers to himself. Yet, on the other hand, aiming to preserve epistemic asymmetry, such reference cannot have any of the ordinary features normally attached to this kind of operation: as Anscombe remarks, 'I' is supposed to have "sure-fire"-reference (1981a, p. 30), a remarkable quality which in turn seems to contradict its professed semantical function. Furthermore, unwilling to confront this precarious dilemma in which epistemic asymmetry undermines referential symmetry and vice versa, the Cartesian is moved to advance an extra-ordinary referent corresponding to his idea of extra-ordinary reference: this is the Cartesian subject or ego.

We pointed out how this referential interpretation of the 'I' seems to saddle us with a mysterious entity which, despite remaining certain in its doubtfulness, is nonetheless entirely useless when it comes to answering even the most basic identity questions. So, in so far as the issue here does not merely pertain to the particularities of the alleged outcome of Descartes' Second Meditation, but, as Lacan suggests, equally concerns the everyday practice of psychoanalysis, then at least one clear undesirable consequence of embracing this figure as the principal term of reference in psychoanalytic theory is that

the dialogical situation between an analyst and his patient would inevitably collapse into an unmanageable state of identity confusion.

Be that as it may, this still leaves us with the task of unravelling in what sense Lacan's peculiar deduction of the meaning of 'unconscious' from the formula of the Cartesian cogito is to be understood. For it is clear, of course, that Lacan is not in the Cartesian business of proving he has a soul, nor intent on shoring up any indubitable entity that could come to function as an initial bedrock for subsequent world-involving epistemological deliberations. The intention here is rather to denounce what he holds to be the "initial error of philosophy" (Lacan, 1987, p. 107)<sup>81</sup>, which, moreover, apparently consists in the ever-renewed attempt to reduce the difference between "subject and consciousness" (Lacan, 1987, p. 106):

At a crucial point of the Cartesian *askesis*, [...] consciousness and subject coincide. It is holding that privileged moment as exhaustive of the subject which is misleading – making of it the pure category that the presence of a gaze (as a mode of opaqueness within the visible) would come to make flesh with its vision [...]. It is, on the contrary, at that moment of coincidence itself, in so far as it is grasped by reflection, that I intend to mark the site through which psychoanalytic experience makes its entrance. At simply being sustained within time, the subject of the "I think" reveals what it is: the being of a fall. I am that which thinks "Therefore I am" [...] noting that the "therefore", the causal stroke, divides inaugurally the "I am" of existence from the "I am" of meaning (Ibid., pp. 106-107).

Yet, here already, and despite the air of radical 'subversion' which is supposed to transpire throughout these lines, it is remarkable that Lacan actually *shares* the detectivist gesture of invoking 'consciousness' as a special medium of epistemic access ("the presence of a gaze", "grasped by reflection", ...) in explaining what allows Descartes, or any other person for that matter, to declare that 'I think ("therefore I am")'. Secondly, we also encounter the referentialist framework within which such detectivist ideas are able to flourish: saying that 'I think ("therefore I am")' is, on that account, to be considered as a linguistic act through which I refer to, and identify with, a person (i.e. 'myself', or, to keep with Lacan's third-personal inversion 'my *self*') under a particular description<sup>82</sup>, in this case 'I am that which thinks "therefore I am"'. So, revealingly, it is in any case *not* this traditional account of first-personal reference

---

<sup>81</sup> See also Miller's diagnosis of this same 'error': "What is serious for Lacan is the logic of the signifier, that is to say the opposite of philosophy, inasmuch as every philosophy rests on the appropriateness, transparency, agreement, harmony of thought with itself. There is always some part hidden, in a philosophy, an I = I, which constitutes what Lacan called at some moment "the initial error in philosophy" which consists in privileging this equality and thus making one believe that the "I" is contemporary with itself, while its constitution is always after the emergence of its cause, of *petit a*" (Miller, 2006, pp. 22-23).

<sup>82</sup> See also Lacan's seminar on *Identification* where this idea is rendered even more explicit: "Therefore – as we have said – this sentence of: "I think" has the interest of showing us – it is the least we can deduce from it – the *voluntary dimension of judgment*" (Lacan, 1961-1962, p. 9, our italics). Elsewhere in the same seminar, Lacan coins the neologistic pun "mihilisme" to qualify this operation: "Thus we see here, in short in our tongue, a sort of identification through the operation of a special significant tendency, that you will allow me to qualify as "mihilisme" in so far as to this act, this experience of the ego is referred" (Ibid., p. 4).

which might be critically questioned and identified as philosophy's "initial error", but rather that of holding this detectivist moment as "exhaustive of the subject" (Lacan, 1987, p. 106). Or, differently put, not the idea that I, in speaking about myself - whether this means saying things like 'I think I exist', or less dramatically, 'I'm not feeling too well', 'I believe it will rain', ... - have to engage in a sort of mysterious auto-reflexive procedure at the end of which, on the one hand, the term 'I' serves to designate or name the person who is currently speaking<sup>83</sup>, and its verbal complements ('not feeling too well', 'believing it will rain'), on the other hand, function as descriptive predicates filling out the first-personal proposition, but the idea that this procedure is actually able to attain its self-reflexive intention.

And indeed, doubts as regards the feasibility of this kind of auto-referential operation are easily generated on account of a famous paradox to which such views unavoidably lead. For if speaking about myself in the first-person is represented as yet another version of third-personal reference, that is, if, in saying 'I know I exist' or 'I believe it will rain', I refer to myself as someone who knows or believes thus, to *who* then exactly does this reference apply? The subject, for example, who, in turning his attention to himself, knows that he believes he exists, does he also believe he exists? Or, to take a more clinically relevant example, the patient who, during his weekly session with the analyst, suddenly avows his intention to change his life, does he also want to change his life? Moreover, given this idea of first-personal reference, could he even be imagined as ever being able to answer that question? Obviously not, for if we are to follow Lacan's reconstruction of the cogito, such a patient is forever condemned to occupy a spectral position of perpetual self-commentary with regard to himself, hence always missing out on his *true* position as commentator.<sup>84</sup> And the reason for this is clear: the effort to align the use of the first-person with the third-person has, as its ultimate consequence, the effect of installing a 'division' of the subject, separating the one 'who is speaking' from the one 'who is spoken of'. Or more precisely: on such an account, when 'I' speak about 'myself', the 'I' and the 'self' do not share the same reference (see also Silveira-Sales, 2007). And here, finally, we also encounter that 'self-differing' movement (cf. *infra*) which, according to Lacan, fundamentally characterizes our position as speaking subjects: wanting to speak about myself in the first-person, and believing to express myself in the first-person present tense, I am irrevocably condemned to speak about myself in a third-personal way, that is, of referring to a 'self' or 'ego' from which, by the simple fact of this act of self-reference, I am immediately detached *qua* speaking subject. Putting this all together, Lacan asks:

---

<sup>83</sup> See also Lacan's "linguistic definition of *I* as signifier": "In a concern for method, we can try to begin here with strictly linguistic definition of *I* as signifier, where it is nothing but the shifter or indicative that, qua grammatical subject of the statement, *designates* the subject insofar as he is currently speaking" (Lacan, 2006, p. 677, our italics).

<sup>84</sup> As indeed was already the conclusion drawn by that other eminent philosopher of reflexive paradoxes, i.e. Sartre: "Thus the consciousness that says "I think" is precisely not the consciousness which thinks. Or, rather, it is not its own thought which it posits by thisthetic act" (Sartre, 2004b, p. 6). The only difference separating Lacan from Sartre with regard to this issue of first-personal reference is therefore related to a disagreement concerning the concept of 'pre-reflexive consciousness'. Cf. *supra*.

Is the place that I occupy as subject of the signifier concentric or eccentric in relation the place I occupy as subject of the signified? That is the question. The point is not to know whether I speak of myself in a way that conforms to what I am, but rather to know whether, when I speak of myself, I am the same as the self of whom I speak (Lacan, 2006, p. 430).

Yet, despite the structuralist tropes invoked in dressing up this question, this is merely to offer a more convoluted version of the optical metaphor which sustains the detectivist idea of the unconscious: in the same way as the ‘eye’, in relation to its visual field, ‘sees everything except itself’ (i.e. Lacan’s “as a mode of opaqueness within the visible”); or, in the same way as the ‘I think’, in order to be able to accompany all my representations, cannot accompany its own act of representation, etc., the speaking subject can only refer to itself on the condition of remaining oblivious to its own position of enunciation.<sup>85</sup> The detectivist unconscious and the paradoxes generated by the idea of first-personal reference are then but two sides of the same coin: wherever self-knowledge is supposed in order to explain referential success, unconscious ignorance comes to haunt this very act of self-reference.

#### **4. Conclusion: expression and the unconscious**

The latter conclusion might be alarming for anyone who takes it upon himself to secure his existence by means of the Cartesian askesis, but for our present purposes, this also means that the account of the unconscious we get from Lacan’s restatement of the cogito in terms of the enchanting formula “I am thinking where I am not, therefore I am where I am not thinking” (2006, p. 340) is, in the end, clearly parasitic on the detectivist picture of first-person authority. That is to say, it is indeed a correction *within* that picture, but not a correction *of* that picture. For it is not a matter of dispute that I can think and talk about myself and that the use of the first-person is the most ordinary way to do so. Whether I say I have the intention to go on a holiday or desire to finish this article, I indeed talk about myself, about my future intentions and current wishes. Nothing in particular has been decided by putting the issue in this way, because no philosophical account of this ability has been advanced either. However, does this necessarily mean, as Lacan seems to suggest, that in order to talk about myself, I have to *refer* to myself? Is speaking or thinking in the first-person to be considered as a referential activity in which I have to identify myself *with* myself and in which the ‘I’ comes to function as a proper name and its psychological complements as descriptive predicates? Or, to put this in terms of Lacan’s quarrel with the philosophies of consciousness: where in fact do we have to situate the ‘initial error of philosophy’?

---

<sup>85</sup> As Lacan puts it: “Is this not designed to make us question ourselves about what is at stake at this radical, archaic point that we must necessarily suppose to be at the origin of the unconscious, namely of this thing through which, in so far as the subject speaks, all he can do is to advance further along the chain, in the unfolding of enunciations, but that, directing himself towards the enunciations, by this very fact in the enunciating, he elides something which is properly *what he cannot know, namely the name of what is qua enunciating subject*” (1961-1962, p. 62, our italics).

In the discussion about whether or not first-personal reference can ever attain that mythical moment of self-coincidence, or, indeed already, as Wittgenstein suggests, in the very idea of self-reference?

To clarify this issue, let us consider the following two sentences:

- (i) It will rain
- (ii) I believe it will rain

The relevant question here is whether there is any difference with regard to the object of these two assertions. At a first impression, and because of the difference in their grammatical form, the two sentences at least *seem* to be different: the first one talking about the weather, the second one, on the other hand, containing a psychological proposition specifying my belief. However, according to Wittgenstein, by themselves, that is, without considering the context of use of these sentences, these grammatical appearances prove nothing. Yet, let us first develop Lacan's suggestion that sentences of the second form are referential descriptions which predicate something about an 'ego' of which it is said that it 'believes it will rain'. Here, it is strictly irrelevant whether this also means 'alienating' oneself in language, the signifier, the mirror image or any other estranging medium held responsible for generating that unpalatable pathos so typical of Lacanian thinking; the same conclusion will hold for the Cartesian who does believe in the contemporaneity of 'subject and consciousness'. Suppose, furthermore, that we have two people, the one affirming 'I believe it will rain', the other maintaining 'No, I don't think so'. If we are to follow Lacan's account, these two speakers do not, and obviously even cannot, speak about the same thing: although, as we have seen, their true positions as speakers might remain unconscious, both can only refer to themselves, to the beliefs they report in describing them. But this is also to say, on the one hand, that the possibility of two speakers ever contradicting each other has been left behind, and indeed, as is now apparent, that the threat of solipsism has become unavoidable.

Assuming that even for the psychoanalyst dealing with psychic reality, the latter consequence is too high a price to pay, we have to reject the picture which seems to hold us captive here, which is in this case the assumption of semantical continuity across the first –and third-person uses of the psychological verb 'believe'. That is, it is only because the sentence 'I believe it will rain' is analyzed as a more private version of 'He believes it will rain' that we seem to get into this philosophical mess of which, as we have tried to show, both unidentifiable subjects and communicative solipsism are the undesirable outcomes. Therefore, rejecting semantical continuity means recognizing that, grammatical appearances notwithstanding, the verb 'believe' does not signify the same thing in the first- and third-person versions of 'believing it will rain'. In the first-person present tense, 'I believe' indeed *expresses* a belief, but this is not a belief referring to myself, nor to some meteorological belief-state, but quite evidently, a belief *about* the weather. But this also means, first of all, that, in contrast to the obvious difference between the sentences 'He believes it will rain' and 'It will rain', sentences (i) and (ii) above are, from the perspective of their use and hence of their meaning, perfectly identical: the prefix 'I believe'



adds nothing to the meaning already found in 'It will rain', which is to say that it is entirely redundant. Yet, secondly, as per the Lacanian idea of the unconscious, this means that the subversive effort to lodge a moment of opacity within the temporal movement of self-reference is equally bound to failure: not because in saying 'I believe it will rain' I am somehow magically protected from errors through misidentification, but simply because in this case there is no self-reference and hence no 'identification' involved in the first place.<sup>86</sup>

Yet, does this observation imply that the effort to offer a philosophical clarification of the adverbial meaning of unconscious should now be abandoned? It seems, on the contrary, that the only thing we have effectively abandoned is a distorted view on the first-person and, by implication, on the meaning of unconsciousness, which keeps producing nothing but conceptual incongruities. Recognizing what distinguishes speaking about oneself in the first person present indicative from third-personal reference is, first of all, to reject the detectivist picture of the unconscious in *all* its different guises, whether this means the deficiency of inner sense as in the earlier quote by Freud, or its more elaborate Lacanian version in terms of the paradoxes of self-reflection and first-personal reference. Speaking about someone else, I indeed refer to and identify someone by predicating something about his intentions and beliefs; but in my own case, none of these epistemic procedures are involved, and so the idea of the unconscious as involving beliefs, intentions or Cartesian subjects which would escape epistemic self-scrutiny cannot be involved either. However, dissociating the adverbial qualification of unconscious from failures of self-knowledge in this way, also means to bring out the truth of Freud's following incisive observation:

From what I have so far said a neurosis would seem to be the result of a kind of ignorance – a not knowing about mental events that one ought to know of. [...] Now it would as a rule be very easy for a doctor experienced in analysis to guess what mental impulses had remained unconscious in a particular patient. So it ought not to be very difficult, either, for him to restore the patient by communicating his knowledge to him and so remedying his ignorance. [...] If only that was how things happened! We came upon discoveries in this connection for which we were at first unprepared. Knowledge is not always the same as knowledge: there are different sorts of knowledge, which are far from equivalent psychologically. [...] If the doctor transfers his knowledge to the patient as a piece of information, it has no result. [...] The patient knows after this what he did not know before – the sense of his symptom; yet he knows it just as little as he did. Thus we learn that there is more than one kind of ignorance. We shall need to have a somewhat deeper understanding of psychology to show us in what these differences consist (Freud, 1966, pp. 280-281).

---

<sup>86</sup> Anscombe sums up this point as follows: "Getting hold of the wrong object *is* excluded, and that makes us think that getting hold of the right object is guaranteed. But the reason is that there is no getting hold of an object at all" (1981a, p. 32).

So, if, as well-attested by Freud, what it is for a desire or belief of mine to be unconscious is not primarily a matter of any straightforward first-personal epistemic shortcoming; moreover, if, on the contrary, knowledge as to my present unconscious inclinations seems perfectly compatible with the very existence of such inclinations, then what is the qualification ‘unconscious’ a qualification *of*? Or, as we put it earlier on in this paper, what exactly would be lost from our descriptive psychological repertoire if we opted to replace the term ‘unconsciously’ by the more innocuous ‘unknowingly’? Would we effectively miss out on certain kinds of human behavior if our conceptual vocabulary was to be impoverished in this way?

We will conclude our article by arguing that this is indeed the case and that such examples can serve as an important step towards clarifying the adverbial meaning of unconsciousness. Here, both the subtle analyses of the phenomenon of ‘belief disavowal’ by the French psychoanalyst Mannoni<sup>87</sup> and Wittgenstein’s grammatical remarks on the concept of unconscious<sup>88</sup> can provide some decisive suggestions. To begin with the former: by means of the formula ‘I know quite well, but still ...’, Mannoni aimed to capture a number of paradoxical, though seemingly entirely common situations in which someone appears to simultaneously entertain two contradictory beliefs. As, for example, the enlightened reader of horoscopes who ‘knows quite well’ that cosmological predictions of future life events are sheer nonsense, yet still seems to take a strange pleasure in reading them anyway. In his own case, Mannoni observed that when the horoscope predicted that a certain day on which he had planned a move was indeed “particularly favorable for changes at home” (1985, 20), he fell into strange fit of laughter. Yet, he also admits that his laugh would have been different if the horoscope had strongly advised against such domestic intentions. Or, in a similar example provided by Pfaller (2014, p.1), imagine that, while reading a newspaper, a friend arrives and says: “Excuse me, can I have a quick look at your newspaper? I know it’s silly, but I just have to know the score from yesterday’s game”. So, with regard to our present discussion, what needs to be highlighted in these examples is the paradoxical relation a subject seems to entertain with its beliefs (i.e. the importance of sports; the significance of horoscope predictions). In a way that is difficult to render conceptually transparent, these sorts of beliefs are never really believed in (‘I know it’s silly’; ‘I know these coincidences are meaningless’), yet in one way or another, they nonetheless exert a particular influence on their subjects (the compulsion to look at the paper – the pleasurable laughter). Moreover, as Pfaller argues (Ibid.), it is not so much their (trivial) content that is primarily of interest here, but rather the *form* in which people seem to refer to these beliefs. That is, as was already apparent from Mannoni’s formula for this type of situations, its form is characterized by a rather unsettling coexistence of ‘better knowledge’ and ‘belief’. The fidget sports

---

<sup>87</sup> For Mannoni’s theory of disavowal, see his two essays “Je sais bien, mais quand même ...” and “L’illusion comique” in Mannoni (1985). For an English translation of the first essay, see Mannoni (2003, pp. 68-92). An important and lucid elaboration of Mannoni’s account can be found in Pfaller (2014).

<sup>88</sup> For some of Wittgenstein’s scattered remarks on a possible grammatical use of unconscious in relation to Moore’s paradox, see Wittgenstein (1988, pp. 65-67). See also Descombes’ brilliant discussion of this connection to which we are indebted for this point, in Descombes (2002).

fanatic knows quite well that yesterday's results are not important, but still he *has* to see them. Despite the better knowledge, and despite the ironical distance between him and his silly practice, he nevertheless acts *as if* sports are of utmost importance.

Importantly, in his own way, Wittgenstein seems to have seized on precisely these sorts of situations in order to justify a possible use of the adverb 'unconsciously'. Starting from the observation that, as we have argued, in the first-person present indicative, the verb 'believe' behaves in a different way compared to its third-personal form, Wittgenstein however also envisaged whether it would be possible to construct an intelligible use of 'I believe' which would be symmetrical to the third-personal 'He believes'. In order to create such a form, Wittgenstein proposes that we could modify the verb 'believe' in the first person by adopting the expression "I seem to believe" or "going by my behavior, this is what I believe":

My attitude to my own words is wholly different from that of others. I could find that variant conjugation of the verb, if only I could say "I seem to believe". If I listened to the words issuing from my mouth, then I could say that someone else was speaking out of it. "Judging from my words, this is what I believe". Now, it would be possible to think up circumstances in which this made sense (Wittgenstein, 2009, II§103-105).

And the sort of circumstances Wittgenstein deems relevant here are exactly those in which it would make reasonable sense to talk about 'unconscious beliefs' or 'unconscious intentions'. That is, 'I seem to believe' is the expression I could use whenever, like in the situations described by Mannoni, I have to conclude, rather to my own surprise and despite better judgment, that I *seem* to behave as someone who believes this or that. Hence, what the qualification 'unconscious' aims to capture in such instances is, on the one hand, the fact that we come to learn about these beliefs or desires in the same way as we would proceed in the case of others, that is, by observing and reporting on what is revealed to me in my own behavior and immediate reactions. Here, therefore, and in contrast to the use of the first person in the present indicative, it *does* make perfect sense to say that I 'identify' myself as someone who unconsciously believes or desires certain things, for it is part of the very meaning of unconsciousness that I have to be informed about my beliefs and desires in such an alienated, third-personal way.<sup>89</sup> Yet, on the other hand, this also means, vindicating Freud's observation, that the idea of unconsciousness which is at stake here is in an important sense *non-epistemic*. On the contrary, since we have shifted to a symmetrical employment of psychological verbs, talk about my unconscious beliefs is precisely talk about what I, in principle, might come to *know* about myself through an attentive examination of my own expressive behavior or through the suggestions provided by the analyst. Nevertheless, as Freud had to acknowledge, acquiring such knowledge seems to be of little avail with regard to the efficacy of the beliefs in question. So, applied to Mannoni's idiom, what seems to single out those unconscious beliefs

---

<sup>89</sup> And not, contra Lacan, the other way around: i.e., because I always *have* to identify myself in one way or another, I necessarily *misidentify* myself qua unconscious subject.

which would complement the expression ‘I know quite well, but still ...’ from those beliefs which, for example, are already expressed in the self-ascription ‘I know quite well’, cannot be related to a lack of detectivist introspection, nor to any paradox of self-reflection, but rather to something I seem unable to *do*. More precisely, although I do *express* these beliefs and desires in my actions, for example, in the urge to read the newspaper, or in the fading smile while reading disappointing cosmological prospects, what makes them specifically *unconscious* seems related to the fact that I am unable to express those beliefs and desires in a particular way, that is, in expressive self-ascriptions while using the first person present. Although I can describe my unconscious mental states through the use of symptomatic expressions like ‘I know, but still ...’, I am unable to talk out about them in linguistic acts of self-expression, that is, by expressing my unconscious belief in simply *saying* ‘I believe’. In sum: on the view we have advanced here, an unconscious state of mind is not primarily a matter of something which I, qua epistemic subject, fail to *acknowledge*, but rather of a specific expressive position which I, as speaker, cannot *occupy*.<sup>90</sup>

---

<sup>90</sup> Thanks to Dominiek Hoens, Russell Grigg and Vincent Descombes for their valuable suggestions.

## References

- Alston, W.P. (1967). "Expressing", in M. Black (Ed.), *Philosophy in America*. Ithaca, NY: Cornell University Press.
- Anscombe, G.E.M. (1963). *Intention*. Cambridge, Massachusetts: Harvard University Press.
- Anscombe, G.E.M. (1981a). "The First Person", in *Metaphysics and the Philosophy of Mind*. Oxford: Basil Blackwell, 21-36.
- Anscombe, G.E.M. (1981b). "Events in the Mind", in *Metaphysics and the Philosophy of Mind*. Oxford: Basil Blackwell, 57-63.
- Bar-On, D. (2004). *Speaking My Mind: Expression and Self-Knowledge*. Oxford: Clarendon Press.
- Bouveresse, J. (1976). *Le Mythe De L'Intériorité. Expérience, signification et langage privé chez Wittgenstein*. Paris: Les Editions de Minuit.
- Bouveresse, J. (1995). *Wittgenstein Reads Freud: The Myth of the Unconscious*. Princeton, New Jersey: Princeton University Press.
- Brentano, F. (2009). *Psychology from an Empirical Standpoint*. London and New York: Routledge.
- Deleuze, G. (1990). *Expressionism in Philosophy: Spinoza*. New York: Zone Books.
- Deleuze, G. & Guattari, F. (1977). *Anti-Oedipus*. Minneapolis: University of Minnesota Press.
- Feyaerts, J. & Vanheule, S. (2015). How to return to subjectivity? Natorp, Husserl and Lacan on the limits of reflection. *Theory & Psychology*, 753-774.
- Dennett, D. (1991). *Consciousness Explained*. New York: Back Bay Books/Little, Brown and Company.
- Descombes, V. (2002). "Wittgenstein face au paradoxe de Moore", in J. Bouveresse, S. Laugier & J-J. Rosat (Eds.), *Wittgenstein, dernières pensées*. Marseille : Agone, 209-235.
- Descombes, V. (2004). *Le Complément de Sujet. Enquête sur le fait d'agir de soi-même*. Editions Gallimard.
- Descombes, V. (2014). *Le Parler de Soi*. Editions Gallimard.
- Doyle, J. (2012). "On Finkelstein's Account of the Distinction between Conscious and Unconscious States of Mind", in G. Abel & J. Conant (Eds.), *Rethinking Epistemology 2*. Berlin/Boston: Walter de Gruyter GmbH & Co, 119-132.
- Eshleman, M.C. (2007). The Cartesian Unconscious. *History of Philosophy Quarterly* 24 (2), 169-187.
- Finkelstein, D. (1999). On the Distinction between Conscious and Unconscious States of Mind. *American Philosophical Quarterly* 36 (2), 79-100.
- Finkelstein, D. (2003). *Expression and the Inner*. Cambridge, Massachusetts: Harvard University Press.
- Freud, S. (1905). Fragment of an analysis of a case of hysteria. *Standard Edition* 7, 1-122.
- Freud, S. (1915). The Unconscious. *Standard Edition* 14, 166-215.
- Freud, S. (1940a). An outline of psycho-analysis. *Standard Edition* 23, 144-207.
- Freud, S. (1966). *Introductory Lectures on Psychoanalysis*. New York: W.W. Norton.
- Gardner, S. (1993). *Irrationality and The Philosophy of Psychoanalysis*. Cambridge: Cambridge University Press.
- Gardner, S. (2012). Psychoanalytic Theory: A Historical Reconstruction. *The Aristotelean Society* 86 (1), 41-60.
- Hacker, P.M.S. (2005). "Of knowledge and knowing that someone is in pain", in A. Pichler & S. Säätelä (Eds.), *Wittgenstein: The Philosopher and his Works*. Bergen: Wittgenstein Archives of the University of Bergen, 203-235.
- Henry, M. (1973). *The Essence of Manifestation*. The Hague: Martinus Nijhoff.
- Henry, M. (1993). *The Genealogy of Psychoanalysis*. Stanford: Stanford University Press.
- Husserl, E. (1970). *Logical Investigations*. New York: Humanities Press.
- Husserl, E. (1976). *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy. Book 1: General Introduction to a Pure Phenomenology*. The Hague: Martinus Nijhoff.
- Kant, I. (1999). *Critique of Pure Reason*. Cambridge: Cambridge University Press.
- Kriegel, U. (2003). Consciousness, higher-order content, and the individuation of vehicles. *Synthese* 134, 477-504.

- Lacan, J. (1961-1962). *The Seminar 1961-1962, Book IX, Identification*. Unpublished translation by Cormac Gallagher.
- Lacan, J. (1981). *The Seminar 1964, Book XI, The Four Fundamental Concepts of Psychoanalysis*. New York & London: W.W. Norton & Company.
- Lacan, J. (1987). Responses to Students of Philosophy concerning the Object of Psychoanalysis. *October* 40, 106-133.
- Lacan, J. (1988). *The Seminar 1954-1955, Book II, The Ego in Freud's Theory and in the Technique of Psychoanalysis*. Cambridge: Cambridge University Press.
- Lacan, J. (2006). *Ecrits*. New York and London: W.W. Norton.
- Laplanche, J. & Pontalis, J-B. (1988). *The Language of Psychoanalysis*. London: Karnac Books.
- Leibniz, G.W. (1981). *New Essays on Human Understanding*. Cambridge: Cambridge University Press.
- MacIntyre, A. (2004). *The Unconscious: A Conceptual Analysis*. New York and London: Routledge.
- Mannoni, O. (1985). *Clefs pour l'Imaginaire ou l'Autre Scène*. Mayenne : Seuil.
- Mannoni, O. (2003). "I know well, but all the same ...", in M.A. Rothenberg, D.A. Foster & S. Zizek (Eds.), *Perversion and the Social Relation*. Durham, NC: Duke University Press, 68-92.
- Miller, J-A. (2006). Profane Illuminations. *Lacanian Ink* 28, 10-25.
- Overgaard, S. (2004). Exposing the Conjuring Trick: Wittgenstein on Subjectivity. *Phenomenology and the Cognitive Sciences* 3 (3), 236- 286.
- Pfaller, R. (2014). *On the Pleasure Principle in Culture: Illusions Without Owners*. London, New York: Verso.
- Redding, P. (1999). *The Logic of Affect*. Ithaca: Cornell University Press.
- Renaudie, P-J. (2009). Expression et description: Remarques sur les limites linguistiques de la phénoménologie. *Bulletin d'analyse phénoménologique* 5 (3).
- Rosenthal, D. (1993). "Thinking That One Thinks", in M. Davies & G.W. Humphreys (Eds.), *Consciousness*. Oxford: Basil Blackwell, 197-223.
- Ryle, G. (2000). *The Concept of Mind*. Chicago: The University of Chicago Press.
- Russell, B. (1949). *The Analysis of Mind*. London: Allen & Unwin.
- Sartre, J.P. (2003). *Being and Nothingness: An Essay in Phenomenological Ontology*. London: Routledge.
- Sartre, J.P. (2004a). *The Imaginary: A Phenomenological Psychology of the Imagination*. London and New York: Routledge.
- Sartre, J.P. (2004b). *The Transcendence of the Ego: A sketch for a phenomenological description*. London and New York: Routledge.
- Schelling, F.W.J. (1978). *System of Transcendental Idealism (1800)*. Charlottesville: University Press of Virginia.
- Shoemaker, S. (1996). *The First-person Perspective and Other Essays*. New York: Cambridge University Press.
- Silveira-Sales, L. (2007). A propos du sujet – note sur le conflit – détermination et subjectivité chez Jacques Lacan. *Essaim* 18 (1), 147-164.
- Smith, D.W. (1986). The structure of (self-)consciousness. *Topoi* 5 (2), 149-156.
- Thomasson, A.L. (2000). After Brentano: A one-level theory of consciousness. *European Journal of Philosophy* 8 (2), 190-210.
- Thomasson, A.L. (2006). Self-Awareness and Self-Knowledge. *Psyche* 12 (2), 1-15.
- Wittgenstein, L. (1966). *Lectures and Conversations on Aesthetics, Psychology and Religious Belief*. Oxford: Blackwell.
- Wittgenstein, L. (1969). *The Blue and Brown Books*. New York: Barnes and Noble.
- Wittgenstein, L. (1981). *Zettel*. Berkeley: University of California Press.
- Wittgenstein, L. (1988). *Wittgenstein's Lectures on Philosophical Psychology, 1946-1947*. Ed. P. T. Geach. Hemel Hempstead: Harvester Wheatsheaf.
- Wittgenstein, L. (2002). *Tractatus Logico-Philosophicus*. London and New York: Routledge.
- Wittgenstein, L. (2009). *Philosophical Investigations (4<sup>th</sup> Edition)*. Oxford: Wiley-Blackwell.
- Wright, C. (2001). "On Making Up One's Mind: Wittgenstein on Intention", in *Rails to Infinity: Essays on Themes from Wittgenstein's Philosophical Investigations*. Cambridge, MA: Harvard University Press.

- Zahavi, D. (2014). *Self and Other: Exploring Subjectivity, Empathy and Shame*. Oxford: Oxford University Press.
- Zizek, S. (1999). *The Ticklish Subject: The Absent Centre of Political Ontology*. London: Verso.
- Zizek, S. (2012). *Less Than Nothing. Hegel and The Shadow of Dialectical Materialism*. London: Verso.
- Zizek, S. (2016). "Afterword: Objects, Objects Everywhere", in A. Hamza & F. Ruda (Eds.), *Slavoj Zizek and Dialectical Materialism*. London: Palgrave Macmillan, 177-192.
- Zupancic, A. (1996). "Philosophers' Blind Man's Buff", in R. Salecl & S. Zizek (Eds.), *Gaze and Voice as Love Objects*. Durham N.C.: Duke University Press, 32-58.





## Discussion and Conclusion

---

Throughout the different chapters comprising our thesis, we have discussed the concept of subjectivity as it emerges in the phenomenological, psychoanalytic and neurocognitive traditions. The specific questions which variably featured in these conceptual analyses included

- (1) the way in which the relation between the first- and third-person perspectives is conceptualized in these traditions;
- (2) the philosophical meaning of the concept of ‘subjectivity’;
- (3) the methods advanced to study subjectivity;
- (4) the concept of ‘the unconscious’ in relation to the sub-personal/personal distinction;
- (5) which philosophical assumptions on subjectivity influence recent accounts of schizophrenic psychopathology.

In the following sections, we will offer a further discussion and contextualization of our conclusions. In the first part, the conclusion for each consecutive chapter will be discussed in relation to the specific research question treated in that chapter. Next, we will give a broader overview of the general themes which emerged from our thesis in terms of the philosophical trajectory we covered throughout the different chapters. Finally, based on the results of our thesis, we will suggest some new avenues for future conceptual and empirical work in the field of psychopathology studies.

### Discussion of our conclusions

In our *first* chapter – ‘Madness, subjectivity and the mirror stage: Merleau-Ponty and Lacan’-, we targeted the different ways Merleau-Ponty and Lacan handled the question of the relation between madness and subjectivity by contrasting their respective phenomenological and psychoanalytic readings of Wallon’s psychogenetic model of the mirror stage. We opened our analysis with the explicit agreement between Merleau-Ponty and Lacan on the need to oppose “scientific cartesianism” (Lacan, 1947/2006; Merleau-Ponty, 1985, p. 58) which reduces the phenomenon of madness to nothing but an ‘organistic’ or ‘physicist’ disturbance. Implicit in that agreement is the shared assumption that such accounts essentially miss out on the fundamental “human value” (Lacan, 1947/2006, p. 135) of madness which transcends the latter condition and reveals something about the metaphysics of subjectivity as

such. For both Merleau-Ponty and Lacan, this particular ‘metaphysics’ is marked by the interlacement of recognition and miscognition. However, not in the sense that madness is the error in relation to which normality represents the criterion of truth, but in the sense that “if a man who thinks he is a king is mad, a king who thinks he is a king is no less so” (Ibid., p. 139). This idea according to which ‘every recognition entails a fundamental miscognition’ led us to question how both thinkers consider the concepts of *identification* and *alienation* in relation to this dialectic, exemplified in our chapter by their respective readings of the mirror stage. Again, both Merleau-Ponty and Lacan basically agree that the specular mirror image functions as the primordial form of identification which is at the same time a misidentification, a jubilant and anticipatory recognition of something which is different from the subject but nonetheless forms the ground for any ulterior identity. Yet, the paths diverge as to the conclusions Merleau-Ponty and Lacan deduce from this observation, both with regard to the idea of subjectivity as such and for the particular case of madness. Whereas Merleau-Ponty regards the illusion of the specular image as something which indeed reveals the ambiguity of perception, of its pre-objective indetermination always ready to be left behind for the exactitude of the world and its naturalistic repression, it is nonetheless this pre-objective, ‘true’ experience which remains the real target of his phenomenological focus. In the same way, the blocked and static posture of the madman is only analyzed the better to discern the dynamic normality which is opposed to it. For Lacan, by contrast, the relation between illusion and truth, and in the same way, the relation between pathology and normality, is reversed. That is, when Lacan claims that it is “the illusion in itself that gives us the gestalt action that is psychology’s true object here” (Ibid., p. 146), he does not want to say that illusion *conceals* this action, but that the illusion *is* the action. For Lacan, it is the illusion qua image which counts, an illusion which will furthermore fulfil a constitutive role for perception itself, whether ‘illusory’ or ‘true’, ‘pathological’ or ‘normal’. More generally, Lacan reverses the relations between the natural and phenomenological attitudes. Whereas, for Merleau-Ponty, in order to reveal and return to our pre-objective ‘lived’ experience, one needs to surpass empiricist and intellectualistic interpretations, and finally even the objective outcome of perception itself, Lacan puts the experience *of* this illusion center stage. In short, for Lacan, lived experience is always already marked by objectification, and therefore, what we are inclined to call my innermost ‘self’ but the supreme synthetic form of this specular alienation.

The latter question about the relation between subjective experience and objectification, and furthermore, about the philosophical method deemed necessary to ‘return’ to subjectivity, was the focus of our *second* chapter – ‘How to return to subjectivity? Natorp, Husserl and Lacan on the limits of reflection’. We started out by shortly revisiting the contemporary debate on the supposed ‘explanatory gap’ separating neurocognitive accounts from first-personal, subjective experience. While basically agreeing with that assessment, we then ventured into the classical problem on how to account for this subjective dimension by discussing how this question was respectively treated in the works of Natorp,

Husserl and Lacan. We developed Natorp's neo-Kantian argument as to the limits of phenomenological reflection and its inability to fully disclose its subjective point of departure. Next, we contrasted Natorp's critique with Husserl's transcendental perspective and the methodological movements of 'epoché' and 'reduction' and concluded by stressing the incompatibility of the phenomenological appeal to 'givenness' and the need for a reflexive thematization of subjectivity. Finally, we offered a new perspective on this discussion between Natorp and Husserl by taking our cue from Lacan's distinction between 'the subject of enunciation' and the 'subject of the utterance'. Again, as in our previous chapter, we pointed out how Lacan stressed the necessity of alienation/objectification in any attempt to reflexively determine our own subjective identity, while being at the same time the necessary and retroactive condition through which we become confronted with our most immediate and elusive, subjective position. In our conclusion, we discussed the implications of Lacan's reformulation of the concept of subjectivity for the debate between phenomenology and neurocognitive science. On the one hand, we stressed that Lacan's conception, though seemingly paradoxically, can also be considered as a perfectly classical one and consistent with those philosophical views for which subjectivity was to be defined as an impossible, negative being that "is what it is not and is not what it is" (Sartre, 1958), and therefore closer to phenomenological readings which stress the "immanent alterity" at the very heart of the phenomenon of self-manifestation (Zahavi, 1999). On the other hand, we pointed out how the Lacanian conception is able to overturn the implicit agreement between neurocognitive and (some) phenomenological approaches according to which the epistemic shift towards 'subjective appearance' is a matter of epistemic certainty.

Our next, *third* chapter 'Cartesian meditations: Lacan and the truth of enunciation' served as a further elaboration of the distinction between 'enunciation' and 'utterance' we touched upon in our previous chapter, this time, however, in order to clarify the specific psychoanalytic notion of 'truth'. We contrasted this idea with correspondence theories according to which the supposed truth-value of a statement is to be evaluated in terms of its conformity with empirical data, be they subjective or objective. What matters in a psychoanalytic approach of subjective truth, on the other hand, is not so much its representational correspondence, i.e. whether or not what I say and claim conforms to what I am or experience, but how this very *act* of speaking is related to *what* I say. As in Freud's formal definition of an illusion (1927, pp. 30-31), I can very well speak the truth about myself on a constative level, while remaining oblivious to the function this speech-act fulfils on a performative level. Therefore, we concluded our analysis of the relation between representational knowledge and performative truth in psychoanalysis by presenting it as a matter of Bleulerian 'double bookkeeping': i.e. as a disjunction between what I confirm on the level of knowledge or belief and the way I nevertheless *act* seemingly against my better knowledge.

The distinct philosophical analyses of our previous chapters were employed in our *fourth* chapter ‘He or it (the thing) thinks, not I: self-monitoring and verbal hallucinations’ to offer an in-depth examination of recent, neurocognitive accounts of the phenomenon of ‘verbal hallucinations’ in schizophrenia. More in particular, we critically discussed Frith’s (1991) ‘self-monitoring model’ in terms of both its explanatory value with regard the ‘sense of ‘agency’ and ‘sense of ownership’ which is presumed to mark self-conscious thought, and in terms of its advanced description of hallucinatory experience. In our overall discussion, we raised two central questions. First, whether the extrapolation of the self-monitoring mechanisms from movement to thinking could be justified on the conditions set forth to explain ownership and agency in the case of movement. Here we argued that in the case of thinking, Frith’s model either leads to infinite regress, or is unable to explain ‘sense of agency’ or Campbell’s (1999) proposed function of thought coherency. Secondly, we raised the phenomenological question whether Frith’s model should offer an explanation of the ‘sense of agency’ in the experience of thinking in the first place. Moreover, this led to the further question whether the differential-diagnostic characterization of verbal hallucinations as a disruption of our ‘sense of agency’ while preserving a ‘sense of ownership’ offers an adequate description of hallucinatory experience. Both questions were answered negatively. On the one hand, we pointed out that agency is not an immediate ‘given’ in our experience of thinking. On the other hand, we argued that the distinction between ‘normal thinking’ and ‘verbal hallucinations’ cannot be made solely in terms of the presence or absence of agency. We ended our conclusion with a tentative proposal for future research on hallucinatory experience according to which the question of self-reference and alienation should be formulated in terms of the concept of *identification*, rather than the (in)correct description of a ‘sense of agency’. However, given the conceptual work done in the later chapters, we believe that we are now in a position to formulate this question in a philosophically more adequate and empirically more promising and fruitful way (cf. *infra*).

Our *fifth* chapter ‘The logic of appearance: Dennett, phenomenology and psychoanalysis’ was to a large extent a resumption of the themes discussed earlier in chapter two. Yet, here the specific question that figured throughout the chapter was in what way both phenomenology and psychoanalytic theory could respond to Dennett’s neurocognitive proposal to reduce ‘subjectivity’ and ‘the first-person perspective’ to a “theorist’s fiction” (Dennett, 1991, p. 79). In the first part, we offered a detailed reading of what is understood by such terms as ‘first-person experience’ and ‘subjective appearance’ in both phenomenological theory and Dennett’s account, and on the basis of that reading, pressed some arguments that could be advanced against Dennett’s fictionalization of subjectivity. We argued that Dennett’s heterophenomenological stance contradicts its professed doxastic, first-person operationalism and is therefore confronted with the following paradox: *either* to make room for the distinction between experience and judgment in which case the methodological plea for heterophenomenology can be saved but at the cost of first-person operationalism, *or* to retain the latter, but at the cost of advocating an

unavoidable return to the philosophical core of Cartesianism, whether or not without a ‘theater’. In the second part, we offered a similar exercise in the case of psychoanalysis. We argued that Sellars’ distinction, exemplified by the discussion between Dennett and phenomenology, touches upon a classic ambiguity in psychoanalytic theory with respect to the question whether it should best be construed as a sub-personal theory on the same explanatory level as neurocognitive science, or as a descriptive theory in which case its concepts have a primary application on the personal level. We discussed this ambiguity in both Freud’s and Lacan’s case, yet argued that psychoanalysis *rejects* the specific way in which that distinction figures in the discussion between Dennett and phenomenology. Here we identified the implicit, though decisive Cartesian agreement on the concept of ‘belief’ shared by both parties and contrasted this with a psychoanalytic account inspired by Mannoni (2003) and Pfaller (2014). On the basis of that account, we rejected the idea that ‘immunity to error’ immediately ensues whenever we shift the terrain to ‘seeming’ in terms of belief. Furthermore, we also rejected what could be considered as the defining gesture of Dennett’s heterophenomenology: i.e. the rejection of “the bizarre category of the objectively subjective” (Dennett, 1991, p. 132). According to Dennett, that obscure philosophical notion should be disposed of because it is once again a clear expression our tenacious captivity in the Cartesian theater. Yet, the examples we discussed made clear that this is not the only way in which that category could be understood. Indeed, rather than believing in ‘appearances’ while not knowing whether or not they correspond to reality, ‘objective beliefs’ as defined in our chapter display the exact opposite structure: we know they are not true, perhaps we even have a thoroughly adequate picture of reality, but the fact that we are nonetheless defined by them, that we act on them, remains undetected. Hence, we concluded that, from a psychoanalytic point of view, it is not so much the illusion of the ‘Cartesian theater’, but Dennett’s ‘belief in belief’ considered as an exhaustive account of subjectivity that needs to be questioned.

This question of what explains our first-person authority when it comes to telling what I believe, feel or intend to do, and conversely, how the limits to this authority should be properly conceived, was explicitly taken up in our ultimate, *sixth* chapter ‘Expression and the unconscious’. We started out by specifying two explanatory requirements which should be met by any convincing philosophical theory dealing with the issues of subjectivity and first-person avowals: (1) they should be able to preserve and explain the basic first/third-person asymmetry in the way psychological terms are ascribed to myself or others; (2) without succumbing to the Cartesian picture of ‘inner observation’ or ‘introspection’. Furthermore, we pointed out the relevance of a philosophical account of first-person authority for the discussions which have traditionally surrounded Freud’s – and in a different, though nonetheless equally decisive way, Lacan’s – understanding and justification of the concept of ‘unconscious’. Here the question was: if first-person authority is not the outcome of observational evidence or any similar third-personal procedure, then when is an intention, fear or belief rightly said to be unconscious? What is it

by virtue of which a state of mind should be characterized as unconscious when the reference to reflexive consciousness, inner sense and similar ideas is rejected?

In the first part of that chapter, we developed an *expressivist* account of first-person authority inspired by Wittgenstein's remarks on the philosophy of psychology (Wittgenstein, 2009). On that account, first-person authority with regard to the ascription of psychological verbs should not be conceived in terms of epistemic *detection* (Finkelstein, 2003), i.e. as a matter of finding out what I hope, feel or believe, but as the capacity to *express* or *manifest* my states of mind in self-ascriptions of them. Similarly, what explains the asymmetry between the first- and third-persons in the use of psychological language is not the result of any epistemic asymmetry, i.e. in terms of the superiority of self-knowledge and the uncertainty plaguing our knowledge of 'other minds', but is grounded in the opposition between present tense expressive uses and all other employments of mental terms. Finally, we pointed out that skeptical attacks on 'folk-psychology' and our ordinary practice of speaking our minds (e.g. Dennett, 1991) are premised on a (materialist) detectivist account of first-person authority and therefore should be rejected on precisely that ground.

In the second part, we used our expressivist account of first-person authority in order to offer a clarification of the meaning of 'unconscious subjectivity'. We argued that one possible way to do so consists in bringing out *why* the reference to 'unconscious' matters in psychological, adverbial descriptions, and whether it would make a considerable difference if the use of such qualifications would henceforth be denied. However, we contrasted this adverbial employment of 'unconscious' with its substantial transformation through which '*the* unconscious' comes to designate a part of the mind and seems to gain an explanatory value. We also discussed Lacan's critique of this idea and considered his alternative according to which the meaning of unconscious could be deduced from Descartes' hyperbolic procedure. Here we rejected two central ideas in Lacan's account: first, that speaking of the unconscious necessitates speaking of a "subject of the unconscious" (Lacan, 1981, p. 126) which is in some sense different from the concept of an empirical person; second, we rejected the idea that the meaning of unconscious could be conceived in terms of the paradoxes of self-reflection and first-personal reference. In our conclusion, we advanced a non-epistemic account of the adverbial meaning of 'unconscious' according to which the qualification 'unconscious' does not designate an epistemic lack or deficiency in self-knowledge, but refers to a specific *expressive* position which I cannot occupy.

### **General considerations: the question of subjectivity**

In this section, we want to give a more general overview of the conclusions presented above by situating them in the philosophical context from which our different questions emerged and in which our discussion should be understood. Although, to a certain extent, there is a considerable overlap and congruence between the themes discussed and the specific conclusions presented throughout our thesis,

there is nonetheless also an important shift to be noted in our current view on how to approach the question of subjectivity in the conceptually most adequate way.

In general, the topics we discussed from various angles and by means of different authors, whether this meant the conflicting readings of the mirror stage by Merleau-Ponty and Lacan, the methodological quarrel between Natorp and Husserl on the limits of phenomenological reflection, or Dennett's skeptical re-description of the 'self' as a "center of narrative gravity" (Dennett, 1992), all concern the discussion about the validity of the concept of subjectivity. In the same way, the most recent expression of this thematic in the discussion between neurocognitive science and phenomenology reflects the fact that, after some years of relative silence, the concept is once again center stage in the contemporary debate. In one way, given its central importance in both our ordinary lives and much of modern philosophy, this is perhaps not too surprising, yet in another way, the current explosion of discussions at the intersection of cognitive neuroscience and philosophy of mind of such topics as consciousness and self-consciousness, embodiment, action and agency, selfhood and person, asks for a more specific elucidation. That is, the latter observation might be surprising in light of the fact that, until recently, especially in the French tradition, the main effort of philosophy or psychoanalysis consisted in what was variably called the 'critique', 'subversion' or 'destruction' of the categories of 'subject' and 'subjectivity'. In the following closing sections, we will therefore offer some reflections on how that 'critique' should be understood in light of the conclusions and arguments we developed. Finally, we will indicate how these reflections on how best to approach subjectivity can inform future research on the nature of schizophrenic pathology.

#### **(a) The concept of 'subject' and its critique**

As mentioned above, for various reasons, it seems that, despite the substantial differences between the philosophical programs of Derrida, Foucault, Lacan or Deleuze, these different authors shared at least one common conviction: it was deemed necessary to put an end to the 'philosophy of the subject' or the 'metaphysics of subjectivity'. What was again essentially at stake here, was not so much a censure of persons as the denunciation of an illusion. That is to say, what was called the 'critique of the subject' was in fact a critique of the *concept* of subject and subjectivity, that is, upon the content of the concept, upon its possible applications, upon its validity. More specifically, according to this critique, it was to be called an 'illusion' – an illusion ascribable to the 'metaphysics of subjectivity' – to believe that a lover is the subject of his desires, a thinker the subject of his thoughts, a writer the subject of his writing, an agent the subject of his action, and so on.

As Balibar observed (2003), as such, the critique implied both a theoretical and practical aspect. As *theory of the subject*, it denounced the idea, explicit in phenomenological philosophy, that the concept of subject was to play a fundamental role in the explanation of a host of other concepts. For

example, the structuralist analyses of language, mythology or kinship, seemed to have no need for the idea of a ‘thinking subject’. In order to distinguish units of meaning within a system, one needs to start from a definition of this system as consisting in distinctive oppositions. It is thus not necessary to look for the origin of meaning in a consciousness or a ‘constituting subject’. Phenomenological philosophers of the existentialist or transcendental stripe, by contrast, would like to promote consciousness, subject and subjectivity as the origin and ground of meaning. It was this frontal opposition about the validity of the concept of subject with regard to the explanation of meaning and language which was best captured in the famous formula of Cavailles (1960, p. 78): “it is not a philosophy of consciousness but a philosophy of the concept that can yield a doctrine of science”<sup>91</sup>. As *ethics or politics of the subject*, on the other hand, it denounced the idea that the concept of subject would render possible an ethical and political thinking. For example, one of the issues discussed here was whether or not we need to be guided by the ideal of a ‘free subjectivity’ (cf. Sartre, 1958) in posing ethical questions, or whether we had to start from those ‘structures’ or ‘systems’ which determine a subject’s imaginary relation to its political surroundings (cf. Althusser, 1972).

Taken together, this typical *querelle* in French philosophy seemed to have the following content. For the critics, the philosophy of the subject was invalid in terms of its theoretical aspect, therefore vain or even disastrous in terms of its practical aspect. As for the defenders, they seemed to assert its validity in its practical aspect, hence justifying its use as a theoretical concept as – more or less – a regulating ideal. However, throughout this quarrel, it is important to note that critics and defenders at least seemed to comprehend the arguments and distinctions presented by their detractors. What was less certain, however, and what we have tried to remedy throughout our thesis, is whether the *philosophy of the subject* was itself properly understood. What do we affirm, for example, when we say that the speaker is the ‘subject’ of his speech? And what do we deny when we say he is not? Nothing immediate comes to mind, but in any case, it is clear that the analysis of the predicate ‘being a subject’ governed the entire discussion.

All too often, in the remarks exchanged during this quarrel, and to a certain extent also in the current discussions, no clear distinction was drawn between the ordinary use of the word subject and the properly philosophical use of this same word. As we made clear in our sixth chapter, in ordinary use, the notion of subject is bound up with the possibility of identifying someone in response to the question ‘who?’<sup>92</sup> For example, we want to know who comes speaking in our cabinet or who it is that we love.

---

<sup>91</sup> Although Cavailles’ formula is often invoked as the programmatic watchword of anti-humanist, structuralist thought, it is important to note that Cavailles himself mainly aimed at two opposing schools in the philosophy of mathematics. The question here was: should one define the concept of number starting from a description of consciousness or from a definition of the predicate of number? The answer should be obvious, but what would a ‘philosophy of the concept’ entail, for example, in the philosophy of language or social philosophy? Here the answer is all but clear.

<sup>92</sup> As Balibar (2003) observes, the possibility of finding a meaningful place for the question ‘who?’ played an important role in the shift between ‘structuralist’ thought *stricto sensu* and what has become known as ‘post-structuralism’. Since, to a large extent, the currents within the latter movement were political, there was indeed an



As long as we hold on to this notion of subject, the question whether the speaker is also ‘the subject’ of his speech or the lover ‘the subject’ of his love appears without any determinate meaning, unless we decide to introduce a new concept of ‘subject’ and ‘subjectivity’. Henceforth, as evidenced in our thesis, the question of the subject takes on a different meaning: it becomes conceivable, for example, that I, Jasper Feyaerts, am speaking without being the subject of my speech. Though it is still not particularly clear what that means, this gesture is easily recognizable as one of the defining moments of the different ‘philosophies of the subject’: it is the exposition of an argument destined to sustain the distinction between the empirical *person* (in the sense of a human being) and the *subject*.

On multiple occasions, most explicitly in our first, two chapters, we have assisted to clarifying that distinction as it is conceived in phenomenological philosophy and psychoanalytic theory, but it is perhaps most appropriate to return to the argument of the *cogito* in Descartes where the distinction is presented in its most distinctive form. In short, when Descartes set out in search for an absolutely certain truth, he arrives at: *Ego sum, ego existo*. At least this existence is certain, but the problems begin when we need to specify of what it is the existence. Who is it, that of whom it is true that he thinks and exists? Regardless of the answer one might prefer here, one point at least seems to be accepted within all philosophical endeavors which in one way or another return to this hyperbolic experiment to launch their different programs: the Cartesian argument requires that we do not understand the pronoun *I* (in the cases when it is used by Descartes) and the name *Descartes* as mutually substitutable terms. Therefore, we are not expected to say: Descartes is certain that Descartes exists, nor that Descartes is Descartes – we must say instead: Descartes is certain of a truth that he expresses to himself by saying “I am, I exist”, but beyond that, Descartes is not certain that there is a person named Descartes. Thus, the philosophy of the subject, in its Cartesian form, uses the method of *skeptical doubt* in order to establish a difference between the person and the subject. Other doctrines employ other methods and at the same time other determinations of the subject. For example, the method of *presupposition* poses a subject distinct from any empirical given as the condition of possibility of experience and science (e.g. Kant and Natorp). The method of *phenomenological experience* does not want the subject to be posed, it should be experienced, presented to itself in and through an experience, whether qua content or qua form (e.g. Husserl and Sartre). Psychoanalytic theory, finally, uses the method of *postulation* to advance the notion of a subject once again *different* from the idea of a person (viz. empirical ‘ego’), but this subject loses its foundational role and refers to that which remains as a negative category ‘internally excluded’ from its empirical determinations (e.g. Lacan and Žižek).

However, given this set-up, one might wonder what a critique of the philosophical subject should entail? Obviously, one could expect that it would bring about a counterargument, so as to be able

---

obvious need to rehabilitate a concept of ‘subject’ because the question ‘who?’ is bound up with the question ‘what should we do?’. That is to say, the latter question is not something which can be left indeterminate, but is always also a question of knowing *who* can do what should be done. Without a place for the ‘who?’, this terrain of practical and political philosophy will remain inaccessible to the philosopher.

to deny that it would be possible to differentiate between the person of Descartes and the thinking subject qua *ego cogito*. But manifestly, and as we have argued for example in the case of Lacan, the critics or subverters of the subject agree, at least on one point, with the defenders of the subject. This philosophy of the subject, through one or another of the methods advocated in the various schools, wishes to establish that Descartes is justified in saying “I am not Descartes”. But this is precisely what also bears out the critics of the subject: Descartes is *not* the subject of his thoughts. We must not say: *Descartes thinks*. We must say: *it thinks in Descartes*. Contrary to what we might expect, the sort of philosophical undertaking which passed itself off as a critique of the philosophy of the subject has not at all sought to restore the identity of Descartes’s person and of the subject to whom we should attribute Descartes’ thoughts. Quite to the contrary: it had sought to *accentuate* this difference. The critique of the subject was not the critique of the philosophical subject, but rather a protest against the tendency to confuse subjectivity (defined by methods of doubt, presupposition, experience or postulation) with a person’s mental life. Therefore, the critique of the subject has nothing to reproach the philosophy of the subject, unless it is that the difference between the subject in the ordinary sense and the subject in the philosophical sense has not been traced in a radical enough way. And this was indeed the recurrent complaint: the *ego* in Descartes resembles too much a human person, the transcendental self was too close to its empirical double, *Dasein* always risks being confused with a being whose mode of presence would be that of things, and so on.

Hence, whether one supports a doctrine of the subject or whether one attacks it, one thing remains constant: when the thought *Ego sum, ego existo* is expressed, it is in any case not Descartes who thinks this thought. Whether we speak of the ‘true subject’ in terms of consciousness, freedom or the unconscious is therefore, considered from this question of the concept of ‘subject’, a difference in terminology but not in thinking. The true subject will always turn out to be different from the category of a human person: Descartes is a man, the true thinking subject must therefore be inhuman; Descartes is a person, the true subject will be impersonal; Descartes is identifiable thanks to his body and history, the true subject escapes identification; Descartes can be described by means of its empirical properties, the true subject will be ineffable; ...

This is why it is safe to say that, in fact, the critique of the philosophical subject never really occurred in French philosophy. Yet, as we have argued in our final chapter, it has occurred, and with profound success, in the tradition we find in Wittgenstein and his successors. The question of the ‘subject’ is posed by Wittgenstein in the *Blue Book* (1969) and in the *Philosophical Investigations* (2009) in grammatical terms, as it is for the meaning of psychological terms in the first-person. This means that the question is one of knowing whether or not thinking and speaking in the first person – e.g. “I want to go on a holiday” – is, from a logical point of view, a singular proposition. In a singular proposition, one needs to find a singular term which permits identification of the referent (object) of which the predicate is true (to which it is applied). The prime example of such a singular term is the proper name: it has as function to designate in the sense of singularizing, selecting, singling out the object of which one wants

to speak. And the question is here: does the word 'I' satisfy this function? For example, in Descartes' famous formula of the 'I think', what is the function of the pronoun? Does the pronoun, in these phrases, introduce a reference to a subject of whom all we know for certain at this point is that it is thinking? Wittgenstein discards this idea. When Descartes, in his soliloquy, declares 'I think' or even 'I am', the predicative proposition that he expresses does not bear upon a referent that would be some ghostly entity. The word 'I', pronounced by Descartes, does not function to help Descartes to know of whom the predicates *think, am, exist* are true. The word 'I' simply does not have here *any* referential function. It is not a special 'first-personal' case of speaking in the third person, of making reference or singling out someone by means of a definitive description (however minimal or lacking). Indeed, the word 'I' does not allow for an identification of the person who speaks: only when we pass to the third-person, by quoting someone's name or empirical identity, can the question of identification and reference be posed.

The relevance of these remarks on the logical function of the 'I' developed extensively by Wittgenstein and Anscombe (1981) is immediately clear if we remember that the discussion on the 'subject' essentially turned around the possibility of finding an acceptable meaning for the proposition: 'I am not René Descartes'. Of course, the latter has a meaning, but it is fairly trivial: it simply means that I am *another* person than Descartes. But does it still have a meaning if we say: I am *other* than the empirical person called Descartes? Here we should ask: another *what*? Again, not another person, since this would bring us back to its trivial meaning. And in fact, not another *anything*, since the possibility of an empirical determination of the 'subject' was already rejected in the very way the question was formulated. Whether one invokes a transcendental consciousness, an existential nothingness or a subject of the unconscious to mark the distinction from the empirical person Descartes' here is strictly irrelevant: neither of those allows one to distinguish one transcendental consciousness, *res cogitans* or subject of the unconscious from another, since we have no idea of what constitutes its identity. But this also means that the question 'who?' loses all meaning since all point of reference are entirely lacking. We know how to distinguish an ordinary subject from another ordinary subject (e.g. by his name, physical appearance or tone of voice), but the effort to identify the philosophical 'subject' can only lead to failure. Hence, all philosophies of the subject are confronted with the same paradox: if the Cartesian cogito (whether conscious or unconscious, purified or impersonal) is read in a way which allows for the question of 'who?', it is assimilated to an individual susceptible of being designated, identified, distinguishes, and so on. But if we do not want it to be identified in such an 'ontic' way, the very question of 'who' can no longer be answered.

### **(b) Psychological verbs and the question of self-reference in schizophrenia**

In our ultimate chapter, we already made clear that rejecting the philosophical problematic of the 'subject' as previously defined, poses no threat for psychoanalytic theory insofar as one of its basic tasks consists in offering a clarification of the concept of 'unconsciousness'. On the contrary, we argued that

abandoning that Cartesian affiliation only amounts to abandoning a distorted view on the first-person and, by implication, on the meaning of unconsciousness, which keeps producing nothing but conceptual incongruities. Yet, here we want to make clear how the philosophy of the cogito also keeps informing recent phenomenological accounts of the problem of self-reference in schizophrenia, and more in particular, in how the issue of misattribution of action and agency in schizophrenic symptoms such as thought-insertion and delusions of control is currently conceived (see also our chapter 4). We'll begin by sorting out some of our agreements and disagreements with how the topic is approached in the recent research literature and end by formulating some concrete ideas for future research.

On a more general level, it is an entirely commendable development in recent psychiatric research that the intimate and complex relation of psychopathology and philosophy is at the forefront of many discussions (e.g. Gallagher, 2007; Parnas, Sass & Zahavi, 2008). It is once again acknowledged, after some years of oversimplification and an exclusive focus on operationalist and empiricist methodology (see Andreasen, 2007), that researchers and mental-health professionals dealing with such terms as 'reality', 'psychosis' or 'subjectivity' are forced to engage, inadvertently or not, in philosophical thinking. Consequently, they have the choice between an unacknowledged and distorting philosophy and a philosophy which critically discloses and discusses its own presuppositions. That is, a psychopathological description involves converting the patient's experience, expressions and/or behavior in the first-person, into specific categories of symptoms and signs that are defined in third-person terms, thus providing objective, sharable information for diagnosis, treatment and research. Obviously, what is at stake here is the nature of the 'mental domain' and 'subjectivity' and the questions of how adequately to address and describe it. This is why, in recent years, there has been an explosive surge of publications at the intersection psychiatry and philosophy, the last with both analytical and phenomenological orientations.

One specific field where this cross-breeding is most thriving is in recent accounts of the nature of schizophrenic pathology. Drawing mainly on phenomenological philosophy, it has become standard to describe schizophrenic symptoms of delusions of control and thought insertion on a first-personal level in terms of problems with the 'sense of agency' and 'sense of ownership' (e.g. Bortolotti, 2010; Langland-Hassan, 2008; Roessler, 2013). Both 'sense of agency' and 'sense of ownership' are meant to refer to the pre-reflective (or first-order) *feeling* or *experience* of agency and ownership. In the context of discussions about schizophrenic delusions of control and symptoms of thought insertion, these phenomenological distinctions are employed to capture problems in the attribution of agency and action in patients' reports. For example, a patient may report:

They inserted a computer in my brain. It makes me turn to the left or right. It's just as if I were being steered around, by whom or what I don't know. (cited in Frith et al., 2000, p. 358)

This problem in the self-ascription of action and movement is thought to result from a distorted *experience* of agency, in this case, the experience as if something (or someone) else is controlling the patient's movement (Synofzik et al., 2008; Gallagher, 2014). More generally, on a phenomenological account, self-ascription of agency and talking in the first-person about our intentions and actions is a matter of describing/reporting our pre-reflective, first-order experience of agency. For example, if I walk over to open the door, and I attribute this action to myself, the ground for my self-attribution depends on experiencing myself initiating the movement, that is, my movement must be accompanied by an implicit 'sense of agency' for this movement; otherwise, I would end up ascribing this action to someone or something else.

Leaving this specific case of how to account for self-ascriptions of agency in ordinary cases and schizophrenic pathology aside for a moment, it is important to note in what sense these phenomenological accounts perpetuate a classical, Cartesian framework. In the same way as a Cartesian account of self-reference necessitates advancing an 'experiential access' to an I/self (through introspection, inner sense or experience) leading to all the sorts of undesirable consequences we noted in our previous section, it is characteristic for this tradition to do the same for the class of 'psychological verbs' (thinking, acting, feeling, ...). As Ricoeur (1950, p.9) remarked, Descartes extended the meaning of the latin verb *cogitare* in order to include *all* these verbs: 'feeling is thinking', 'willing is thinking', 'intending is thinking', and finally, even 'acting is thinking'. More in particular, 'thinking' or 'cogitare' in this extended sense refers more to something like 'feeling' or 'having the experience of'. That is, my consciousness cannot tell me whether I really see what I believe to see or whether I'm really walking or only have the impression of doing so, but nevertheless, it can at least guarantee me that I have the *experience* of seeing/walking.

In our final article, we have discussed Wittgenstein's objection against this decision to include all psychological verbs in the same category with the generic term 'experience'. Wittgenstein's objection is not that these verbs which are part of the Cartesian modes of 'cogitare' cannot be put in the same class. Indeed, they can, but only because of the asymmetry they manifest between the first- and third-persons in the present indicative. In the third-person, the use of the verb rests on the observation of the person who one describes. In the first-person, however, the subject who says 'I'm thinking about what to make for dinner this evening' or 'I want to go to Italy for vacation' doesn't need to observe himself in order to say what he is thinking or what his current intention is. However, from the fact that all these verbs have this asymmetric property, it does not follow that they are therefore all *experience* verbs of the sort 'I see a color', 'I'm hearing a noise'. Characteristic for 'experience verbs' is the fact that I can describe their content or *Erlebnisinhalt* (the perceived of my perception, the seen of my vision, ...). Yet, someone who has the intention to do something can experience all sorts of sentiments relative to what he is planning to do (e.g. excitement, restlessness, ...), but *nothing* of what he experiences constitutes his intention to carry out his project. In short: intention and action verbs and their related auxiliaries as "willing" and "intending" are precisely not experience verbs; therefore,

speaking in the first-person about my voluntary actions and movements is not a matter of describing ‘voluntary experiences’ or, indeed, a ‘sense of agency’. Being asked to describe the experience of my action, e.g. whether or not it *felt* voluntary, is precisely to neglect the differences between the language games of *action* and *experience*.

The implications of Wittgenstein’s critique of the Cartesian tendency to reduce everything we do in the first-person to the generic ‘conscious experience’ for current and future research on schizophrenic symptoms like delusions of control, verbal hallucinations and thought insertion should be clear: the difference between what is voluntary or involuntary, whether an action, movement or thought is active or passive, cannot be made in terms of the distinction between the presence or absence of a ‘sense of agency’. Wittgenstein contends that what characterizes one’s action as voluntary is not an ‘experience of voluntariness’, but a particular surrounding of intention, motivation and behavior (1988, p. 213). Similarly, in order to determine what patients mean when they describe a thought as ‘inserted’ in one’s mind or an action performed against one’s will, it is insufficient to approach these complex questions by simply introducing spurious, phenomenological variables as a ‘feeling of insertion’ or a ‘sense of passivity’. Here too, a more sensitive approach takes into account what patients describe as the circumstances, environment and surroundings of their verbal hallucinations or delusions of control.

Finally, we believe that these introductory remarks on Wittgenstein’s philosophy of action, and more specifically, the perspective it opens for future, qualitative research on the nature of schizophrenic psychopathology, could profit from the work that is done on these issues from a Lacanian perspective (see, e.g., Vanheule, 2011). Like Wittgenstein, Lacan moved away from exclusive, phenomenological accounts of agency and perception as a framework to approach the clinic of psychosis and proposed a model in which the relation of intersubjectivity and self-experience is central for understanding what characterizes, for example, the sudden emergence of ego-dystonic hallucinations. Rather than reducing the nature of hallucinatory intentionality to formal disruptions in the experience of consciousness or our sense of control, Lacan (1959/2006) recommended to take into account the different faces of the ‘Other’ with which patients become confronted. Here again, and like Wittgenstein, Lacan proposes to include the particular surroundings and context of the patient in order to locate and understand the complaints of destabilization, nonsensicality and suspension so characteristic for psychosis. Given the current resurgence of the combination of philosophical approaches of subjectivity and detailed, qualitative research of first-person accounts, the general research program we presently discussed is in our view one of the most promising to follow in the years to come.

## References

- Althusser, L. (1972). *Lénine et la philosophie*. Paris: Maspero.
- Andreasen, N.C. (2007). DSM and the death of phenomenology in America: An example of unintended consequences. *Schizophrenia Bulletin* 33, 108-112.
- Anscombe, G.E.M. (1981). "The First Person", in *Metaphysics and the Philosophy of Mind*. Oxford: Basil Blackwell, 21-36.
- Balibar, E. (2003). Structuralism: A Destitution of the Subject? *Differences: A Journal of Feminist Cultural Studies* 14 (1), 1-21.
- Bortolotti, L. (2010). *Delusions and other irrational beliefs*. Oxford: Oxford University Press.
- Campbell, J. (1999). Schizophrenia, The Space of Reasons and Thinking as a Motor Process. *The Monist* 82 (4), 609-625.
- Cavaillès, J. (1960). *Sur la logique et la théorie des sciences*. Paris : Presses universitaires de France.
- Dennett, D. (1991). *Consciousness Explained*. New York: Back Bay Books/Little, Brown and Company.
- Dennett, D. (1992). "The self as a center of narrative gravity", in F.S. Kessel, P.M. Cole & D.L. Johnson (Eds.), *Self and Consciousness: Multiple Perspectives*. Hillsdale, NJ: Erlbaum, 4-237.
- Finkelstein, D. (2003). *Expression and the Inner*. Cambridge, Massachusetts: Harvard University Press.
- Freud, S. (1927). The Future of an Illusion. *Standard Edition* 21, 5-56.
- Frith, C.D. (1991). *The Cognitive Neuropsychology of Schizophrenia*. Hillsdale: Erlbaum.
- Frith, C.D., Blakemore, S.J. & Wolpert, D.M. (2000). Abnormalities in the awareness and control of action. *Philosophical Transactions of the Royal Society of Biological Sciences* 355, 1771-1788.
- Gallagher, S. (2007). The natural philosophy of agency. *Philosophy Compass* 2 (2), 347-357.
- Lacan, J. (1981). *The Seminar 1964, Book XI, The Four Fundamental Concepts of Psychoanalysis*. New York & London: W.W. Norton & Company.
- Lacan, J. (2006 [1947]). "Presentation on Psychical Causality", in J. Lacan & J.-A. Miller (Eds.) *Ecrits*. New York and London: W.W. Norton, 123-160.
- Lacan, J. (2006 [1959]). "On a Question Prior to any Possible Treatment of Psychosis", in J. Lacan & J.-A. Miller (Eds.) *Ecrits*. New York and London: W.W. Norton, 445-488.
- Langland-Hassan, P. (2008). Fractured phenomenologies: thought insertion, inner speech, and the puzzle of extraneity. *Mind & Language* 23 (4), 369-401.
- Mannoni, O. (2003). "I Know well, But All the Same ...", in M.A. Rothenberg, D.A. Forster & S. Zizek (Eds.), *Perversion and the Social Relation*. Durham, NC: Duke University Press, 68-92.
- Merleau-Ponty, M. (1985). *L'oeil et l'esprit*. Paris: Gallimard.
- Parnas, J., Sass, L.A. & Zahavi, D. (2008). Recent developments in philosophy of psychopathology. *Current Opinion in Psychiatry* 21 (6), 578-584.
- Pfaller, R. (2014). *On the Pleasure Principle in Culture: Illusions Without Owners*. London, New York: Verso.
- Ricoeur, P. (1950). *Philosophie de la volonté, t. I, Le volontaire et l'involontaire*. Paris: Aubier.
- Roessler, J. (2013). "Thought insertion, self-awareness, and rationality", in K.W.M. Fulford et al. (Eds.) *The Oxford handbook of philosophy and psychiatry*. Oxford: Oxford University Press, 658-672.
- Sartre, J.P. (1958). *Being and Nothingness: An Essay on Phenomenological Ontology*. London: Methuen.
- Synofzik, M., Vosgerau, G. & Newen, A. (2008). Beyond the comparator model: a multifactorial two-step account of agency. *Consciousness and Cognition* 17, 219-239.
- Vanheule, S. (2011). *The Subject of Psychosis: A Lacanian Perspective*. Hampshire: Palgrave Macmillan.
- Wittgenstein, L. (1969). *The Blue and Brown Books*. New York: Barnes and Noble.
- Wittgenstein, L. (1988). *Wittgenstein's Lectures on the Philosophy of Psychology 1946-7*, ed. P.T. Geach. Chicago: UP.
- Wittgenstein, L. (2009). *Philosophical Investigations* (4<sup>th</sup> Edition). Oxford: Wiley-Blackwell.
- Zahavi, D. (1999). *Self-Awareness and Alterity: A Phenomenological Investigation*. Evanston: Northwestern University Press.





## Subjectiviteit in de psychoanalyse, fenomenologie en cognitieve neurowetenschap: een conceptuele studie

---

### Probleemstelling

In recente jaren zijn we getuige geweest van een terugkeer van het vraagstuk naar subjectiviteit in diverse onderzoeksdomeinen, gaande van de neuro-cognitieve wetenschappen, de analytische ‘philosophy of mind’, de studie van psychopathologie tot de zogenaamde ‘consciousness studies’. Een van de belangrijkste redenen voor een dergelijke terugkeer ontspringt aan wat we freudiaans zouden kunnen omschrijven als een ‘Unbehagen in die kognitive Neurowissenschaft’: i.e., een groeiend onbehagen met wat algemeen beschouwd wordt als het dominante paradigma in het wetenschappelijk onderzoek naar de menselijke geest sinds de jaren ‘60, de cognitieve (neuro)wetenschap. Hoewel zij initieel in staat werd geacht om Hume’s befaamde ambitie te verwezenlijken de ‘Newton of mental phenomena’ te worden, stellen onderzoekers vandaag dat er nog steeds fundamentele vragen onbeantwoord zijn gebleven. Het belangrijkste vraagstuk is deze naar de relatie tussen cognitieve beschrijvingen van het mentaal functioneren, zoals deze typisch wordt gedacht als een of andere vorm van ‘informatieverwerking’ in onze black-box, en subjectiviteit, losjes en wat onbeholpen gedefinieerd in de literatuur aan de hand van Thomas Nagels befaamde uitdrukking “what it is like” om een cognitieve geest te zijn, dewelke ruwweg verwijst naar het fenomenologische begrip van subjectiviteit als hoe de dingen ‘lijken’, ‘schijnen’ of ‘ervaren’ worden vanuit het subjectief of eerste-persoonsperspectief. Er wordt meer bepaald beargumenteerd dat de cognitieve wetenschap geconfronteerd zou worden met een ‘verklaringskloof’ ten aanzien van deze subjectiviteit, dat ze, niettegenstaande de groeiende complexiteit in haar beschrijvingen van ons cognitief apparaat zoals deze functioneert ‘in zichzelf’, er desondanks niet in slaagt in rekening te brengen hoe dit cognitief apparaat op haar beurt verschijnt ‘aan zichzelf’, hoe het cognitief functioneren subjectief wordt ervaren. Of om het met de treffende woorden van Roy et al. (2000, p. 12) te stellen: “explaining what is happening in the black box is not explaining what is happening for the black box”. Volgens de pleitbezorgers van dit verklaringskloof-argument is het dus één ding om hypotheses te ontwikkelen over wat er zoal omgaat in ons brein – op eender welk niveau van verklaring, hetzij neurobiologisch of functioneel beschouwd – voor het begrijpen van diverse vormen van cognitief gedrag (bv. aandacht, perceptie, geheugen, ...),

het is iets fundamenteels anders wanneer we bovendien trachten in rekening te brengen wat er *lijkt* te gebeuren bij dergelijke vormen van cognitief gedrag vanuit het eerste-persoonsperspectief.

Het is deze opmerkelijke omkering in een soort verbasterde lezing van een kantiaans wetenschapsideaal — waarbij de cognitieve wetenschap er wel in slaagt om het cognitief/noumenale ‘Ding an sich’ in kaart te brengen, maar in gebreke blijft ten aanzien van haar fenomenale verschijning — die onderzoekers er dus toe gebracht heeft te stellen dat een louter functionele analyse van ons cognitief functioneren geen exhaustieve en adequate verklaring kan bieden zolang de eerste-persoonsdimensie qua subjectieve ervaring niet opgenomen wordt als een expliciet domein van (neuro)cognitief onderzoek. In het licht van deze verklaringskloof en voortbouwend op het daaraan verbonden fenomenologisch argument, roepen diverse auteurs dan ook op om terug te keren naar klassieke fenomenologische benaderingen van subjectiviteit in een poging om de relatie tussen cognitieve processen en fenomenale manifestaties te verhelderen (Schmicking & Gallagher, 2010). Nog volgens deze auteurs zou de fenomenologie, zoals deze haar oorsprong kende in het werk van Husserl en later verder opgenomen werd door fenomenologen zoals Heidegger, Merleau-Ponty en Sartre, de cognitieve wetenschap kunnen voorzien van een verfijnde descriptieve benadering van deze subjectiviteit als een noodzakelijke en propedeutische voorwaarde voorafgaand aan elke naturalistische poging deze kloof te dichten.

Echter, niet alle auteurs delen ditzelfde vertrouwen met betrekking tot wat een ware “fenomenologische renaissance” genoemd wordt (Parnas, Sass & Zahavi, 2008, p. 578) en haar mogelijkheid om daadwerkelijk bij te dragen aan het klassieke project het mentale te naturaliseren. Deze kritiek werd geuit langs twee verschillende lijnen. Enerzijds wordt er gesteld dat hoewel een naturalistisch perspectief op subjectiviteit mogelijk is, een fenomenologische benadering in termen van een geprivilegieerd eerste-persoonsperspectief en fenomenale ervaring dit niet is. Met name Daniel Dennett heeft het idee verworpen dat subjectiviteit en bewustzijn ‘data’ zouden impliceren die enkel toegankelijk zijn vanuit een eerste-persoonsperspectief en, meer nog, dat er in feite niet zoiets is als een ‘werkelijke fenomenologie’ (Dennett, 1991) of de mogelijkheid van een wetenschap in de eerste-persoon. Een andere lijn van kritiek stelt dat hoewel een wetenschappelijke benadering van subjectiviteit inderdaad mogelijk is, dit niet kan gebeuren vanuit een naturalistisch wetenschapsbegrip. Deze kritiek is geworteld in de transcendentale aard van de Husserliaanse fenomenologie. Husserl’s punt is dat er steeds een dimensie zal zijn die zowel vooronderstelt als ontoegankelijk is voor de natuurwetenschappen en dewelke een specifieke fenomenologische benadering vereist (Husserl, 1900-1901/2002, p. 254).

Een intellectuele traditie die verassend genoeg bijna volledig afwezig is in deze discussies is de Freudiaanse en Lacaniaanse psychoanalyse. Dit is des te opmerkelijker aangezien de spanning tussen de claims van de manifeste en wetenschappelijke beelden (Sellars, 1963), en meer in het bijzonder, de meest recente manifestatie van deze ongemakkelijke verhouding tussen de eerste- en derde-persoonsperspectieven, de eigenlijke kern raakt van wat er in de psychoanalyse op het spel staat. Gegeven de specifieke distributie van het onderscheid tussen neurocognitieve realiteit en manifeste

verschijning, zijn er inderdaad geen waardevolle lessen meer te leren van Freuds originele inzicht dat, zelfs in het veld van verschijning, “het psychische in werkelijkheid niet zo hoeft te zijn als het ons toeschijnt” (Freud, 1915, p. 69)? Of wat te denken van Lacans paradoxale claim dat het Cartesiaanse cogito niet zozeer een fictie is die achtergelaten zal worden doorheen de wetenschappelijke vooruitgang (Dennett, 1991), maar zelf de uitgelezen figuur is waaruit de betekenis van ‘het onbewuste’ kan afgeleid worden? Echter, noch dient deze opvallende stilte weggezet te worden als een triviale observatie: zoals Ricoeur (1997) en Laplanche (1999) opmerken, van bij haar oorsprong is de psychoanalyse gegrond in de systematische verwerping van de disjunctie tussen de sub-personele en persoonlijke niveaus van verklaring. Los van het feit of dit gezien moet worden als haar “onschatbare bijdrage” (Ricoeur, 1997, p. 66 – onze vertaling) of eerder als een “abominabele verwarring” (Wittgenstein, 1978, p. 313 – onze vertaling), deze tweespalt verklaart ten dele de opmerkelijke psychoanalytische timiditeit ten aanzien van de filosofische uitdagingen die oprijzen vanuit het actuele debat.

### **Onderzoeksvragen**

Samengenomen reveleert de algemene discussie die we net schetsten met betrekking tot de zogenaamde ‘verklaringskloof’ tussen de cognitieve wetenschappen en de fenomenologie, en voorts, de ambigue positie van de psychoanalyse ten aanzien van deze filosofische situatie, de nood aan

- (1) een filosofische verheldering van hoe en op welke grond de rivaliserende claims van neurocognitieve en fenomenologische theorieën omtrent subjectiviteit beoordeeld dient te worden;
- (2) de explicitering van de filosofische assumpties die de verschillende tradities sturen in de respectievelijke benaderingen van subjectiviteit en de vraag van haar naturalistische verklaring en/of eliminatie;
- (3) een verheldering van het precaire statuut van psychoanalytische claims met betrekking tot het wezen van subjectiviteit via een vergelijking met de hedendaagse fenomenologie en neurocognitieve wetenschap.

Deze algemene problemen leiden vervolgens tot de volgende, meer specifieke onderzoeksvragen:

- (1) Op welke manier wordt de verhouding tussen de eerste- en derde-persoonspectieven geconceptualiseerd in de neurocognitieve, fenomenologische en psychoanalytische tradities? Is deze verhouding een kwestie van epistemische asymmetrie, en indien zo, op welke wijze wordt

de kwestie van eerste-persoonskennis gedacht met betrekking tot de vraag naar epistemische verantwoording?

- (2) Wat wordt er precies bedoeld in elke traditie met populaire termen als ‘subjectiviteit’, ‘manifeste verschijning’, ‘lijken/schijnen’ en de ‘eerste-persoon’? Bijvoorbeeld, vooronderstelt ‘subjectiviteit’ de introductie van het concept ‘subject’? Indien zo, wat zijn de verschillen/overeenkomsten tussen de verschillende disciplines in hoe dit ‘subject’ wordt begrepen en op welke manier verschilt het van het concept van de ‘empirische persoon’?
- (3) Wanneer subjectiviteit erkend wordt als een legitiem en noodzakelijk veld van onderzoek, welke methoden worden er dan geschikt geacht om deze te onderzoeken en de subjectieve dimensie toegankelijk te maken? Hoe zijn deze methoden op hun beurt verbonden met de wijze waarop subjectiviteit gedacht wordt in de verschillende tradities?
- (4) Is het verschil tussen eerste-persoons beschrijvingen van ervaring en derde-persoons verklaringen in neurocognitieve benaderingen, reduceerbaar tot het verschil tussen ‘bewustzijn’ en ‘onbewustzijn’? Specifiek voor de psychoanalyse: hoe verhoudt haar idee van ‘het onbewuste’ zich tot sub-personele neurocognitieve theorieën? Laat het psychoanalytische onbewuste de autoriteit van de eerste-persoon intact of vormt ze daarvoor een uitdaging?
- (5) Wat is de relevantie van deze discussies voor concrete voorbeelden uit de onderzoeksliteratuur over psychopathologie? Meer in het bijzonder, welke filosofische assumpties over subjectiviteit beïnvloeden de wijze waarop schizofrene pathologie wordt benaderd?

## **Methode**

Aangezien deze vragen allen filosofisch en conceptueel van aard zijn, worden zij behandeld en uitgewerkt via een reeks van conceptuele onderzoeksartikels. Deze artikels zijn papers die reeds gepubliceerd zijn in wetenschappelijke tijdschriften of als hoofdstukken in academische collecties, aanvaard werden voor publicatie of recent ingediend. Doorheen deze artikels zal er bijzondere aandacht worden besteed aan de conceptuele studie van zowel klassieke als meer recente filosofische auteurs in de respectievelijke tradities die invloedrijke ideeën hebben ontwikkeld over hoe subjectiviteit dient te worden benaderd. Voor de fenomenologische discipline omvatten deze auteurs Husserl, Merleau-Ponty en Sartre; voor de neurocognitieve traditie en analytische filosofie wordt aandacht besteed aan Dennett, Frith en Wittgenstein; voor de psychoanalyse zullen Freud en Lacan onze focus uitmaken. In elk artikel wordt er gepoogd om elke perspectief op zichzelf zo gedetailleerd mogelijk uit te werken, terwijl de verschillende claims die oprijzen uit deze lezing slechts vervolgens in wederzijds perspectief worden

geplaatst om de relatieve sterktes en zwaktes te bespreken in de uiteindelijke conclusies. Waar mogelijk en relevant zullen de conceptuele analyses besproken en geïllustreerd worden met empirische voorbeelden uit de onderzoeksliteratuur over psychopathologie en de implicaties voor toekomstig onderzoek duidelijk worden gemaakt. Hoewel onze algemene onderzoeksvragen het achterliggende kader vormt voor elk artikel, zal er genoeg ruimte worden voorzien om de specifieke details en vragen te behandelen die, hoewel niet expliciet vooropgesteld in onze introductie, niettemin kunnen opduiken doorheen onze discussie van verschillende auteurs en onderwerpen. In dat geval zal de relevantie en de implicaties van deze kwesties voor onze algemene onderzoeksvragen duidelijk worden gemaakt en bediscussieerd.

## Opzet

In het *eerste* hoofdstuk van ons proefschrift – ‘Madness, subjectivity and the mirror-stage: Merleau-Ponty and Lacan’ – wordt de intieme relatie tussen subjectiviteit en waanzin besproken zoals deze verschijnt in Lacans vroege geschriften en Merleau-Ponty’s klassieke werk *Fenomenologie van de waarneming* (2002). Centraal in deze paper staat het verschil tussen Lacans en Merleau-Ponty’s respectievelijke lezingen van Wallons psycho-genetische model van het ‘spiegelstadium’. Beide auteurs waren geïnteresseerd in dit model omwille van het besloten potentieel om de eigen theoretische posities aan te scherpen: waar het spiegelbeeld voor Merleau-Ponty een objectiverende illusie vormt dewelke kan en moet gereduceerd worden tot haar pre-objectieve grond in onze geleefde ervaring, beschouwt Lacan het speculaire beeld als onreducerbaar en constituerend voor subjectiviteit als dusdanig. We bespreken vervolgens hoe het verschil tussen Lacans en Merleau-Ponty’s ideeën aangaande de relatie tussen waanzin en subjectiviteit verbonden is met de functie die ze toeschrijven aan *miskenning* in de formatie van subjectiviteit, als ook de mogelijkheid (Merleau-Ponty) of onmogelijkheid (Lacan) om deze miskenning te overstijgen.

Beide concepten van miskenning en objectivering in verhouding tot subjectiviteit worden verder uitgewerkt in ons *tweede* hoofdstuk – ‘How to return to subjectivity? Natorp, Husserl and Lacan on the limits of reflection’ -, waarin de hedendaagse discussie omtrent de ‘verklaringskloof’ gebruikt wordt om deze concepten te verhelderen. In het eerste deel van dit hoofdstuk geven we een meer gedetailleerde discussie van de fenomenologische argumenten die wijzen op de vooronderstelde beperkingen van neurocognitieve verklaringen ten aanzien van subjectiviteit. In de volgende delen wordt de methodologische vraag behandeld aangaande de mogelijkheid om subjectiviteit toegankelijk te maken via fenomenologische reflectie. Deze vraag wordt uitgewerkt door middel van een lezing van de neokantiaanse filosoof Paul Natorp, zijn discussie en onenigheid met de grondlegger van de transcendentale fenomenologie Husserl, en besloten met een hegeliaanse synthese van deze posities via Lacans conceptuele koppel van uitspreken/uitspraak. In het laatste deel worden de implicaties en

relevantie van deze discussie voor het verklaringkloof-argument uitgewerkt en de kritische positie van de psychoanalyse ten aanzien van dat argument verduidelijkt.

Lacan's conceptuele onderscheid tussen het 'subject van het uitspreken' en het 'subject van de uitspraak' wordt verder opgenomen in onze *derde* hoofdstuk – 'Cartesian meditations: Lacan and the truth of enunciation' -, waar we deze gebruiken om een lezing te bieden van Freuds ambigue therapeutisch advies om in analyse "meer te zeggen dan men weet" (Freud, 2006 [1926e], p. 281). Vertrekkende van deze opmerking wordt een specifieke psychoanalytische versie van de cartesiaanse en husserliaanse procedures van *epoché* ontwikkeld om de rol van 'waarheid' in de psychoanalyse te verduidelijken. Overeenkomsten en verschillen ten aanzien van de fenomenologische notie van waarheid en de sartreanse figuur van 'bewuste negativiteit' worden uitgewerkt.

In het *vierde* hoofdstuk – 'He or it (the thing) thinks, not I: self-monitoring and verbal hallucinations' – worden de resultaten van onze filosofische analyses gebuikt om een kritische lezing te bieden van recente, neurocognitieve verklaringen van verbale hallucinaties in de schizofrenie. Meer in het bijzonder wordt Frith's predictieve, 'self-monitoring' model bevestigd in termen van enerzijds, het algemene idee omtrent subjectiviteit gegrond in het onderscheid tussen 'agency' en 'ownership' dat vooronderstelt wordt door het model, en anderzijds, de specifieke beschrijving die gegeven wordt van de fenomenologische aard van hallucinatoire ervaring. Op basis van deze kritische discussie schetsen we in het laatste deel enkele voorwaarden waaraan een neurocognitieve verklaring van verbale hallucinaties moet voldoen.

In het *vijfde* hoofdstuk – 'The logic of appearance: Dennett, phenomenology and psychoanalysis' – vervolgen we onze discussie uit hoofdstuk twee met betrekking tot de rivaliserende claims van neurocognitieve en fenomenologische theorieën omtrent subjectiviteit. Deze discussie wordt gekaderd aan de hand van Sellars' befaamde onderscheid tussen de manifeste en wetenschappelijke beelden over de mens. Vertrekkende van dit onderscheid trachten we drie verschillende manieren te ontwikkelen om te antwoorden op Sellars' diagnose: Dennetts filosofische reconstructie van de neurocognitieve wetenschap, hedendaagse fenomenologische modellen en de psychoanalyse. We suggereren dat de respectievelijke tradities en de substantiële verschillen tussen deze tradities begrepen kunnen worden aan de hand van een 'logica van verschijning'. Hieraan verbonden zijn verschillende ideeën omtrent de rechten en limieten van het eerste-persoonsperspectief, de relatie tussen bewuste ervaring en geloof, en de kwestie van naturalisering. In het laatste deel van ons hoofdstuk trachten we duidelijk te maken, op basis van een gedetailleerde lezing van de discussie tussen Dennett en de fenomenologie, op welke manier de psychoanalyse theorie kan antwoorden op deze verschillende punten.

De conclusie van ons vijfde hoofdstuk aangaande de psychoanalytische kritiek op het idee van ‘eerste-persoons immuniteit’ volgend uit de fenomenologische wending naar ‘verschijnen’ en ‘lijken’ wordt verder opgenomen in ons finaal, *zesde* hoofdstuk – ‘Expression and the unconscious’. De focus van dit hoofdstuk is tweërlei: enerzijds, een bespreking van Wittgensteins behandeling van ‘eerste-persoons autoriteit’ en de ‘asymmetrie tussen de eerste- en derde-persoon’ met betrekking tot het gebruik van psychologische expressies in de tegenwoordige tijd. Deze lezing wordt in kritisch perspectief gezet met fenomenologische en neurocognitieve, materialistische voorstellen om eerste-persoons autoriteit te verklaren dan wel te verwerpen op basis van het (gebrek aan) epistemische asymmetrie en zelf-referentie. Anderzijds gebruiken we Wittgensteins concept van ‘expressie’ teneinde een non-epistemische verklaring te bieden van de bijwoordelijke betekenis van ‘onbewust’. Zowel Freuds als Lacans respectievelijke manieren om het gebruik van ‘onbewust’ te rechtvaardigen worden kritisch besproken en een alternatieve oplossing voor de logische betekenis van onbewust wordt voorgesteld.

### **Resultaten per hoofdstuk**

In ons *eerste* hoofdstuk focusten we op de verschillende wijzen waarop Merleau-Ponty en Lacan de relatie tussen waanzin en subjectiviteit behandelen door de respectievelijke fenomenologische en psychoanalytische lezingen van Wallons spiegelstadium te contrasteren. We begonnen met de gedeelde assumptie van beide auteurs over de noodzaak om het wetenschappelijk cartesianisme (Lacan, 1949/2006; Merleau-Ponty, 1985) te verlaten dewelke het fenomeen van waanzin reduceert tot een organische of fysische disfunctie. Impliciet in deze gedeelde overtuiging is de assumptie dat dergelijke benaderingen de essentieel humane waarde van waanzin miskennen, een waarde die het specifieke geval van waanzin overstijgt en ons iets van de metafysica van subjectiviteit reveleert. Voor zowel Merleau-Ponty als Lacan wordt deze ‘metafysica’ gemarkeerd door het complex samenspel van herkenning en miskennen. De wegen scheiden echter wanneer het gaat over de conclusies die beide auteurs verbinden aan deze vaststelling: waar Merleau-Ponty de speculaire illusie beschouwt als iets wat gereduceerd dient te worden tot onze meest onmiddellijke, geleefde ervaring, ziet Lacan deze objectiverende illusie als fundamenteel en onvermijdelijk.

Deze laatste kwestie aangaande de relatie tussen subjectieve ervaring en objectivering, en voorts, over de filosofische methode die geschikt wordt geacht om ‘terug te keren’ naar subjectiviteit, was het onderwerp van ons *tweede* hoofdstuk. We begonnen met een korte schets van het actuele debat omtrent de zogenaamde ‘verklaringskloof’ tussen neurocognitieve verklaringen en subjectiviteit. Hoewel we akkoord gingen met deze diagnose, openden we vervolgens het klassieke vraagstuk over hoe deze subjectieve dimensie dan bestudeerd dient te worden via een discussie van Natorp, Husserl en Lacan. We ontwikkelden Natorps neokantiaanse argument met betrekkingen tot de grenzen van

fenomenologische reflecteren en contrasteerde deze met Husserls transcendentale perspectief. We beargumenteerden dat de fenomenologische methode geplaagd wordt door de onverenigbaarheid tussen het fenomenologische appél op ‘gegevenheid’ en de noodzaak aan een reflexieve thematisering van subjectiviteit. Tenslotte boden we een nieuw perspectief op deze discussie vertrekkende van Lacans conceptuele koppel van ‘subject van het uitspreken’ en ‘subject van de uitspraak’. Zoals in het vorige hoofdstuk beklemtoonden we dat voor Lacan aliënatie/objectivering noodzakelijk is voor onze subjectieve identiteit te denken, hoewel deze tegelijkertijd functioneert als datgene waartegen onze subjectieve positie negatief bepaald kan worden. In onze conclusie bespraken we de implicaties van Lacans subjectiviteitsopvatting voor de discussie tussen de fenomenologie en neurocognitieve wetenschap. Enerzijds beargumenteerden we dat Lacans opvatting consistent is met deze filosofische visies waarvoor subjectiviteit gedefinieerd dient te worden als een paradoxaal gegeven dat “is wat het niet is en niet is wat het is” (Sartre, 1958). Anderzijds wezen we op het feit dat Lacan de impliciete overeenkomst tussen neurocognitieve en fenomenologische benaderingen - volgens dewelke de shift naar ‘subjectieve verschijning’ een kwestie is van epistemische zekerheid - op de helling zet.

In ons volgende, *derde* hoofdstuk boden we een verdere uitwerking van het onderscheid tussen ‘uitspreken’ en ‘uitspraak’, dit keer echter om de psychoanalytische notie van ‘waarheid’ te verduidelijken. We contrasteerden dit idee met klassieke correspondentietheorieën door te wijzen op de rol van performatieve waarheid en de disjunctie tussen representatieve kennis en onbewust gedrag.

De uitgesproken filosofische analyses van onze voorgaande hoofdstukken werden gebruikt in ons *vierde* hoofdstuk om een kritische lezing te bieden van recente, neurocognitieve modellen van ‘verbale hallucinaties’ in de schizofrenie. Meer in het bijzonder werd Friths ‘self-monitoring model’ besproken en werd er gewezen op zowel de tekortkomingen op het vlak van de verklaring van ‘agency’ in geval van denken en op de ontoereikende beschrijving van de hallucinatoire ervaring. We besloten onze conclusie met een voorstel voor toekomstig onderzoek waarbij de vraag van zelf-referentie en aliënatie in de schizofrenie geformuleerd dient te worden op basis van het concept van *identificatie*, eerder dan de (in)correcte beschrijving van een ervaring van ‘agency’.

Ons *vijfde* hoofdstuk was in belangrijke mate een voortzetting van de thema’s die we besproken in hoofdstuk twee. De vraag die ons hier echter leidde was hoe zowel de fenomenologie als psychoanalyse dienden te antwoorden op Dennetts neurocognitieve voorstel om ‘subjectiviteit’ en het ‘eerste-persoons perspectief’ te begrijpen als een *theoretische fictie* (Dennett, 1991, p. 79). In het eerste deel ontwikkelden we enkele fenomenologische argumenten die ingezet kunnen worden tegen Dennetts fictionalisering van subjectiviteit op basis van de incompatibiliteit tussen Dennetts heterofenomenologie en zijn eerste-persoons operationalisme. In het tweede deel beargumenteerden we dat de discussie tussen Dennett en de fenomenologie raakt aan een klassieke ambiguïteit in de psychoanalytische theorie met



betrekking tot de vraag of deze laatste best gezien dient te worden als een subpersonele theorie of een descriptieve theorie met een primaire toepassing op het niveau van de persoon. We wezen er echter op dat de psychoanalyse dit onderscheid in de discussie tussen Dennett en de fenomenologie *verwerpt* en ontwikkelden een psychoanalytische visie op ‘verschijning’ waar dit concept verwijst naar die overtuigingen die we onbewust tonen in onze onmiddellijke reacties.

In ons *zesde* en laatste hoofdstuk bogen we ons over de vraag hoe het fenomeen van eerste-persoons autoriteit dient te worden begrepen en wat de implicaties daarvan zijn voor psychoanalytische concepties van ‘het onbewuste’. Hier maakten we gebruik van Wittgensteins concept van ‘expressie’ om eerste-persoons autoriteit en de asymmetrie tussen de eerste en derde personen te begrijpen. Hier beargumenteerden we dat de zelf-attributie van psychologische termen niet gegrond is epistemische detectie (Finkelstein, 2003), maar begrepen dient te worden als de mogelijkheid om mentale toestanden te *manifesteren* of *uit te drukken* in linguïstische attributies. In het volgende deel gebruikten we onze expressivistische verklaring van eerste-persoons autoriteit om de betekenis van ‘onbewuste subjectiviteit’ te verhelderen. We contrasteerden het bijwoordelijke gebruik van ‘onbewust’ met het substantieel gebruik en ontwikkelden Lacans kritiek op deze laatste. Voorts bekritiseerden we twee centrale ideeën in Lacans benadering: ten eerste, dat ‘onbewust’ betrekking heeft op een ‘subject van het onbewuste’ (Lacan, 1981, p. 126); ten tweede, dat de betekenis van onbewust afgeleid kan worden uit de paradoxen van zelfreflectie en eerste-persoons referentie. In ons besluit stelden we een niet-epistemische benadering van de bijwoordelijke bepaling van ‘onbewust’ voor volgens dewelke de kwalificatie ‘onbewust’ niet verwijst naar een gebrek aan zelfkennis, maar eerder naar een specifieke, expressieve positie die ik niet kan innemen.

## **Discussie**

De onderwerpen die we vanuit verschillende perspectieven en door middel van verschillende filosofische auteurs onderzochten in ons proefschrift – gaande van de conflicterende lezingen van Merleau-Ponty en Lacan over het spiegelstadium, de discussie tussen Natorp en Husserl over de grenzen van fenomenologische reflectie, tot Dennetts sceptische herwaardering van het ‘zelf’ als een “centrum van narratieve zwaartekracht” (Dennett, 1992) hebben allen betrekking op de validiteit van het concept ‘subjectiviteit’. Ook de meest recente expressie van deze probleemstelling in de discussie tussen de neurocognitieve wetenschap en de fenomenologie wijst op het feit dat het concept eens te meer een centrale plaats opeist in het actuele debat. Gezien de belangrijke rol die het speelt in ons dagdagelijkse leven en een groot deel van de moderne filosofie hoeft dit enerzijds niet te verbazen. Anderzijds, echter, vraagt de recente explosie van discussies op het kruisvlak van de cognitieve neurowetenschap en de ‘philosophy of mind’ over (zelf)bewustzijn, identiteit, actie en de eerste-persoon, om een meer specifieke verklaring. Deze laatste observatie kan inderdaad verrassend zijn gezien het feit dat tot

recentelijk, bij uitstek in de Franse traditie, de dominante tendens van de filosofie of psychoanalyse bestond uit wat de ‘kritiek’, ‘subversie’ of ‘destructie’ van de categorieën van ‘subject’ en ‘subjectiviteit’ genoemd werd.

Op verschillende momenten in ons proefschrift hebben we gezien dat zowel de verdedigers als critici van het concept subjectiviteit een onderscheid maken tussen de empirische *persoon* en *subject*. Afhankelijk van de traditie wordt dit onderscheid anders ingevuld: de methode van *vooronderstelling* postuleert een subject dat verschillend is van elk empirisch gegeven als de mogelijkheidsvoorwaarde van ervaring en wetenschap (bv. Kant en Natorp); de methode van fenomenologische reductie wil geen subject vooronderstellen maar ervaren, i.e. deze moet gegeven zijn aan zichzelf doorheen een specifieke vorm van ervaring (bv. Husserl en Sartre); psychoanalytische theorie, tenslotte, vooronderstelt opnieuw een subject dat verschillend is van het idee van een empirisch ego, maar deze verliest haar funderende rol en refereert naar dat wat overblijft als negatieve categorie in spanning met haar empirische determinering (bv. Lacan en Žizek). We hebben erop gewezen dat het enige twistpunt tussen deze tradities gelegen is in hoe radicaal dit onderscheid wordt gemaakt – wat echter overeind blijft is het cartesiaanse idee dat ‘subject’ of ‘subjectiviteit’ een bijzonder gegeven aanduidt dat verschilt van hoe we de concepten ‘subject’ of ‘persoon’ dagdagelijks gebruiken.

In onze lezing van Wittgenstein en Anscombe (1981) hebben we beargumenteerd dat dit onderscheid leidt tot een impasse: los van het feit of men een transcendentaal bewustzijn, een existentiële leegte of een subject van het onbewuste evoceert, dergelijke categorieën blijven onbruikbaar om zelfs de meest eenvoudige identiteitsvragen te beantwoorden. Enerzijds wordt de subjectfiguur gebruikt op een manier die nog steeds relevant wordt geacht om ‘wie’-vragen te beantwoorden, maar anderzijds wil men niet dat deze verward wordt haar empirische tegenhanger. Echter, indien elke empirische identificatie wordt verworpen, dan kan de eigenlijke ‘wie’-vraag niet langer beantwoord worden. In ons laatste hoofdstuk maakten we reeds duidelijk dat het achterlaten van deze manier om het probleem van het subject te definiëren, geen probleem stelt voor de psychoanalytische theorie in zoverre deze als taak heeft een verheldering te bieden van het concept ‘onbewuste’. In tegendeel, ons argument was dat het opgeven van deze cartesiaanse affiliatie enkel neerkomt op het opgeven van een vertekend beeld van de eerste-persoon, en bijgevolg, op de betekenis van ‘onbewust’.

Voorts is het ook duidelijk dat de filosofie van het cogito nog steeds invloedrijk is in recente, fenomenologische benaderingen van het probleem van zelf-referentie in de schizofrenie, en meer in het bijzonder, in hoe de misattributie van actie en ‘agency’ bij schizofrene symptomatologie zoals gedachteninsertie of controlewanen momenteel wordt begrepen. Het is gebruikelijk geworden om deze symptomen op het eerste-persoons niveau te omschrijven als problemen in de ‘sense of agency’ en ‘sense of ownership’ (e.g. Bortolotti, 2010; Langland-Hassan, 2008; Roessler, 2013). Beide variabelen worden geacht te refereren naar de prereflexieve (of eerste-orde) *ervaring* of *gevoel* van agency/ownership, dewelke in schizofrenie vooronderstelt wordt verstoord te zijn. Meer algemeen

worden de zelf-attributie van agency en het spreken in de eerste-persoon over intenties en acties in fenomenologische modellen begrepen als een kwestie van het beschrijven/rapporteren van de *ervaring* van agency. Het is belangrijk om op te merken hoe dergelijke benaderingen een klassiek, cartesiaans kader bestendigen. Op dezelfde wijze waarop in een cartesiaanse benadering de mogelijkheid van zelf-referentie gedacht wordt in termen van een ‘ervaring’ van een ik/zelf, is het kenmerkend voor deze traditie om hetzelfde te doen in het geval van psychologische termen (denken, voelen, doen, ...). Zoals Ricoeur opmerkt (1950, p.9) breidde Descartes de betekenis van het latijnse werkwoord ‘cogitare’ uit om al deze werkwoorden te omvatten: ‘voelen is denken’, ‘willen is denken’, ‘bedoelen is denken’, en tenslotte, zelfs ‘ageren is denken’. Meer specifiek verwijst ‘denken’ of ‘cogitare’ in deze uitgebreide betekenis eerder naar ‘de indruk hebben dat’ of ‘de ervaring hebben alsof’. Hoewel mijn bewustzijn me niet kan vertellen dat ik daadwerkelijk zie wat ik geloof te zien, kan het mij tenminste garanderen dat ik de *impressie of ervaring* heb te zien.

In ons laatste artikel hebben we Wittgensteins objectie besproken tegen deze beslissing om alle psychologische werkwoorden in dezelfde categorie van ‘ervaring’ te plaatsen. Deze objectie luidt niet dat dit onmogelijk is, maar enkel omwille van de asymmetrie tussen de eerste –en derde personen in de tegenwoordige tijd. Echter, deze asymmetrie betekent niet dat deze werkwoorden daarom allemaal ervarings-werkwoorden zijn. Kenmerkend voor deze laatste is dat ik de inhoud van mijn ervaring kan beschrijven (het waargenomen van mijn waarnemen, het gevoelde van mijn voelen, ...). In het geval van intentie of agency kan ik echter allerlei dingen ervaren in relatie tot wat ik van plan ben om te doen (opgewondenheid, onrust, ...), maar geen van deze ervaringen constitueert mijn intentie om mijn handeling uit te voeren. Gevraagd worden om de ervaring van mijn actie te beschrijven – bv. of het vrijwillig of onvrijwillig *voelt* – is precies het verwaarlozen van het verschil tussen de taalspellen van *actie* en *ervaring*.

De implicaties van Wittgensteins kritiek op de cartesiaanse tendens alles wat we doen in de eerste-persoon te herleiden tot de algemene term ‘bewuste ervaring’ voor actueel en toekomstig onderzoek naar schizofrenie is duidelijk: het verschil tussen wat vrijwillig of onvrijwillig is, of een actie, beweging of gedachte actief of passief is, kan niet gemaakt worden op basis van het de aan –of afwezigheid van een ‘sense of agency’. Wittgenstein stelt dat hiervoor een ruimere beschouwing nodig is van de intentie, motivatie en context van gedrag. Net zoals Wittgenstein stelt Lacan dat fenomenologische variabelen ontoereikend zijn voor het begrijpen van schizofrene symptomatologie. Eerder dan het wezen van hallucinatoire intentionaliteit te herleiden tot formele verstoringen van het bewustzijn, suggereert Lacan dat we de verschillende facetten van de Ander waarmee de schizofrene patiënt geconfronteerd wordt in rekening dienen te brengen. Gegeven de hedendaagse opleving van eerste-persoons benaderingen van subjectiviteit, is een dergelijk onderzoeksprogramma naar onze mening een van de meest beloftevolle voor de jaren die volgen.

## Referenties

- Anscombe, G.E.M. (1981). "The First Person", in *Metaphysics and the Philosophy of Mind*. Oxford: Basil Blackwell, 21-36.
- Bortolotti, L. (2010). *Delusions and other irrational beliefs*. Oxford: Oxford University Press.
- Dennett, D. (1991). *Consciousness Explained*. New York: Back Bay Books/Little, Brown and Company.
- Dennett, D. (1992). "The self as a center of narrative gravity", in F.S. Kessel, P.M. Cole & D.L. Johnson (Eds.), *Self and Consciousness: Multiple Perspectives*. Hillsdale, NJ: Erlbaum, 4-237.
- Finkelstein, D. (2003). *Expression and the Inner*. Cambridge, Massachusetts: Harvard University Press.
- Freud, S. (1915). The unconscious. *Standard Edition* 14, 166-215.
- Husserl, E. (2002 [1910-1911]). Philosophy as Rigorous Science. *New Yearbook for Phenomenology and Phenomenological Philosophy* 2, 249-295.
- Lacan, J. (1981). *The Seminar 1964, Book XI, The Four Fundamental Concepts of Psychoanalysis*. New York & London: W.W. Norton & Company.
- Lacan, J. (2006 [1947]). "Presentation on Psychical Causality", in J. Lacan & J.-A. Miller (Eds.) *Ecrits*. New York and London: W.W. Norton, 123-160.
- Langland-Hassan, P. (2008). Fractured phenomenologies: thought insertion, inner speech, and the puzzle of extraneity. *Mind & Language* 23 (4), 369-401.
- Laplanche, J. (1999). *Essays on Otherness*. London: Routledge.
- Merleau-Ponty, M. (1985). *L'oeil et l'esprit*. Paris: Gallimard.
- Parnas, J., Sass, L.A. & Zahavi, D. (2008). Recent developments in philosophy of psychopathology. *Current Opinion in Psychiatry* 21 (6), 578-584.
- Ricoeur, P. (1950). *Philosophie de la volonté, t. I, Le volontaire et l'involontaire*. Paris: Aubier.
- Ricoeur, P. (1977). *Freud and Philosophy: An Essay on Interpretation*. New Haven: Yale University Press.
- Roessler, J. (2013). "Thought insertion, self-awareness, and rationality", in K.W.M. Fulford et al. (Eds.) *The Oxford handbook of philosophy and psychiatry*. Oxford: Oxford University Press, 658-672.
- Roy, J.M., Petitot, J., Pachaud, B. & Varela, F.J. (1999). "Beyond the Gap: An Introduction to Naturalizing Phenomenology", in J. Petitot, F.J. Varela, B. Pachaud & J.M. Roy (Eds.), *Naturalizing Phenomenology*. Stanford (California): Stanford University Press, 1-82.
- Sartre, J.P. (1958). *Being and Nothingness: An Essay on Phenomenological Ontology*. London: Methuen.
- Sellars, W. (1991). "Philosophy and the Scientific Image of Man," in *Science, Perception and Reality*, Atascadero: Ridgeview Publishing Company, 1-40.
- Wittgenstein, L. (1988). *Wittgenstein's Lectures on the Philosophy of Psychology 1946-7*, ed. P.T. Geach. Chicago: UP.

