

Subjectivity and Selfhood

Investigating the First-Person Perspective

Dan Zahavi



Subjectivity and Selfhood

Investigating the First-Person Perspective

Dan Zahavi

A Bradford Book
The MIT Press
Cambridge, Massachusetts
London, England

© 2005 Massachusetts Institute of Technology

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

MIT Press books may be purchased at special quantity discounts for business or sales promotional use. For information, please email special_sales@mitpress.mit.edu or write to Special Sales Department, The MIT Press, 55 Hayward Street, Cambridge, MA 02142.

This book was set in Sabon by SNP Best-set Typesetter Ltd., Hong Kong.
Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Zahavi, Dan.

Subjectivity and selfhood: investigating the first-person perspective / Dan Zahavi.

p. cm.

“A Bradford book.”

Includes bibliographical references (p.) and index.

ISBN 0-262-24050-5 (hc : alk. paper)

1. Self (Philosophy). 2. Consciousness. 3. Subjectivity. 4. Self-knowledge, Theory of.

I. Title.

BD438.5.Z34 2006

126—dc22

2005047918

10 9 8 7 6 5 4 3 2 1

Contents

Acknowledgments vii

Introduction 1

1 | Self-Awareness and Phenomenal Consciousness 11

- I Varieties of Self-Awareness 13
- II Higher-Order Theories of Consciousness 17
- III A One-Level Account of Consciousness 20
- IV The Problem of Infinite Regress 25

2 | The Concept(s) of Consciousness in Early Phenomenology 31

- I Three Concepts of Consciousness 31
- II The Stream of Consciousness 32
- III Inner Consciousness and Self-Awareness 37
- IV Some Shortcomings 44

3 | The Structure of Time-Consciousness 49

- I Subjectivity of Experience 50
- II Temporality 55
- III The Internal Object Account 59
- IV *Urbewußtsein* and Self-Affection 65

4 | Reflection and Attention 73

- I Natorp's Challenge 73
- II The Criticism of Reflective Phenomenology 76
- III A Hermeneutical Alternative 78

IV Pure and Impure Reflection 86
V Reflection and Alteration 89
VI Reflective versus Hermeneutical Phenomenology 96

5 | Consciousness and Self 99

I The Non-egological Challenge 99
II Different Notions of Self 103
III The Narrative Concept of Self 106
IV The Self as an Experiential Dimension 115
V Empirical Implications 132
VI A Sense of Self 146

6 | Self and Other 147

I Expression and Empathy 148
II Embodied Subjectivity and Internal Otherness 156
III Beyond Empathy 163
IV The Transcendence of the Other 168
V A Multidimensional Approach 174

7 | Theory of Mind, Autism, and Embodiment 179

I Theory of Mind 179
II Theory-Theory of Self-Awareness 183
III Autism 189
IV A Critical Rejoinder 197
V Autism Revisited 215

Notes 223

References 241

Index 261

What is the relation between (phenomenal) consciousness and the self? Must we evoke a subject of experience in order to account for the unity and continuity of experience, or rather, are experiences anonymous mental events that simply occur without being anybody's states? When we speak of self-awareness, do we then necessarily also speak of a self? Is there always a self involved in self-awareness, or is it also possible to speak of self-awareness without assuming the existence of anybody being self-aware? Answers to these questions are of obvious importance when it comes to a proper understanding both of the structure of consciousness and of what it means to be a self.

I The Non-egological Challenge

Let me illustrate two alternatives by means of Gurwitsch's classical distinction between an egological and a non-egological theory of consciousness (Gurwitsch 1941). An *egological* theory would claim that when I watch a movie by Bergman, I am not only intentionally directed at the *movie*, nor merely aware of the movie being *watched*, I am also aware that it is being watched by *me*, that is, that *I am watching the movie*. In short, there is an object of experience (the movie), there is an experience (the watching), and there is a subject of experience, myself. Thus, an egological theory would typically claim that it is a conceptual and experiential truth that any episode of experiencing necessarily includes a subject of experience (see Shoemaker 1968, 563–564). In contrast, a *non-egological* theory, also known as the *no-ownership* view (see Strawson 1959, 95), would deny that every experience is for a subject. It would, in other words, omit any reference to a subject of experience and simply say that there is an awareness of the watching of the

movie. Experiences are egoless; they are anonymous mental events that simply occur, and minimal self-awareness should, consequently, be understood as the acquaintance that consciousness has with *itself* and *not* as an awareness of an experiencing *self*.

It is not difficult to find arguments against an egological theory of self-awareness and in favor of a non-egological position in twentieth-century philosophy. According to Henrich and Pothast, for instance, to speak of a self or an ego is to speak of an *agent*, that is, some principle of activity and volition. Pre-reflective self-awareness, however, is not something that we initiate or control; it is something that precedes all performances and should, consequently, not be attributed to an ego, but rather be understood as an egoless occurrence. Moreover, if one conceives of the ego qua subject of experience as something that *has* the experience, one obviously makes a distinction between the ego and the experience; they are not identical. In this case, however, it is difficult to understand why the ego's awareness of the experience should count as a case of *self*-awareness. Thus, Henrich and Pothast conclude that it is better to avoid introducing any ego into the structure of basic self-awareness (Henrich 1970, 276, 279; Pothast 1971, 76, 81; cf. Frank 1991b, 252; Cramer 1974, 573).

This view has affinities with the position advocated by Sartre, though he frequently phrased his discussion in terms of the relation between consciousness and self, rather than in terms of the relation between self-awareness and self. In chapter 2, I presented Sartre's position in some detail, so let me just quickly recapitulate a few of his main points. It has often been argued that mental life would dissipate into a chaos of unstructured and separate sensations were it not supported by the unifying, synthesizing, and individuating function of a central and atemporal ego. However, as Sartre pointed out, this reasoning misjudges the nature of the stream of consciousness. The stream of consciousness does not need an exterior principle of individuation, since it is, *per se*, individuated. Nor is it in need of any transcendent principle of unification, since the stream of experiences is self-unifying (Sartre 1936, 21–23). Sartre then went on to point out that an unprejudiced phenomenological description of lived consciousness will simply not include an ego, understood as an inhabitant in or possessor of consciousness. Lived pre-reflective consciousness has no egological structure. As long as we are absorbed in the experience, *living* it through, no ego appears. It is only when we adopt a distancing and objectifying attitude toward the experience in

question, that is, when we reflect upon it, that an ego appears. Even then, however, we are dealing not with an I-consciousness, since the reflecting pole remains non-egological, but merely with a consciousness *of* an ego. The ego is not the subject, but the object of consciousness. It is not something that exists in or behind consciousness, but in front of it (Sartre 1936, 34, 43–44). Thus, Sartre accepted Lichtenberg’s famous critique of Descartes: the traditional rendering of the cogito affirms too much. What is certain is not that “I am aware of this chair,” but only that “there is awareness of this chair” (Sartre 1936, 31–32, 37).

Sartre’s position was not original. Not only was it anticipated—as we have already seen—by Husserl in *Logische Untersuchungen*, but similar views were also advocated by Hume and Nietzsche, who both insisted that the positing of a conscious self or subject is descriptively unwarranted. If we describe the content of our consciousness accurately and pay attention to that which is given, we will not find any self. As Hume famously wrote in *A Treatise of Human Nature*:

For my part, when I enter most intimately into what I call *myself*, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never can catch *myself* at any time without a perception, and never can observe any thing but the perception. (Hume 1888, 252)

One finds the following statement in one of Nietzsche’s manuscripts from the 1880s:

The “subject” is not something given, it is something added and invented and projected behind what there is. (Nietzsche 1960, 903 [1968, 267])

Thus, rather than having experiential reality, the self must be classified as a linguistic construct or as a product of reflection.

Recently, a rather different type of skepticism regarding the self has gained popularity. According to this approach, it is not at all crucial whether or not the self is a given. Whether something is real is not a question of its appearance or of whether it is experienced as real; rather it is a question of whether it can be naturalized and explained by means of the principles and methods employed by natural science. According to this criteria, the self has been weighed and has been found wanting.

One exponent of this view is Metzinger, who in his recent book *Being No One* offers a representationalist and functionalist analysis of what a consciously experienced first-person perspective is. The conclusion he reaches is

quite unequivocal: “no such things as selves exist in the world: Nobody ever *was* or *had* a self” (Metzinger 2003a, 1). Thus, for all scientific and philosophical purposes, the notion of a self can safely be eliminated. It is neither necessary nor rational to assume the existence of a self, since it is a theoretical entity that fulfills no indispensable explanatory function. In reality, the self is not an actually existing object or an unchangeable substance, but a representational construct. All that has been explained previously by reference to a phenomenological notion of “self” can consequently be better explained with the notion of a phenomenally transparent self-model whose representational, or more important, *misrepresentational* (hallucinatory) nature cannot be recognized by the system using it (Metzinger 2003a, 337, 563, 626). The way in which we are given to ourselves on the level of conscious experience must consequently count as a deficit. Biological organisms exist, but an organism is not a self. Some organisms possess self-models, but such self-models are not selves, but merely complex brain states (Metzinger 2003a, 563). Whenever we speak of a “self,” we consequently commit what Metzinger alternately calls the phenomenological fallacy, or the error of phenomenological reification. We confuse the content of an ongoing subpersonal self-representational process with a real existing entity (Metzinger 2003a, 268). All that really exist are certain types of information-processing systems engaged in operations of self-modeling, and we should not commit the mistake of confusing a model with reality (Metzinger 2003b, 370, 385, 390). To be more precise (since there is no I, you, or we), owing to an autoepistemic closure or lack of information, owing to a special form of epistemic darkness, the self-representing system is caught up in a naive-realistic self-misunderstanding (Metzinger 2003a, 332, 436–437, 564). Properly speaking, there is no one who confuses herself with anything, since there is no one who could be taken in by the illusion of a conscious self (Metzinger 2003a, 634). The self is a mere appearance, and on several occasions Metzinger compares the recognition of the illusionary or fictitious character of one’s own self with the kind of insight that is one of the main goals of Buddhist enlightenment (Metzinger 2003a, 550, 566).

We will have occasion to return to Metzinger later on, but it is already appropriate now to present a few critical remarks. Metzinger argues that there are no such things as selves or subjects of experience in the world. All that exist are phenomenal selves, that is, selves that are nothing but properties of complex representational processes (Metzinger 2003a, 577).

Granted that this is true, however, why does Metzinger adopt a no-self doctrine? Why does he take the self to be an illusion? Why does he not, rather, argue like Churchland, who writes “The brain makes us think that we have a self. Does that mean that the self I think I am is not real? No, it is as real as any activity of the brain. It does mean, however, that one’s self is not an ethereal bit of ‘soul stuff’” (Churchland 2002, 124)? Part of the reason for this seems to be that Metzinger, himself, remains committed to the rather classical definition of the self, according to which the self is a mysteriously unchanging essence, a process-independent ontological substance that could exist all by itself, that is, in isolation from the rest of the world (Metzinger 2003a, 577, 626). Metzinger denies the existence of such an entity and then concludes that no such things as selves exist. The only reason to accept his conclusion would be if this notion of self were the only one available. And as we will shortly see, this is by no means the case. It is obviously possible to speak of the self or ego the way Henrich, Pothast, Sartre, Hume, Nietzsche, and Metzinger do. One problem with their skeptical reservations, however, is that they all presuppose rather specific concepts of self, which they then proceed to criticize. Yet is it at all clear what, precisely, a self is?

II Different Notions of Self

On closer examination, it should be obvious that it is an exaggeration to claim that the notion of “self” is unequivocal and that there is widespread consensus about what, exactly, it means to be a self. Quite to the contrary, if one looks at contemporary discussions one will find them to be literally bursting with completing and competing definitions of the self. In a well-known article from 1988, Neisser distinguished five types of self: the ecological, interpersonal, extended, private, and conceptual self (Neisser 1988, 35). Eleven years later, Strawson summed up a recent discussion on the self in the *Journal of Consciousness Studies* by enumerating no fewer than twenty-one concepts of self (Strawson 1999, 484). Given this escalating abundance, it is quite easy to talk at cross-purposes, particularly in an interdisciplinary context. It is a simple fact that the concept of self connotes different things in different disciplines—sometimes radically different things. What is urgently needed is, consequently, a clarification of the relationships between these various conflicting and/or complementary notions of self. Moreover, such a taxonomic clarification is indispensable if one is to

evaluate the merits of the no-self doctrine. What I intend to do in the following is to contrast a rather classical understanding of the self that, to a large extent, is targeted by the non-egological criticism with two alternate and more contemporary ways of conceiving of the self.

A Kantian Perspective: The Self as a Pure Identity-Pole

This traditional view insists on distinguishing between the identical self on the one hand and the manifold of changing experiences on the other. In turn, I can taste an ice cream, smell a bunch of roses, admire a statue by Michelangelo, and recollect a hike in the Alps. We are here faced with a number of different experiences, but they also have something in common; they all have the same subject, they are all lived through by one and the same self, namely myself. Whereas the experiences arise and perish in the stream of consciousness, the self remains as one and the same through time. More specifically, the self is taken to be a distinct *principle of identity* that stands apart from and above the stream of changing experiences and which, for that very reason, is able to structure it and give it unity and coherence (cf. Kant 1956: B 132–133).

The notion of self at work here is obviously a formal and abstract one. Every experience is always lived through *by* a certain subject; it is always an experience *for* a certain subject. The self is, consequently, understood as the pure subject, or ego-pole, that any episode of experiencing necessarily refers back to. It is the subject of experience rather than the object of experience. Instead of being something that can itself be given as an object for experience, it is a necessary condition of the possibility for (coherent) experience. We can infer that it must exist, but it is not itself something that can be experienced. It is an elusive principle, a presupposition, rather than a datum or something that is itself given. Were it given, it would be given for someone, that is, it would be an object, and therefore no longer a self (see Natorp 1912, 8, 40). As Kant wrote in *Kritik der reinen Vernunft*: “It is . . . evident that I cannot know as an object that which I must presuppose to know any object” (Kant 1956, A 402).

A Hermeneutical Perspective: The Self as a Narrative Construction

A quite different way of conceiving the self takes its point of departure in the fact that self-comprehension and self-knowledge, rather than being something that is given once and for all, is something that has to be appropriated and can be attained with varying degrees of success. As long as life goes on,

there is no final self-understanding. The same, however, can also be said for what it means to be a self. The self is not a thing; it is not something fixed and unchangeable, but rather something evolving. It is something that is realized through one's projects, and it therefore cannot be understood independently of one's own self-interpretation. As Jopling puts it: "Selfhood is best viewed as a kind of ongoing project that serves as a response to the question of how to be" (Jopling 2000, 83). In short, one is not a self in the same way as one is a living organism. One does not have a self in the same way that one has a heart or a nose (Taylor 1989, 34). To have a self, or better, to be a self, is something in which one is existentially involved.

According to this view, which has become increasingly popular lately, the self is assumed to be a construction. It is the product of conceiving and organizing one's life in a certain way. When confronted with the question "Who am I?" we will tell a certain story and emphasize aspects that we deem to be of special significance, to be that which constitutes the *leitmotif* in our life, that which defines who we are, that which we present to others for recognition and approval (Ricoeur 1985, 442–443). This narrative, however, is not merely a way of gaining insight into the nature of an already existing self. On the contrary, the self is first constructed in and through the narration. Who we are depends on the story we (and others) tell about ourselves. The story can be more or less coherent, and the same holds true for our self-identity. The narrative self is, consequently, an open-ended construction that is under constant revision. It is pinned on culturally relative narrative hooks and organized around a set of aims, ideals, and aspirations (Flanagan 1992, 206). It is a construction of identity starting in early childhood and continuing for the rest of our life, which involves a complex social interaction. Who one is depends on the values, ideals, and goals one has; it is a question of what has significance and meaning for one, and this, of course, is conditioned by the community of which one is part. Thus, as has often been claimed, one cannot be a self on one's own, but only together with others, as part of a *linguistic* community. As Taylor puts it, "There is no way we could be inducted into personhood except by being initiated into a language" (Taylor 1989, 35).

A Phenomenological Perspective: The Self as an Experiential Dimension

The phenomenological alternative I now consider can be seen as a replacement of the first notion of self and as a necessary founding supplement for the second notion of self. The crucial idea is that an understanding of what

it means to be a self calls for an examination of the structure of experience, and vice versa. In other words, the investigations of self and experience have to be integrated if both are to be understood. More precisely, the self is claimed to possess experiential reality, is taken to be closely linked to the first-person perspective, and is, in fact, identified with the very first-personal *givenness* of the experiential phenomena. As Michel Henry would have put it, the most basic form of selfhood is the one constituted by the very self-manifestation of experience (Henry 1963, 581; 1965, 53). To be conscious of oneself, consequently, is not to capture a pure self that exists in separation from the stream of consciousness, but rather entails just being conscious of an experience in its first-personal mode of givenness; it is a question of having first-personal access to one's own experiential life. Thus, the self referred to is not something standing beyond or opposed to the stream of experiences but is rather a feature or function of its givenness. In short, the self is conceived neither as an ineffable transcendental precondition, nor as a mere social construct that evolves through time; it is taken to be an integral part of our conscious life with an immediate experiential reality.

This third notion of self, just like the Kantian notion, is a very formal and minimalist notion, and it is obvious that far more complex forms of selves exist. With this said, however, the phenomenological notion nevertheless strikes me as being of pivotal significance. It is fundamental in the sense that nothing that lacks this dimension deserves to be called a self. Thus, in my view, this experiential sense of self deserves to be called the *minimal self* or the *core self*. In order to substantiate this suggestion, we must look closer at the last two concepts of self, the hermeneutical and the phenomenological. Let us start with the former.

III The Narrative Concept of Self

It has recently been argued that it is impossible to discuss the issues of selfhood and personal identity in abstraction from the temporal dimension of human existence (Ricoeur 1990, 138). Human time, however, is neither the subjective time of consciousness nor the objective time of the cosmos. Rather, human time bridges the gap between phenomenological and cosmological time. Human time is the time of our life stories; a narrated time structured and articulated by the symbolic mediations of narratives (Ricoeur 1985, 439). What contributions do such narratives make to the constitution of the

self? It has been suggested that they make up the essential form and central constitutive feature of self-understanding and self-knowledge.

In order to know who you are, in order to gain a robust self-understanding, it is not enough to simply be aware of oneself from the first-person perspective. It is not sufficient to think of oneself as an I; a narrative is required. To answer the question “Who am I?” is to tell the story of a life (Ricoeur 1985, 442). I attain insight into who I am by situating my character traits, the values I endorse, the goals I pursue within a life story that traces their origin and development; a life story that tells where I am coming from and where I am heading. This narrative, however, is not merely a way of gaining insight into the nature of an already existing self. On the contrary, the self is the product of a narratively structured life. As MacIntyre puts it, the unity of the self “resides in the unity of a narrative which links birth to life to death as narrative beginning to middle to end” (MacIntyre 1985, 205). Thus, for MacIntyre, personal identity is the identity presupposed by the unity of the character that the unity of the narrative requires, or, to put it differently, the notion of a character in a story is more fundamental than the concept of a person. The latter concept is simply the concept of a character abstracted from its history (MacIntyre 1985, 217–218).

Why is it natural for us to think of the self in terms of narrative structures? This is because human activities are enacted narratives; our actions gain intelligibility by having a place in a narrative sequence. We live out narratives in our lives and we understand our own lives in terms of such narratives: “Stories are lived before they are told—except in the case of fiction” (MacIntyre 1985, 212). Stories involve agents and patients, people who act and suffer. It is within the framework of such narratives that we can ask the central who-questions: “Who is this?”; “Who did this?”; “Who is responsible?” The answers to these questions are provided by the narrative itself. The self is the “who” of the story, the one upon whom the story confers an identity (Villela-Petit 2003, 3). To ask for the identity of the one who is responsible is, consequently, to ask for his narrative identity. This is why, according to MacIntyre, any attempt to elucidate the notions of selfhood or personal identity independently of and in isolation from the notions of narrativity, intelligibility, and accountability is bound to fail (MacIntyre 1985, 218).

Ricoeur has sought to clarify the concept of narrative identity by means of two further concepts of identity: identity as sameness (*mêmeté*) and

identity as selfhood (*ipséité*). The first concept of identity, the identity of the same (Latin: *idem*), conceives of the identical as that which can be reidentified again and again, as that which resists change. The identity in question is that of an unchangeable substance, or substrate, that remains the same over time. As Ricoeur points out, not all problems of personal identity can be conceived as problems concerning the possibility of reidentification and tackled by means of the concept of sameness. Thus, the second concept of identity, the identity of the self (Latin: *ipse*), has very little to do with the persistence of some unchanging personality core. It is, primarily, not a question concerning the kinds of causal links that are required if we are to identify P_2 at t_2 as the same as P_1 at t_1 . Rather, its identity condition is linked to the question of self-understanding, to the question “who am I” (Ricoeur 1990, 12–13, 140). When confronted with this question, I am forced to reflect on and evaluate my way of living, the values I honor, and the goals I pursue. I am forced to confront the life I am living. Thus, the answer to the question is not immediately accessible; rather it is the fruit of an examined life.

Whereas questions such as “What is x ” or “Is x at t_1 identical to y at t_2 ”—questions regarding *idem*-identity—can be answered from a third-person perspective and be given definite and informative answers, questions regarding *ipse*-identity, such as “Who am I,” must include an approach from the first-person perspective and will never find an exhaustive answer.

Ricoeur has occasionally presented his own notion of narrative identity as a solution to the traditional dilemma of having to choose between the Cartesian notion of the self as a principle of identity that remains the same throughout the diversity of its various states and the positions of Hume and Nietzsche, who held an identical subject to be nothing but a substantialist illusion (Ricoeur 1985, 443). Ricoeur suggests that we can avoid this dilemma if we replace the notion of identity that they respectively defend and reject with the concept of narrative identity. The identity of the narrative self rests on narrative configurations. Unlike the abstract identity of the same, the narrative identity can include changes and mutations within the cohesion of a lifetime. The story of a life continues to be reconfigured by all the truthful or fictive stories a subject tells about him- or herself. It is this constant reconfiguration that makes “life itself a cloth woven of stories told” (Ricoeur 1985, 443 [1988, 246]).

Any consideration of narrative identity obviously entails a reference to others, since there is a clear social dimension to the achievement of narra-

tive self-understanding. Narrative self-understanding requires maturation and socialization and the ability to access and issue reports about the states, traits, and dispositions that make one the person one is. I come to know who I am and what I want to do with my life by participating in a community. To come to know oneself as a person with a particular life history and particular character traits is, consequently, both more complicated than knowing one's immediate beliefs and desires and less private than it might initially seem (Jopling 2000, 137). When I interpret myself in terms of a life story, I might be both the narrator and the main character, but I am not the sole author. Whereas I, as the author of a literary text, am free to determine the beginning, middle, and end of the story, the beginning of my own story has always already been made for me by others, and the way the story unfolds is determined only in part by my own choices and decisions. As MacIntyre points out:

[W]e are never more (and sometimes less) than the co-authors of our own narratives. Only in fantasy do we live what story we please. In life, as both Aristotle and Engels noted, we are always under certain constraints. We enter upon a stage which we did not design and we find ourselves part of an action that was not of our making. (MacIntyre 1985, 213)

Who we are depends on the stories told about us, both by ourselves and by others. Our narrative self is multiple-authored and under constant revision. The story of any individual life is not only interwoven with the stories of others (parents, siblings, friends, etc.), it is also embedded in a larger historical and communal meaning-giving structure (MacIntyre 1985, 221). The concepts I use to express the salient features of the person I take myself to be are concepts derived from tradition and theory that will vary widely from one historical period to the next and across social class and culture. To think of oneself as a citizen, an academic, a European, as hot tempered, handsome, clever, weak willed, amblyopic, anorectic, or anemic is to think of oneself by means of concepts that are embedded within diverse theoretical frameworks, be they of a sociological, biological, psychological, or religious provenance (see Neisser 1988, 53–54).

Ricoeur has frequently been regarded as one of the main proponents of a narrative approach to the self. Although it is undeniable that he has made decisive contributions to the discussion, Ricoeur himself has also pointed to some of the limitations of this approach. As he states in *Temps et récit*, narrative identity is the name of a problem at least as much as it is that of a

solution (Ricoeur 1985, 446). Let us, then, take a closer look at this problem, or, to be more precise, at some of these problems.

Fiction and Confabulation

It is possible to tell different, even incompatible, stories about one and the same life, but not all of them can be true. The fact that our narration can, and does, include fictional components gives rise to at least two questions. First, how do we distinguish true narratives from false narratives? It is obvious that a person's sincere propagation of a specific life story does not guarantee its truth. In fact, in some cases the stability of our self-identity might be inversely proportional to the fixed stories we tell about ourselves. Elaborate storytelling might serve a compensatory function as an attempt to make up for the lack of a coherent self-identity.

Since the internal coherency of a story is no guarantee of its accuracy, it must be complemented by other constraints. Jopling has recently suggested some additional constraints. A self-narrative should not only be (1) internally coherent, it should also be (2) externally coherent, that is, it should fit with the narratives told about me by other people, and it should be (3) applicable to my current life situation, since a self-narrative is not meant to be relevant only for the understanding of the past, but is also meant to entail a forward-looking commitment to a broadly unified set of possible actions. Finally, it should (4) fit with narrative-transcendent facts (Jopling 2000, 50).

This reference to narrative-transcendent facts, however, calls attention to the second, more worrying issue: What is a narrative self-understanding an understanding of? What is the question "Who am I?" a question about? Is the self an independently existing entity that makes the questions we ask about it true or false? Is it something whose nature we gradually unearth, or rather, is it wholly constituted and constructed by our descriptions? Some defenders of a narrative approach to selfhood have argued that the self is nothing but a linguistic and social invention. As Dennett puts it, biological organisms with brains like ours cannot prevent themselves from inventing selves. We are hardwired to become language users, and the moment we make use of language, we begin spinning our stories. The self is produced in this spinning, but it has no reality; it is merely a fictional center of narrative gravity (Dennett 1991, 418; see also 1992). It is the abstract point where various stories about us intersect. Thus, on this reading, the narrative account turns out to be a variant of the no-self doctrine.

Dennett has compared the notion of self to the notion of a center of gravity. The latter is an abstractum, and although it has a well-defined role to play within physics, it remains a theorist's fiction; it is not a real thing in the universe. The situation is quite similar when it comes to selves, Dennett claims. They are theoretical fictions that we find it perspicuous to employ when we engage in the interpretation and prediction of behavior. It facilitates our predictions to organize our interpretations around such a central abstraction, but this does not change the fact that the self is, and remains, a fiction.

According to Dennett, selves are theoretical fictions that differ from theoretically inferred entities, such as subatomic particles, by having only the properties endowed by the theory that constitutes them. There are no theory-transcendent constraints; there is nothing to be discovered. To ask *what* a self really is, or to ask—as some neuroscientists do—*where* the self is, is quite simply a category mistake (Dennett 1992).

Some might argue that fictional selves depend on real selves for their creation. Dennett, however, considers this suggestion mistaken. He asks us to consider the following scenario: Let us imagine a highly sophisticated computer that has been designed to write novels. The first novel it writes is the apparent autobiography of some fictional person called Gilbert. Gilbert is a fictional self, yet its creator—the computer—is not a real self. There may have been human designers who built the computer, but the narrative construction of Gilbert did not involve any selves. Dennett then asks us to expand upon the story. Let us turn the novel-writing computer into a robot and assume that it is outfitted with wheels and cameras and that it can move around in the world. It still writes a novel about Gilbert, but strangely enough, the adventures of Gilbert bear a striking relationship to the adventures of the robot. If you hit the robot with a baseball bat, the story about Gilbert will, shortly thereafter, include a section that describes how he was hit with a baseball bat by somebody looking like you. If you instead help the robot, it will send you a note saying “Thank you. Love, Gilbert.” Thus, Gilbert seems to be real, seems to be the robot. In truth, however, there is no Gilbert, there is no self. All that exists are patterns of behavior that can be interpreted by means of a narrative that includes a reference to a self. Needless to say, on Dennett's view, this holds true not only in the case of the robot and Gilbert, but in the case of each of us as well (Dennett 1992).

Let us grant that the narrative self is a construction. It is not something innate, and the material used for its construction consists not only of real-life materials, but also ideals and fictive ideas. Let us grant that our narrative identity is subject to constant revisions and that it is organized around numerous narrative hooks that differ from culture to culture. Does this, then, justify the claim that the self is nothing but a fiction? Let us not forget that there are constraints; some self-narratives are more true than others. As Flanagan points out, it is undeniable that the self plays a crucial role in our psychological and social life by giving it organization, meaning, and structure. Thus, in his view, the narrative self might be a construction, but that does not make it unreal (Flanagan 1992, 205–210). Ricoeur and MacIntyre would, obviously, agree. Although both reject the idea of a substantial self, they would insist that human life has a natural narrative structure. To declare everything peculiar to human life fictitious simply because it cannot be naturalized, because it cannot be grasped by a certain mode of scientific comprehension, merely reveals one's prior commitment to a naive scientism, according to which (natural) science is the sole arbiter of what there is.

I fully share this view with Flanagan, Ricoeur, and MacIntyre; yet a lingering doubt remains. Is it possible to resist Dennett's conclusion as long as the self is taken to be nothing but a narrative construction? Does the narrative self not require some kind of experiential support?

Finitude

We are finite creatures. No finite, fallible creature can explicate the full story of its life. Self-narratives may capture something important about who we are, but are they capable of capturing the full complexity of the self? Is it legitimate to reduce our selfhood to that which can be narrated? Could it not be argued that we actually learn more about ourselves when confronted with situations that make us step out of smooth, unifying narratives, that make us act "out of character"? Furthermore, it might not only be objected that stories are told rather than lived and that narratives differ from life, it might also be argued that the very attempt to present human life in the form of a narrative will necessarily transform it. The storyteller will inevitably impose an order on the life events that they did not possess while they were lived. To form a self-narrative more must be done than simply recall and

recount certain life events. One must also consider these events reflectively and deliberate on their meaning to decide how they fit together. All of this involves a certain element of confabulation that goes beyond the lived life itself (see Gallagher 2003a).

MacIntyre's reply to the latter kind of objection is that the only picture we can envisage of a human life, prior to its alleged misinterpretation by a narrative, is in the form of a sequence of disjointed fragments or snapshots, which, in his view, simply proves his point. To talk of the sequence as being fragmented is to measure it against the narrative that continues to remain the framework of intelligibility (MacIntyre 1985, 212–215). This rejoinder, however, is too easy. Although it might be true that many of our actions easily lend themselves to narrative articulation, human life is made up of more than just actions. Moreover, it is one thing to claim that actions can be narrated, and something quite different to claim that they can all be fitted into one unifying narration without thereby imposing more unity upon them than they had to start with. As Drummond has recently put it:

Narratives are reflective selections and organizations of a life. In this sense the narrative captures less than an individual's life, for not all of a life as pre-reflectively lived can be fitted into a narrative, which best suits goal-directed action. From the opposite perspective, narratives, by virtue of their selectivity, impose more unity than life itself has manifested. . . . [W]e should not confuse the reflective, narrative grasp of a life with an account of the pre-reflective experience that makes up that life prior to that experience being organized into a narrative. (Drummond 2004, 119)

Ethics and Beyond

Despite his being heralded as one of the leading protagonists of the narrative approach to selfhood, one of Ricoeur's conclusions, reached in *Temps et récit* but only fully developed in *Soi-meme comme un autre*, is that the discussion of narrative identity does not exhaust the question concerning the identity of the self. Selfhood cannot be reduced to narrative identity since the identity of the self is only fully revealed the moment we include the *ethical dimension*.

As Ricoeur argues, personal identity has two poles, and narrative identity is what links the two (Ricoeur 1990, 195). The first pole is constituted by *character*, which is, as he puts it, the *what* in the *who* (Ricoeur 1990, 147). The character is the totality of our enduring dispositions and habits, those

distinctive traits that enable others to recognize and reidentify us. Our character, consequently, has a temporal dimension; it expresses permanence in time. As a second, acquired nature my character is I, myself, but it is also a dimension of self that announces itself as *idem*. To speak of character is to speak of the self in the manner of the same. It is the limit point where a discussion of *ipse* is indiscernible from a discussion of *idem* (Ricoeur 1990, 143). The second pole is constituted by what Ricoeur calls *faithfulness of self*, that is, by the ethical dimension of the self. This is the dimension of pure *ipseity* where *ipse* and *idem* are completely dissociated and where the question of self-identity is a question of *accountability* and *responsibility* (Ricoeur 1990, 143, 179, 195). As Ricoeur already pointed out in 1950, in *Philosophie de la volonté*:

I form the consciousness of being the author of my acts in the world and, more generally, the author of my acts of thought, principally on the occasion of my contacts with an other, in a social context. Someone asks, who did that? I rise and reply, I did. Response-responsibility. To be responsible means to be ready to respond to such a question. (Ricoeur 1950, 55 [1966, 56–57])

Thus, to be a self is not simply a question of storytelling. It is also a question of adopting certain norms as binding; to be bound by obligation or loyalty. It is to remain true to oneself in promise keeping. It is to be somebody others can count on. It is to assume responsibility for one's past actions and for the future consequences of one's present actions, regardless of how much one's self-narrative might change (Ricoeur 1990, 341–342). Thus, Ricoeur ends up arguing that the narrative take on selfhood must be complemented by a different perspective that includes the issue of *ethical responsibility*.¹

Although I fully agree with Ricoeur's concession that a discussion of narrative identity is insufficient if we want to understand the full complexity of the self, I will not follow Ricoeur on his excursion into ethics. Rather, I want to suggest that the narrative or hermeneutical take on self must be complemented by an experiential or phenomenological take on the self. To put it very simply, it takes a self to experience one's life as a story. In order to begin a self-narrative, the narrator must be able to differentiate between self and nonself, must be able to self-attribute actions and experience agency, and must be able to refer to him- or herself by means of the first-person pronoun. All of this presupposes that the narrator is in possession of a first-person perspective.

IV The Self as an Experiential Dimension

The term “*ipseity*” (selfhood, from the Latin, *ipse*) has gained a recent popularity as a result of Ricoeur’s writings on narrative and what he calls his “herméneutique de l’ipséité” (Ricoeur 1990, 357). However, Ricoeur is by no means the first French thinker to employ the term, and if we look briefly at the way the term has been used by some of his phenomenological predecessors (Merleau-Ponty, Sartre, and Henry) we will begin to gain a better understanding of what the phenomenological concept of self amounts to.

Merleau-Ponty occasionally spoke of the subject as realizing its *ipseity* in its embodied being-in-the-world (Merleau-Ponty 1945, 467). He also, however, referred to Husserl’s investigations of inner time-consciousness and wrote that the original temporal flow involves a self-manifestation. Consciousness constitutes itself in terms of itself and, as Merleau-Ponty then stated, the temporal explosion of the present toward the future counts as the archetypical relationship of self to self and traces out an interiority or *ipseity* (Merleau-Ponty 1945, 487–488).

We have already come across Sartre’s dismissal of an egological account of consciousness in *La transcendance de l’ego*. Whereas Sartre, in that early work, had characterized non-egological consciousness as *impersonal*, he described this view as mistaken in both *L’être et le néant* and in his important 1948 article “Conscience de soi et connaissance de soi.” Although no ego exists on the pre-reflective level, consciousness remains personal because consciousness is, at bottom, characterized by a fundamental self-givenness or self-referentiality that Sartre called *ipseity*:

Thus, the ego appears to consciousness as a transcendent in-itself, as an existent in the human world, not as *of* the nature of consciousness. Yet we need not conclude that the for-itself is a pure and simple “impersonal” contemplation. The ego is far from being the personalizing pole of a consciousness which, without it, would remain in the impersonal stage; on the contrary, it is consciousness in its fundamental *ipseity* which, under certain conditions, allows the appearance of the ego as the transcendent phenomenon of that *ipseity*. (Sartre 1943, 142 [1956, 103; translation modified]. see also 1943, 162, 284; 1948, 63)

Sartre’s crucial move was, consequently, to distinguish between ego and self. From the context, it is obvious that Sartre had nothing like narrative identity in mind when he spoke of *ipseity*. He was referring to something much more basic, something characterizing consciousness as such. It is something

that distinguishes my very mode of existence, and, although I can fail to articulate it, it is not something I can fail to be. As he also wrote, “pre-reflective consciousness is self-consciousness. It is this same notion of *self* which must be studied, for it defines the very being of consciousness” (Sartre 1943, 114 [1956, 76]).

The most focused discussion of *ipseity*, however, is to be found in the work of Michel Henry. Henry repeatedly characterized selfhood in terms of an interior self-affection (Henry 1963, 581, 584, 585). Insofar as subjectivity reveals itself to itself, it is an *ipseity* (Henry 2003, 52). As he put in his early work, *Philosophie et phénoménologie du corps*: “The interiority of the immediate presence to itself constitutes the essence of ipseity” (Henry 1965, 53 [1975, 38]). For Henry, there was a clear connection between being a self and being self-aware. It is because consciousness is characterized, as such, by self-awareness that we can ascribe it a fundamental type of *ipseity*.

What we find in all three thinkers is, consequently, an attempt to link a basic sense of self to the first-personal givenness of experiential life. Let us take a closer look at the structure of first-personal givenness in order to better understand this line of thought.

What It Is Like

Whereas we cannot ask what it feels like to be a piece of soap or a radiator, we can ask what it is like to be a cat, a wolf, or another human being, because we take them to be conscious and to have experiences. Experiences are not something that one simply has, like coins in the pocket. On the contrary, experiences have a subjective “feel” to them, that is, a certain (phenomenal) quality of “what it is like” or what it “feels” like to have them. This is obviously true of bodily sensations like pain or nausea. It is also the case for perceptual experiences, as well as desires, feelings, and moods. There is something it is like to taste an omelet, touch an ice cube, crave chocolate, have stage fright, or to feel envious, nervous, depressed, or happy. Should one limit the phenomenal dimension of experience to *sensory* or *emotional* states alone? Is there nothing it is like to simply think of, rather than perceive, a green apple? What about abstract beliefs; is there nothing it is like to believe that the square root of nine equals three?

In *Logische Untersuchungen* (1900–1901), Husserl argued that conscious thoughts have experiential qualities and that episodes of conscious thoughts are experiential episodes. In arguing for this claim, Husserl drew some dis-

inctions that remain relevant. According to Husserl, every intentional experience possesses two different but inseparable moments. Every intentional experience is an experience of a specific type, be it an experience of judging, hoping, desiring, regretting, remembering, affirming, doubting, wondering, fearing, and so on. Husserl called this aspect of the experience the *intentional quality* of the experience. Every intentional experience is directed at something, is also about something, be it an experience of a deer, a cat, or a mathematic state of affairs. Husserl called the component that specifies what the experience is about the *intentional matter* of the experience (Hua 19/425–426). The same quality can be combined with different matters, and the same matter can be combined with different qualities. It is possible to doubt that “the inflation will continue,” that “the election was fair,” or that “one’s next book will be an international bestseller,” just as it is possible to deny that “the lily is white,” to judge that “the lily is white,” or to question whether “the lily is white.” Husserl’s distinction between the intentional matter and the intentional quality, therefore, bears a certain resemblance to the contemporary distinction between propositional content and propositional attitudes, though it is important to emphasize that Husserl, by no means, took all intentional experiences to be propositional in nature.

Furthermore, and this is of course the central point, Husserl considered these cognitive differences to be *experiential* differences. Each of the different intentional qualities has its own phenomenal character. There is an *experiential* difference between affirming and denying that Hegel was the greatest of the German idealists, just as there is an *experiential* difference between expecting and doubting that Denmark will win the 2006 FIFA World Cup. What it is like to be in one of these occurrent intentional states differs from what it is like to be in another of these occurrent intentional states.² Similarly, the various intentional matters each have their own phenomenal character. There is an *experiential* difference between entertaining the occurrent belief that “thoughts without content are empty” and the belief that “intuitions without concepts are blind,” just as there is an *experiential* difference between denying that “the Eiffel Tower is higher than the Empire State Building” and denying that “North Korea has a viable economy.” To put it differently, a change in the intentional matter entails a change in what it is like to undergo the experience in question.³ These experiential differences, these differences in what it is like to think different thoughts, are not simply sensory differences.⁴

In the same work, Husserl also called attention to the fact that one and the same object can be given in a variety of modes. This is not only the case for spatiotemporal objects—the same tree can be given from this or that perspective, as perceived or recollected, and so on—but also for ideal or categorical objects. There is an experiential difference between thinking of the theorem of Pythagoras in an empty and signitive manner, without really understanding it, and doing so in an intuitive and fulfilled manner by actually thinking it through with comprehension (Hua 19/73, 667–676). In fact, as Husserl pointed out, our understanding of signs and verbal expressions can illustrate these differences most clearly:

Let us imagine that certain arabesques or figures have at first affected us merely aesthetically, and that we then suddenly realize that we are dealing with symbols or verbal signs. In what does this difference consist? Or let us take the case of a man attentively hearing some totally strange word as a sound-complex without even dreaming it is a word, and compare this with the case of the same man afterwards hearing the word, in the course of conversation, and now acquainted with its meaning, but not illustrating it intuitively. What in general is the surplus element distinguishing the understanding of a symbolically functioning expression from the uncomprehended verbal sound? What is the difference between simply looking at a concrete object *A*, and treating it as a representative of “any *A* whatsoever”? In this and countless similar cases it is the act-characters that differ. (Hua 19/398 [2001, II/105])

Strawson has argued more recently in a similar fashion. He asks us to consider a situation wherein Jacques, a monoglot Frenchman, and Jack, a monoglot Englishman, are both listening to the same French news program. The experiences of Jacques and Jack are certainly not the same, for only Jacques is able to understand what is being said. Only Jacques is in possession of what might be called an experience of understanding. To put it another way, there is normally something it is like, experientially, to understand a sentence. There is an experiential difference between hearing something that one does not understand, and hearing and understanding the very same sentence. This experiential difference is not a sensory difference, but a cognitive one (Strawson 1994, 5–6). This is why Strawson can write that

the apprehension and understanding of cognitive content, considered just as such and independently of any accompaniments in any of the sensory-modality-based modes of imagination or mental representation, is part of experience, part of the flesh or content of experience, and hence, trivially, part of the qualitative character of experience. (Strawson 1994, 12)

Every conscious state, be it a perception, an emotion, a recollection, or an abstract belief, has a certain subjective character, a certain phenomenal quality of “what it is like” to live through or undergo that state. This is what makes the mental state in question *conscious*. In fact, the reason we can distinguish occurrent conscious mental states from each other is exactly because there is something it is like to be in those states. The widespread view that only sensory and emotional states have phenomenal qualities must, therefore, be rejected. Such a view is not only simply wrong, phenomenologically speaking, but its attempt to reduce phenomenality to the “raw feel” of sensation marginalizes and trivializes phenomenal consciousness and is detrimental to a correct understanding of its cognitive significance.⁵

First-personal Givenness

When asked to exemplify the “what it is like” quality of experience, one will often find references to what have traditionally been called secondary sense qualities, such as the smell of coffee, the color of red silk, or the taste of a lemon. These answers reveal an ambiguity in the notion of “what it is like.” Normally, the “what it is like” aspect is taken to designate experiential properties. If, however, our experiences are to have qualities of their own, they must be qualities over and above whatever qualities the intentional object has. It is exactly the silk that is red, and not my perception of it. Likewise, it is the lemon that is bitter, and not my experience of it. The *taste* of the lemon is a qualitative feature of the lemon and must be distinguished from whatever qualities my *tasting* of the lemon has. Even if there is no other way to gain access to the gustatory quality of the lemon than by tasting it, this will not turn the quality of the object into a quality of the experience. In this situation, however, a certain problem arises. There is definitely something it is like to taste coffee, just as there is an experiential difference between tasting wine and water. When one asks for this quality and for this qualitative difference, it seems hard to point to anything beside the taste of coffee, wine, or water, though this is not what we are looking for. Should we consequently conclude that there is, in fact, nothing in the tasting of the lemon apart from the taste of the lemon itself?

Recently a number of philosophers have defended what might be called an *intentionalistic* interpretation of phenomenal qualities. The point of departure has been the observation that it can often be quite difficult to distinguish a description of certain objects from a description of an experience

of these very same objects. Back in 1903, G. E. Moore called attention to this fact and dubbed it the peculiar *diaphanous* quality of experience: When you try to focus your attention on the intrinsic features of experience, you always seem to end up attending to that *of* which it is an experience. As Tye argues, the lesson of this transparency is that “*phenomenology ain’t in the head*” (Tye 1995, 151). To discover what it is like, you must look at what is being intentionally represented. Thus, as the argument goes, experiences do not have intrinsic and nonintentional qualities of their own; rather, the qualitative character of experience consists entirely, as Dretske writes, in the qualitative properties objects are experienced as having (Dretske 1995, 1). In other words, the phenomenal qualities are qualities of that which is represented. Differences in what it is like are, in fact, intentional differences. An experience of a red apple is subjectively distinct from an experience of a yellow sunflower in virtue of the fact that different kinds of objects are represented. Experiences acquire their phenomenal character simply by representing the outside world. Consequently, all phenomenal qualities are intentional.

Dretske’s and Tye’s intentionalistic interpretation of phenomenal qualities bears a certain resemblance to views found in phenomenology. Phenomenologists would not interpret phenomenal experience as some kind of internal movie screen that confronts us with mental representations. Rather, we are “*zunächst und zumeist*” directed at real, existing objects. The so-called qualitative character of experience, the taste of a lemon, the smell of coffee, the coldness of an ice cube—these are not at all qualities belonging to some spurious mental objects, but qualities of the presented objects. Instead of saying that we experience *representations*, it would be better to say that our experiences are *presentational*, that they *present* the world as having certain features (see also Sartre 1943, 26–28, 363).⁶

Both Tye and Dretske explicitly criticize the attempt to draw a sharp distinction between the intentional or (re)presentational aspects of our mental lives and their phenomenal, subjective, or felt aspects. They deny the existence of epiphenomenal qualia and relocate the phenomenal from the “inside” to the “outside.” To repeat the earlier question, does this justify the claim that there is nothing in the tasting of the lemon apart from the taste of the lemon itself?

I think such a conclusion would be overhasty, since it fails to realize that there are two sides to the question of “what it is like.” In *Ideen I*, Husserl

distinguished between the intentional object in “the how of its determinations” (*im Wie seiner Bestimmtheiten*) and in “the how of its givenness” (*im Wie seiner Gegebenheitsweisen*) (Hua 3/303–304). Although this distinction is introduced as one that falls within the noematic domain, rather than a distinction between the noetic and the noematic domain, it nevertheless points us in the right direction. There is a difference between asking about the property the object is experienced as having (what does the object seem like to the perceiver) and asking about the property of the experience of the object (what does the perceiving feel like to the perceiver). Both questions pertain to the phenomenal dimension, but whereas the first question concerns a worldly property, the second concerns an experiential property.⁷ Contrary to what both Dretske and Tye are claiming, we consequently need to distinguish between (1) what the object is like for the subject and (2) what the experience of the object is like for the subject (see also Carruthers 1998; McIntyre 1999). Insisting on this distinction, however, is not enough; the tricky part is then to respect the lesson of transparency and avoid misconstruing the experiential properties as if they belong to some kind of mental objects. It is not the case that worldly properties, such as blue or sweet, are matched one to one by experiential doublets of an ineffable nature, let us call them *blue or *sweet, or that both kinds of properties are present in ordinary perception. How, then, is the distinction to be phenomenologically redeemed?

We are never conscious of an object *simpliciter*, but always of the object as appearing in a certain way; as judged, seen, described, feared, remembered, smelled, anticipated, tasted, and so on. We cannot be conscious of an object (a tasted lemon, a smelt rose, a seen table, a touched piece of silk) unless we are aware of the experience through which this object is made to appear (the tasting, smelling, seeing, touching). This is not to say that our access to, say, the lemon is *indirect*, or that it is mediated, contaminated, or blocked by our awareness of the experience; the given experience is not itself an object on a par with the lemon, but instead constitutes the access to the appearing lemon. The object is given through the experience; if there is no awareness of the experience, the object does not appear at all. If we lose consciousness, we, or more precisely our bodies, will remain causally connected to a number of different objects, but none of these objects will appear to us. In short, the red cherry is present for me, through my seeing it. Experiences are not objects, but rather, they provide us with access to objects; I attend

to the objects through the experiences. Experiential properties are not properties like red or bitter; they are properties pertaining to these various types of access. These accesses can take different forms; the same object, with the exact same worldly properties, can present itself in a variety of manners. It can be given as perceived, imagined, or recollected, and so on.

The moment we are dealing with manifestation or appearance we are faced with the phenomenal dimension. In fact, “what it is like” is exactly a question of how something appears to me, that is, it is a question of how it is given to and experienced by me. When I imagine a unicorn, desire an ice cream, anticipate a holiday, or reflect upon an economic crisis, all of these experiences bring me into the presence of different intentional objects. What this means is that not only am I phenomenally acquainted with a series of worldly properties such as blue, sweet, or heavy, but also the object is there *for me* in different modes of givenness (as imagined, perceived, recollected, anticipated, etc).

Whereas the object of John’s perception, along with all its properties, is intersubjectively accessible in the sense that, in principle, it can be given to others in the same way that it is given to John, it is different with John’s perceptual experience. Whereas John and Mary can both perceive the exact same cherry, each of them have his or her own distinct perception of it and can share these just as little as Mary can share John’s bodily pain. Mary might certainly realize that John is in pain, she may even sympathize with John; but she cannot actually feel John’s pain in the same way John does. It is here customary to speak of an epistemic asymmetry and say that Mary has no access to the *first-personal givenness* of John’s experience.

This first-personal givenness of experiential phenomena is not something incidental to their being, a mere varnish that the experiences could lack without ceasing to be experiences. On the contrary, this first-personal givenness makes the experiences *subjective*. To put it another way, their first-personal givenness entails a built-in self-reference, a primitive experiential self-referentiality.

In contrast to the redness of the tomato or the bitterness of the tea, both of which are worldly properties, the first-personal givenness of the perception of the redness or bitterness is not a worldly property, but an experiential property. When asked to specify “what the experience of the object is like for the subject,” this first-personal givenness is precisely one of the features to mention. In short, the experiential dimension does not have to do

with the existence of ineffable qualia; it has to do with the dimension of first-personal experiencing.

To reiterate: the “what it is like” question has two sides to it: “what is the object like for the subject” and “what is the experience of the object like for the subject.” Although these two sides can be distinguished conceptually, they cannot be separated. It is not as if the two sides or aspects of phenomenal experience can be detached and encountered in isolation from one other. When I touch the cold surface of a refrigerator, is the sensation of coldness that I then feel a property of the experienced object or a property of the experience of the object? The correct answer is that the sensory experience contains two dimensions, namely one of the *sensing* and one of the *sensed*, and that we can focus on either. Phenomenology pays attention to the givenness of the object, but it does not simply focus on the object exactly as it is given; it also focuses on the subjective side of consciousness, thereby illuminating our subjective accomplishments and the intentionality that is at play in order for the object to appear as it does. When we investigate appearing objects, we also disclose ourselves as datives of manifestation, as those to whom objects appear.

To put it differently, when speaking of a first-person perspective, or of a dimension of first-personal experiencing, it would be a mistake to argue that this is something that exclusively concerns the type of access a given subject has to his or her own experiences. Access to objects in the common world is independent of a first-person perspective, precisely in that it involves a third-person perspective. This line of thought will not do; obviously, I can be directed at intersubjectively accessible objects, and although my access to these objects is of the very same kind available to other persons, this does not imply that there is no first-person perspective involved. Rather, intersubjectively accessible objects are intersubjectively accessible precisely insofar as they can be accessed directly from each first-person perspective. They thereby differ from experiences, which are accessible in a unique way from the very same first-person perspective they, themselves, help constitute. Phrased another way, every givenness, be it the givenness of mental states or the givenness of physical objects, involves a first-person perspective. There is no pure third-person perspective, just as there is no view from nowhere. To believe in the existence of such a pure third-person perspective is to succumb to an objectivist illusion. This is, of course, not to say that there is no third-person perspective, but merely that such a perspective is exactly a

perspective from somewhere. It is a view that *we* can adopt on the world. It is a perspective founded upon a first-person perspective, or to be more precise, it emerges out of the encounter between at least two first-person perspectives; that is, it involves intersubjectivity.⁸

Let me summarize the line of argumentation. The phenomenal dimension covers both the domains of (1) what the object is like for the subject, and (2) what the experience of the object is like for the subject. Both the worldly properties of the appearing object and the experiential properties of the modes of givenness are part of the phenomenal dimension. They are not to be separated, but neither are they to be confused.

Mineness and Selfhood

Although the various modes of givenness (perceptual, imaginative, recollective, etc.) differ in their experiential properties, they also share certain features. One common feature is the quality of *mineness*, that is, the fact that the experiences are characterized by a first-personal givenness that immediately reveals them as one's own. When I (in nonpathological standard cases) am aware of an occurrent pain, perception, or thought from the first-person perspective, the experience in question is given immediately, noninferentially and noncriterially as *mine*. If I feel hunger or see a sunrise, I cannot be in doubt or be mistaken about who the subject of that experience is, and it is nonsensical to ask whether I am sure that I am the one who feels the hunger. Whether a certain experience is experienced as mine or not, however, depends not on something apart from the experience, but precisely on the givenness of the experience. If the experience is given in a first-personal mode of presentation, it is experienced as *my* experience, otherwise not (see James 1890, I/226–227). Obviously, this form of egocentricity must be distinguished from any explicit I-consciousness. I am not (yet) confronted with a thematic or explicit awareness of the experience as being owned by or belonging to myself. The mineness is not something attended to; it simply figures as a subtle background presence. Nevertheless, the particular first-personal givenness of the experience makes it mine and distinguishes it for me from whatever experiences others might have (Klawonn 1991, 5, 141–142; Hua 8/175; Hua 13/28, 56, 307). As Husserl put it in a manuscript now published in *Zur Phänomenologie der Intersubjektivität II*:

What is most originally mine is my life, my “consciousness,” my “I do and suffer,” whose being consist in being originally pre-given to me qua functioning I, i.e., in the

mode of originality, in being experientially and intuitively accessible as itself. (Hua 14/429)

It could be argued that it is misleading to suggest that experiences can be given in more than one way. Either an experience is given from a first-person perspective, or it is not given at all. However, I think this objection is mistaken. It is correct that experiences must always be given from a first-person perspective, for otherwise they would not be experiences; this does not, however, prevent them from being given from a second-person perspective as well. Let us assume that I get into a car accident, and that I am being scolded by the driver whose car I have just damaged. That the driver is angry is not something I establish by way of a hypothesis; it is something I experience. That I experience the anger of the driver does not imply that the experience is infallible (perhaps the driver is actually happy about the accident, since he can now finally get a new car, but he does not show his real feelings), nor does it imply that the driver's anger is given to me in the same way that it is given to the driver himself. The anger is exactly given from a second-person perspective for me. To deny the possibility of this is to face the threat of solipsism. (I will return to these issues in detail in chapters 6 and 7.)

Contrary to what some of the self-skeptics are claiming, one does not need to conceive of the self as something standing apart from or above experiences, nor does one need to conceive of the relation between self and experience as an external relation of ownership. It is also possible to identify this pre-reflective sense of *mineness* with a minimal, or core, sense of self. To again quote Henry, the most basic sense of self is the one constituted by the very self-givenness of experience (Henry 1963, 581; 1965, 53). In other words, the idea is to link an experiential sense of self to the particular first-personal givenness that characterizes our experiential life; it is this first-personal givenness that constitutes the *mineness* or *ipseity* of experience. Thus, the self is not something that stands opposed to the stream of consciousness, but is, rather, immersed in conscious life; it is an integral part of its structure.

One advantage of this view is that, incidentally, it makes it clear that self-awareness is not to be understood as an awareness of an isolated, worldless self, nor is the self located and hidden in the head. To be self-aware is not to interrupt the experiential interaction with the world in order to turn the gaze inward; on the contrary, self-awareness is always the self-awareness of

a world-immersed self. The self is present to itself precisely and indeed only when worldly engaged. It would consequently be a decisive mistake to interpret the present notion of a core, or minimal, self as a Cartesian-style mental residuum, that is, as some kind of self-enclosed and self-sufficient interiority.⁹ The phenomenological notion of an experiential self is fully compatible with a strong emphasis on the fundamental intentionality, or being-in-the-world, of subjectivity.¹⁰ It is no coincidence that even Heidegger employed such a minimal notion of self (see chapter 4 above).

An effective way to capture this basic point is to replace the traditional phrase “subject of experience” with the phrase “subjectivity of experience.” Whereas the first phrasing might suggest that the self is something that exists apart from, or above, the experience and, for that reason, is something that might be encountered in separation from the experience or even something the experience may occasionally lack, the second phrasing excludes these types of misunderstanding. It makes no sense to say that the subjectivity of the experience is something that can be detached or isolated from the experience, or to say that it is something the experience can lack. To stress the subjectivity of experience is not an empty gesture; it is to insist on the basic *ipseity* of the experiential phenomena.

In order to have a self-experience, it is, consequently, not necessary to apprehend a special self-object, it is not necessary to have a special experience of self alongside yet different from other experiences; rather what is required is simply an episode of pre-reflective self-awareness. What is needed is an acquaintance with the experience in its first-personal mode of presentation.¹¹ Thus, from Hume’s famous passage in *A Treatise on Human Nature*, wherein he declared that he could not find a self when he investigated his own mental life, but only particular perceptions or feelings, it would be natural to conclude that he had overlooked something in his analysis, namely the specific givenness of his own experiences. He was looking for the self in the wrong place, so to speak. As Evans states, “from the fact that the self is not an object of experience it does not follow that it is non-experiential” (Evans 1970, 145).

One possible countermove would be to insist that first-personal access to individual mental states is not sufficient for self-experience. Self-experience involves some reference to self, but one can be aware of a mental happening from the first-person perspective and fail to realize that the happening occurs to oneself. As already mentioned, the non-egological theory would

claim that our experiences are normally impersonal, in the sense that they do not include any reference, not even an implicit reference, to oneself as the subject of the experience. Thus, even if one has to concede that two persons who have two simultaneous and qualitatively identical experiences would still have two distinct experiences, the fact that they are distinct is not due to each of the experiences having a different *subject*. To quote Parfit, “one of these experiences is *this* experience, occurring in *this* particular mental life, and the other is *that* experience, occurring in *that* other particular mental life” (Parfit 1987, 517).

However, is it true that the primary difference between my perception and my friend’s perception is that my perception is *this* one and his *that* one? Is this not, as Klawonn has argued, a parasitic and derived characterization? Is it not, rather, the case that an experience is *this* one exactly because it is *mine*, that is, given in a *first-personal mode of presentation*, whereas the other’s experience is not given in a first-personal mode for *me*, and precisely therefore, is no part of *my* mental life (Klawonn 1991, 28–29)?

For the same reason, the validity of Sartre’s revision of the cogito in *La transcendance de l’ego* must also be questioned. It does not seem adequate to render the cogito as “there is a perception of a chair,” nor for that matter as “somebody perceives a chair” (Merleau-Ponty 1945, 249, 277), since both formulations overlook one significant detail. If the reader and I look at the same chair, these two perceptions of the chair might very well be impersonal or anonymous in the sense of lacking any *thematic* self-reference. In fact, on the pre-reflective level there is no explicit awareness of the experience being mine. The two perceptions, however, are definitely *not* anonymous in the sense of being undifferentiated and indistinguishable, regardless of whether this is taken to imply strict numerical identity or mere qualitative identity. On the contrary, the moment we take the first-person perspective seriously, it is obvious that there is a vital difference between the two perceptions; only one of them is given in a first-personal mode of presentation for me.

It might be objected that the current proposal makes the thesis concerning the experiential reality of the self acceptable but also quite trivial. However, as long as the thesis is routinely denied by advocates of the different impersonality theses, that is, by adherents to the no-ownership view or the non-egological account, it does not seem superfluous to make the point. Moreover, as both Wittgenstein and Heidegger remarked, one of the tasks of philosophy is exactly to call attention to and elucidate those

fundamental aspects that are so familiar to us, so taken for granted, that we often fail to realize their true significance and even deny their existence.

Another countermove would be to follow Metzinger and argue that it is a phenomenological fallacy to conclude to the literal properties of the self from the content and structure of phenomenal self-experience. In his view, a phenomenological account of selfhood has no metaphysical impact. Our self-experience, our primitive, pre-reflective feeling of conscious selfhood, is never truthful in that it does not correspond to any single entity inside or outside of the self-representing system (Metzinger 2003a, 565). Why, however, should the reality of the self depend on whether it faithfully mirrors either subpersonal mechanisms or external (mind-independent) entities? If we were to wholeheartedly endorse such a restrictive metaphysical principle, we would declare illusory most of the world we live in and know and care about. Why not rather insist that the self is real if it has experiential reality and that the validity of our account of the self is to be measured by its ability to be faithful to experience, by its ability to capture and articulate (invariant) experiential structures?

As Strawson has recently argued, if we wish to answer the metaphysical question concerning whether or not the self is real, we will first need to know what a self is supposed to be. In order to establish this, our best chance will be to look at self-experience, since self-experience is what gives rise to the question in the first place by giving us a vivid sense that there is something like a self. Thus, as Strawson readily concedes, the metaphysical investigation of the self is subordinate to the phenomenological investigation. The latter places constraints on the former. Nothing can count as a self unless it possesses those properties attributed to the self by some genuine form of self-experience (Strawson 2000, 40).¹²

Although the minimal notion of self might seem overly inclusive, it does, in fact, exclude several nonexperiential candidates. Some have argued that no organism can survive or act without being able to distinguish between self and nonself (see Dennett 1991, 174, 414). According to this proposal, however, whether to ascribe selfhood to, say, bacteria would depend on whether bacteria can be said to possess phenomenal experiences. It would not be sufficient that the bacterium was able to nonconsciously differentiate itself from the environment.

Let me return to the three notions of self that I distinguished in the beginning of the chapter. How does the self as experiential dimension stand to the

self as a pure identity-pole, and to the self as a narrative construction? As already mentioned, it can replace the first and supplement the second. If we start with the latter, that is, with the relation between the self as experiential dimension and the self as narrative construction, the case is relatively straightforward. The two notions of self are so different that they can easily complement each other. In fact, on closer consideration it should be clear that the notion of self introduced by the narrative model is not only far more complex than, but also phenomenologically and ontologically dependent, on the experiential self. Only a being with a first-person perspective could make sense of the ancient dictum “know thyself”; only a being with a first-person perspective could consider her own aims, ideals, and aspirations *as* her own and tell a story about them. To avoid unnecessary confusion, one might opt for a terminological differentiation. When dealing with the experiential self, one might retain the term “self,” since we are dealing precisely with a primitive form of self-givenness or self-referentiality. By contrast, it may have been better to speak not of the self, but of the *person* as a narrative construction. After all, what is being addressed by this model is the nature of my personal character or personality, a personality that evolves through time and is shaped by the values I endorse and by my moral and intellectual convictions and decisions.

The fact that narrative personhood presupposes experiential selfhood (but not vice versa) does not diminish the significance of the former. Owing to the first-personal givenness of experience, our experiential life is inherently individuated. It is, however, a purely formal kind of individuation. In contrast, a more concrete kind of individuality manifests itself in my personal history, in my moral and intellectual convictions and decisions. It is through such acts that I define myself; they have a character-shaping effect. I remain the same as long as I adhere to my convictions; when they change, *I* change (Hart 1992, 52–54). Ideals can be identity defining; acting against one’s ideals can mean the disintegration (in the sense of a dis-integrity) of one’s wholeness as a person (see Moland 2004). These convictions and endorsed values are all intrinsically social; it is no coincidence that Husserl distinguished the subject taken in its bare formality from the personalized subject and claimed that the origin of personality must be located in the social dimension (see p. 95 above). I am not simply a pure and formal subject of experience, but also a person with abilities, dispositions, habits, interests, character traits, and convictions, and to focus exclusively on the first is to

engage in an abstraction (Hua 9/210). Given the right conditions and circumstances, the self acquires a personalizing self-apprehension, that is, it develops both into a person and as a person. This development depends heavily on social interaction. To exist as a person is to exist socialized into a communal horizon, where one's bearing to oneself is appropriated from the others. To put it differently, the self is fully developed only when personalized intersubjectively; I become a person exclusively through my life with others in our communal world (Hua 4/265; 14/170–171).¹³

Usually, the self under consideration is already personalized or at least in the process of developing into a full-blown person. Although a narrow focus on the experiential core self might, therefore, be said to involve an abstraction, there is no reason to question its reality; it is not a *mere* abstraction. Not only does it play a foundational role, but, as we will shortly see, the notion of an experiential core self has also found resonance in empirical science. There are limit situations where this minimal self might, arguably, be encountered in its purity.

Whereas the relation between the self as experiential dimension and the person as narrative construction is relatively straightforward, the situation is slightly more complicated when it comes to the relation between the self as an experiential phenomenon and a pure identity-pole. One advantage of the view just outlined is that it may be capable of accounting for some of the features normally associated with the pure identity-pole model, particularly its ability to account for the identity of the self through time without actually having to posit the self as a separate entity over and above the stream of consciousness. Although the phenomenological account sketched above is intended mainly as an account of the conditions of selfhood (what properties must x have in order to count as a self), it could, ultimately, also have something to say concerning the conditions of persistency (the conditions required for x to remain the same from t_1 to t_2).

To show why this is so, let us briefly return to Husserl. As already mentioned, Husserl operated with a whole range of various notions of self. He not only spoke of it in terms of the first-personal givenness of an experience, he also introduced the notion of an act-transcendent ego. This was explained as the self considered as an identity-pole, a principle of focus, shared by all experiences belonging to the same stream of consciousness (Hua 13/248; 9/207; 4/277). What is new in this characterization is obviously the attempt to *differentiate* between the self and the experiences. Such a differentiation

seems warranted the moment we pass beyond a narrow investigation of the self-givenness of a single experience and consider the case of a plurality of experiences. After all, not only is it possible to be aware of one's own toothache, it is also possible to be aware of oneself as the common subject of a manifold of simultaneous experiences. The same holds true wherein one can be self-aware across temporal distance and recall a past experience as one's own. In these latter cases, it is necessary to distinguish the self from any single experience, as the self can preserve its identity whereas experiences arise and perish in the stream of consciousness, replacing one other in a permanent flux (Hua 4/98; 17/363). Husserl then went on to emphasize, however, that although the ego must be distinguished from the experiences in which it lives and functions, it cannot exist in any way independently of them. It is a transcendence, but in Husserl's now famous phrase, it is *a transcendence in the immanence* (Hua 3/123–124, 179; 4/99–100; 13/246; 14/43).

If we relate the question concerning the act-transcendent self to the discussion of self-awareness, the obvious question to ask is: When does my self-awareness contain a reference to such an overarching identity? I think a plausible answer would be that the self-givenness of a single experience is a necessary, but not a sufficient, condition for this type of self-awareness to occur. The latter entails more than a simple and immediate self-awareness; it also entails a difference or distance that is bridged, that is, it involves a *synthesis*. This is so because the self cannot be given as an act-transcendent identity in a *single* experience (Kern 1989, 60–62; 1975, 66; Marbach 1974, 110, 112). It is only by comparing several experiences that we can encounter something that retains its identity through changing experiences (Hua 13/318; 4/208; 11/309–310).

What has all of this to do with the earlier discussion of the self as an experiential dimension? At first glance, the answer might be “nothing.” After all, whereas the current point seems to be that one should distinguish self and experience, the earlier discussion attempted to abolish this difference. Such a response would be premature, however; as it was already pointed out in the discussion of Husserl's theory of inner time-consciousness from chapter 3, it is quite legitimate to insist on the *difference* between our singular and transitory experiences and the abiding dimension of first-personal experiencing (see Hua 23/326; 14/46). In other words, the moment we expand the focus to include more than a single experience, it becomes not only

legitimate but also highly appropriate to distinguish the strict singularity of the field of first-personal givenness from the plurality of changing experiences. Although the act-transcendent identity of the self is revealed only in acts of synthesis, it does not arise out of the blue, but is grounded in the pervasive dimension of first-personal experiencing. Whereas we live through a number of different experiences, the dimension of first-personal experiencing remains the same. In short, although the self, as an experiential dimension, does not exist in separation from the experiences, and is identified by the very first-personal givenness of the experiences, it may still be described as the *invariant* dimension of first-personal givenness throughout the multitude of changing experiences.

The pertinence of this account for the problem concerning the diachronic persistency of the self should be obvious. To determine whether a past experience is mine, I do not first need to assure myself of the uninterrupted, temporal continuity between my present recollection and the past experience. If I have first-personal access to the past experience, it is automatically given as *my* past experience. There is more to episodic memory than the simple retrieval of information about the past. The subjective experience of remembering involves the conviction that the remembered episode was once experienced by me. Obviously, this is not to say that episodic memory is infallible (I might have false beliefs about myself), but only that it is not subject to the error of misidentification (see Campbell 1994, 98–99). To question the unity of mind by pointing to alleged interruptions in the stream of consciousness (dreamless sleep, coma, etc.) is consequently pointless, since one thereby makes the erroneous assumption that it is the *continuity* and *contiguity* between two experiences that makes them belong to the same self, rather than their shared mineness, or their shared manner of givenness.¹⁴

V Empirical Implications

The narrative concept of self has found resonance not only in different philosophical traditions (Ricoeur, MacIntyre, Dennett), but also in a variety of empirical disciplines, such as developmental psychology, neuroscience, and psychiatry (see Gallagher 2000a). The same holds true for the phenomenological concept of an experiential core self. In the following, I will briefly discuss some relevant psycho- and neuropathological findings.

The study of pathological phenomena might not only serve as a demonstration of the empirical relevance of the phenomenological analysis of self; on its own it might enrich our understanding of the nature of selfhood. Pathological cases can function as a heuristic device that shocks one into an awareness of what is normally taken for granted. They may be employed as a means of gaining distance from the familiar, in order better to explicate it. To put it another way, core features of subjectivity, including fundamental aspects of self-experience, can be sharply illuminated through a study of their pathological distortions. These distortions may reveal, through their very absence, aspects of normal existence that frequently remain unnoticed. In using pathology as a contrast, it will also become clear that normality cannot be taken for granted; it is, itself, an achievement.

The Case of Schizophrenia

In his *Allgemeine Psychopathology*, Jaspers famously wrote that schizophrenia is characterized by its un-understandability or incomprehensibility. What he meant by this was that schizophrenic symptoms are so strange and bizarre that they remain inaccessible to empathy and meaningful reconstruction. More specifically, Jaspers distinguished between *static un-understandability*, which refers to their inaccessibility to empathy, and *genetic un-understandability*, which refers to the impossibility of understanding the development or emergence of psychotic symptoms (Jaspers 1959, 24, 251, 483–486). By implication, there is not much to be won by paying close attention to the first-person accounts of schizophrenic patients. Their disturbed self-descriptions do not present us with a key to an understanding of schizophrenia; rather they must be seen as senseless ravings or morbid eruptions of a malfunctioning brain.

Jaspers's claims were based on a study of the chronic stages of schizophrenia. In recent years, however, Parnas and Sass have argued that a study of the advanced stages of the illness is of limited value if one wishes to understand the core features of the illness. This is so not only because of the apparent incomprehensibility of the symptoms, but also because the advanced stages confront one with the results of a long-standing interaction between multiple factors, such as the effects of medication, social isolation, and stress. This complexity inevitably makes it much harder to isolate the primary pathogenetic factors.

Parnas and Sass have suggested that it would make more sense to examine the highly informative antecedent stages of schizophrenia. One should study the early symptoms detectable in the first (initial) prodromal stage, that is, the stage immediately preceding and leading to the onset of a schizophrenic psychosis, as well as the abnormalities present in the even earlier so-called premorbid phase, since these symptoms might, in a much sharper manner, express the essential core of the illness. This change of focus may, incidentally, also be of direct benefit to the patient, since it allows for earlier detection and therapeutic intervention.

What will one find if one investigates these prepsychotic stages? According to Parnas and Sass, one will find a diverse assortment of self-disorders involving a variety of alterations and transformations to the very basic sense of self, including disturbances of the first-person perspective, the first-personal givenness of experience, and the dimension of mineness. Parnas and Sass argue that these self-disorders may even be ascribed a generating, pathogenic role. They antecede, underlie, and shape the emergence of later and psychotic pathology and may thus unify what, from a purely descriptive psychiatric standpoint, may seem to be unrelated or even antithetical syndromes and symptoms (Sass and Parnas 2003, 428). To put it differently, a focus on the psychopathology of chronic schizophrenia might present one with such a diversity of apparently unconnected symptoms that it thereby raises doubts about the unity of the diagnostic category. A focus on the earlier stages of the illness will reveal the sought-after unity. The heterogeneity of the symptoms, both negative and positive, encountered in advanced stages is merely ostensible; at its root, schizophrenia is a disorder of the self (Sass and Parnas 2003, 427–428). The concept of self in use, however, is not the narrative concept, but the phenomenological concept of *ipseity*, the concept of an experiential core self. This commitment to a phenomenological understanding of selfhood is articulated by Parnas and Sass themselves (see Sass 2000, 152; Parnas 2003, 219).

Although the most recent versions of the psychiatric diagnostic systems (DSM-IV and ICD-10) do not include a reference to the self, varieties of self-disorders have always figured, at least implicitly, as an important component in the clinical picture of schizophrenia. As early as 1913, the concept of “*Ichstörungen*” (self-disturbances) was introduced by Jaspers. One year later, Berze proposed that a basic transformation of self-consciousness was at the root of schizophrenia. As Parnas and Sass point out, however, the

most detailed analysis of schizophrenic self-disorders is to be found in phenomenologically oriented psychiatry (Minkowski 1927; Conrad 1958; Laing 1960; Blankenburg 1971; Tatossian 1979). As Minkowski wrote:

The madness . . . does not originate in the disorders of judgment, perception or will, but in a disturbance of the innermost structure of the self. (Minkowski 1997, 114)

Patients will frequently complain about having lost something fundamental. The phrasing may range from “I don’t feel myself,” “I am not myself,” “I have lost contact with myself,” to “My I is disappearing for me” or “I am turning inhuman.” The patients may sense an inner void, a lack of an undefinable “inner nucleus,” a diminished sense of presence, or an increased distance from the world, and an incipient fragmentation of meaning. As described by an eighteen-year old patient:

I am more and more losing contact with my environment and with myself. Instead of taking an interest in what goes on and caring about what happens with my illness, I am all the time losing my emotional contact with everything including myself. What remains is only an abstract knowledge of what goes on around me and of the internal happenings in myself. (Quoted in Frith and Johnstone 2003, 2)

Parnas has argued that all these complaints point to a diminished *ipseity*, where the sense of self no longer automatically saturates the experience (Parnas 2003; Parnas, Bovet, and Zahavi 2002). We are faced with an experiential disturbance on a pre-reflective level that is far more basic than the kind of feelings of inferiority, insecurity, and unstable identity that we find in personality disorders outside the schizophrenic spectrum.

Some patients are able to articulate these subtle disturbances better than others. One of Parnas’s patients reported that the feeling that his experiences were his own always came with a split-second delay; another that it was as if his self was somehow displaced a few centimeters back. A third explained that he felt an indescribable inner change that prevented him from leading a normal life. He was troubled by a very distressing feeling of not being really present or even fully alive. This experience of distance or detachment was accompanied by a tendency to observe or monitor his inner life. He summarized his affliction by saying that his first-personal life was lost and replaced by a third-person perspective (Parnas 2003, 223).

Contrary to a traditional view according to which phenomenology is a descriptive rather than an explanatory enterprise—its task is to describe and define experiential structures rather than to account for the causal

mechanisms that bring them about—Sass and Parnas have also argued for the explanatory relevance of phenomenological psychiatry (Sass and Parnas 2005). Some grounds for this claim are as follows:

- First, phenomenological descriptions must act as constraining conditions for any neuroscientific explanation. That is, the neuroscientific explanation must be compatible with the facts about the subjective dimension. After all, the subjective dimension is precisely the explicandum that a satisfactory causal explanation is supposed to account for. The phenomenological investigation can be seen as an unfolding of the various facets of conscious life in order to gain a richer insight into its lived texture and internal structure. In this sense, a descriptive account remains indispensable for any causal account.
- Second, a phenomenological understanding of the fundamental structures of self-experience and self-disturbance may allow one to make sense of seemingly incomprehensible actions and beliefs. “One may, e.g., come to see how the person’s actions or beliefs are in some respect *inspired* or *justified* by the kinds of experiences the person is having, or one may see these actions or beliefs in the light of general features of the person’s experience of time, space, causality, or selfhood” (Sass and Parnas 2005). To put it differently, and contrary to Jaspers’s claim, many so-called bizarre delusions pathognomonic of schizophrenia, are, in fact, psychologically comprehensible. A phenomenological approach might allow one to understand these bizarre experiences as arising from, and in a sense, expressing aspects of the profoundly altered form of consciousness characterizing schizophrenia (Sass and Parnas 2005). Following is Sass and Parnas’s attempt to account for the symptom of schizophrenic perplexity:

normal ipseity . . . provides a point of orientation: it is what grounds human motivation and organizes our experiential world in accordance with needs and wishes, thereby giving objects their “affordances,” their significance for us as obstacles, tools, objects of desire, and the like. In the absence of this vital yet implicit self-affection, and the lines of orientation it establishes, the structured nature of the worlds of both thought and perception will be altered or even dissolved. For then there can no longer be any clear differentiation of means from goal; any reason for certain objects to show up in the focus of awareness while others recede; or any reason for attention to be directed outward toward the world rather than inward toward one’s own body or processes of thinking. Without normal self-affection, the world will be stripped of all the affordances and vectors of concern by which the fabric of normal, common-sense reality is knitted together into an organized and meaningful

whole. This, we believe, is the basis of the distinctively schizophrenic “perplexity” (Ratlosigkeit) described in classic German psychopathology. (Sass and Parnas 2005)

- Finally, and related to the previous point, experiential disturbances should not be regarded as mere epiphenomena or something that can be discounted in the scientific search for the core of schizophrenia. Rather, the experiential disturbances found at the prodromal stage can even partially explain the subsequent progression of the disease. Once the field of experience is disturbed, quite possibly as a result of neurobiological abnormalities, new types of experience arise in reaction to the changes in question. In this sense, subjective experience can play an important causal role in the progressive experiential transformation that we encounter in the development of schizophrenia (Sass and Parnas 2005). As a case in point, one might mention the fact that many schizophrenic patients engage in compulsive self-monitoring. This is what Blankenburg called a convulsive reflection or a reflective spasm (*Reflexionskrampf*) and is an aspect of what Sass has more recently dubbed *hyperreflexivity*. According to Sass, this hyperreflexivity manifests itself in a variety of ways, depending on whether it occurs (1) as a facet of the basic defect itself, (2) as a consequence of the more basic disturbance, or (3) as defensive compensation for the more basic disturbance (Sass 2000, 153). At first, hyperreflexivity is not a volitional kind of self-consciousness; it occurs in a more or less automatic manner and has the effect of disrupting experiences and actions that would normally remain in the background of awareness. Thus, the normal stream of consciousness is interrupted by sensations, feelings, or thoughts that suddenly become the focus of attention with an objectlike quality (*basal hyperreflexivity*). These primary disruptions and disturbances then attract further attention and thereby elicit a process of self-scrutiny and self-objectification, or reflective turning-inward of the mind (*consequential hyperreflexivity*). Finally, such patients might voluntarily engage in reflective self-monitoring in an attempt to compensate for their diminished self-presence (*compensatory hyperreflexivity*):

I forgot myself at the Ice Carnival the other night. I was so absorbed in looking at it that I forgot what time it was and who and where I was. When I suddenly realized I hadn't been thinking about myself I was frightened to death. The unreality feeling came. I must never forget myself for a single minute. (Patient quoted in Laing 1960, 109)

Needless to say, rather than restoring what has been lost, such excessive self-monitoring may only exacerbate the problem by further objectifying, alienating, and dividing the experiential life (Sass 1994, 12, 38, 91, 95). A patient, studied and described in detail by Sass, offered the following insights:

“My downfall was insight,” he explained, “. . . too much insight can be very dangerous, because you can tear your mind apart.” “Well, look at the word ‘analysis,’” he said on another occasion. “That means to break apart. When it turns in upon itself, the mind would rip itself apart.” Lawrence spoke of “doing six self-analyses simultaneously” and of how he needed to change his living environment often, because he knew that, once everything around him had been scrutinized, his mind would then turn inward and begin undoing itself, leading him eventually to the feeling of having no real mind at all: “Once I started destroying [my mind], I couldn’t stop.” (Sass 1992, 337–338)

Neurological Findings

Damasio argues in his recent book, *The Feeling of What Happens*, that a sense of self is an indispensable part of the conscious mind. As he writes: “If ‘self-consciousness’ is taken to mean ‘consciousness with a sense of self,’ then all human consciousness is necessarily covered by the term—there is just no other kind of consciousness as far as I can see” (Damasio 1999, 19). When I think thoughts, read a text, perceive a windowsill, a red book, or a steaming teacup, my mind is configured in such a manner that I automatically and relentlessly sense that I, rather than anyone else, am doing it. I sense that the objects I now perceive are being apprehended from my perspective and that the thoughts formed in my mind are mine and not anyone else’s. Thus, as Damasio puts it, there is a constant but quiet and subtle presence of self in my conscious life, a presence that never falters as long as *I* am conscious. Were it absent, there would no longer be a self (Damasio 1999, 7, 10, 127).

Consciousness is not a monolith, however, and Damasio finds it reasonable to distinguish a simple, foundational kind, which he calls *core consciousness*, from a more complex kind, which he calls *extended consciousness*. Core consciousness has a single level of organization and remains stable across the lifetime of the organism. It is not exclusively human and does not depend on conventional memory, working memory, reasoning, or language. In contrast, extended consciousness has several levels of organization. It evolves across the lifetime of the organism and depends on both conventional and working memory. It can be found in a basic form in some nonhumans, but attains its highest peak only in language-using humans.

According to Damasio, these two kinds of consciousness correspond to two kinds of self. He calls the sense of self that emerges in core consciousness *core self* and refers to the more elaborate sense of self provided by extended consciousness as *autobiographical self* (Damasio 1999, 16–17, 127).¹⁵

The relation between core consciousness and extended consciousness and thus, between core and autobiographical self, is foundational. Extended consciousness is built on the foundation of core consciousness. It presupposes a core, which it then extends by linking it to both the lived past and the anticipated future. From a developmental perspective, there are little more than simple states of core self in the beginning, but as experience accrues, memory grows and the autobiographical self can be deployed (Damasio 1999, 175).

When we speak of a persisting personal identity, we are referring to the autobiographical self. This self is based on a repository of memories that can be reactivated and, thereby, provide continuity to our lives (Damasio 1999, 217). By contrast, Damasio takes the core self to be a transient and ephemeral entity. It is generated anew, in a pulsative fashion, for every set of contents of which we are to be conscious. It also possesses a remarkable degree of structural invariance, as it is remade time and again in essentially the same form across a lifetime. It is this invariance that allows it to provide stability to the mind (Damasio 1999, 17, 126, 135, 173–176).

As Damasio points out, neuroscience, particularly neuropathology, provides empirical evidence in support of his distinction. The investigation of neurological diseases permits us to tease apart the layers and functions of consciousness:

The results of neurological disease validate the distinction between core consciousness and extended consciousness. The foundational kind of consciousness, core consciousness, is disrupted in akinetic mutisms, absence seizures, and epileptic automatism, persistent vegetative state, coma, deep sleep (dreamless), and deep anesthesia. In keeping with the foundational nature of core consciousness, when core consciousness fails, extended consciousness fails as well. On the other hand, when extended consciousness is disrupted, as exemplified by patients with profound disturbances of autobiographical memory, core consciousness remains intact. (Damasio 1999, 121–122).

This shows that neuropathology can reveal that impairments of extended consciousness allow core consciousness to remain intact, whereas impairments that begin at the level of core consciousness cause extended consciousness to collapse as well (Damasio 1999, 17). In illustration, Damasio presents data from a patient whose temporal lobes had both sustained major

damage from a case of encephalitis. This patient's memory was limited to a window of less than one minute; he was unable to learn any new facts and unable to recall many old facts. In fact, the recall of virtually any unique thing, individual, or event from his entire life was denied to him. Whereas his autobiographical memory had been reduced to a skeleton, and the autobiographical self that could be constructed at any moment was severely impoverished, he retained a core consciousness for the events and objects in the here and now and, thereby, also a core self (Damasio 1999, 115–119).

The Use and Misuse of Pathology

One of the customary methods of testing the validity of philosophical analyses has been to look for invalidating counterexamples. If none can be found, so much the better for the proposed thesis. This search has often been carried out by means of imagination. We don't necessarily have to come across (f)actual counterexamples; it is sufficient if we can imagine them. Thus, *imaginability* has often been taken as a mark of *possibility*: If something is imaginable, then it is, if not practically or physically possible, at least possible in principle, that is, conceptually or metaphysically possible. If this is the case, then the exceptions are relevant and should be taken into account when assessing the validity of the philosophical claims.

Much contemporary philosophy, particularly analytical philosophy of mind, abounds with thought experiments meant to test and challenge our habitual assumptions about the nature of consciousness, the mind–body relation, personal identity, and so on. Thus, one often comes across references to zombies, brain transplants, Twin Earths, and teletransporters. This way of doing philosophy has, to put it mildly, not been met with universal approval.¹⁶ One understandable reaction has been to ask whether it is legitimate to draw substantial philosophical conclusions from the mere fact that certain scenarios are imaginable. Is our imagination always trustworthy; does it always attest to metaphysical possibility, or might it occasionally reflect nothing but our own ignorance?

As Wilkes has pointed out, if thought experiments are to be of value, they must be performed with as much attention to detail and as many stringent constraints as real experiments conducted in the laboratory. One important requirement is that we are clear about the background conditions against which the experiment is set. In other words, we need to know exactly what is being altered and what remains the same when the imagined scenario is compared to the actual world. If there are too many variables, if too many

parameters are changed, we would not know which of them were responsible for the outcome, and it would, consequently, be impossible to draw any clear conclusion from the experiment (Wilkes 1988, 2, 6). Another obvious prerequisite is that we actually know something about the topic under discussion. Otherwise, we might easily end up in a situation where we believe that we have succeeded in imagining a possible state of affairs, yet, in reality, we have done nothing of the sort, as we will realize when we acquire more information and are able to think the scenario through more carefully.

To illustrate the importance of these requirements: If we ask somebody whether he can imagine a candle burning in a vacuum, or a gold bar floating on water, and the answer is yes, should we then conclude that there must be some possible world where gold bars have a different molecular weight while remaining gold bars, and where candles can burn despite a lack of oxygen, or should we rather conclude that the person has succeeded only in imagining something that superficially resembles gold bars and burning candles? It certainly seems necessary to distinguish between imagining something in the sense of having a loose set of fantasies and imagining it in the sense of thinking it through carefully. Surely only the latter is of any value if we wish to establish whether a certain scenario is possible or not. The lesson to learn is, undoubtedly, that the more ignorant we are, the easier it will seem to imagine something since “the obstructive facts are not there to obtrude” (Wilkes 1988, 31). What seems to be an imaginable possibility might, on closer examination, turn out to be an impossibility in disguise. If we wish to derive any interesting conclusions from our thought experiments, we must assure ourselves that we are not faced with such impossibilities. As Dennett puts it, “When philosophical fantasies become too outlandish—involving time machines, say, or duplicate universes or infinitely powerful deceiving demons—we may wisely decline to conclude *anything* from them. Our conviction that we understand the issues involved may be unreliable, an illusion produced by the vividness of the fantasy” (Dennett 1981, 230).

This criticism should not be misunderstood. Thinking about exceptional cases “is indispensable if we wish to avoid mistaking accidental regularities for regularities which reflect a deeper truth about the world” (Gendler 1999, 463). Yet since so many details must be attended to if a thought experiment is to be truly conclusive, it might, occasionally, be better to abandon fiction altogether and instead pay more attention to the startling facts found in the actual world. Real-life deviations can serve the same function as thought experiments. If we are looking for phenomena that can shake our ingrained

assumptions and force us to refine, revise, or even abandon our habitual way of thinking, all we have to do is to turn to psychopathology, along with neurology, developmental psychology, and ethnology; all of these disciplines present us with rich sources of challenging material. In other words, if we wish to test our assumptions about the unity of mind, the privacy of mental states, the nature of agency, or the role of emotions, far more may be learned from a close examination of pathological phenomena such as depersonalization, thought insertion, multiple personality disorder, cases of apraxia, or states of anhedonia than from thought experiments involving swapped brains or teletransporters. The former phenomena can also probe and test our concepts and intuitions and in a far more reliable way, since the background conditions are known to us. As they are real phenomena, they do not harbor any concealed impossibilities.

This said, a word of caution is appropriate. Pathological phenomena and other empirical findings are, of course, open to interpretation. Their interpretation usually depends on the framework within which one is operating. Thus, the theoretical impact of an empirical case is not necessarily something that is easily determined. One might agree with Metzinger that it is important not to underestimate the richness, complexity, and variety of conscious phenomena and that nonstandard cases of conscious self-experience can test the validity of a theory of self. He may, however, be overstating his point when he writes that “many classical theories of mind, from Descartes to Kant, will have to count as having been refuted, even after consideration of the very first example” of such pathological cases (Metzinger 2003a, 429). Contrary to what Metzinger suggests, it is rather doubtful that one will find many classical philosophers who subscribed to the thesis that unnoticed errors about the content of one’s own mind are logically impossible (Metzinger 2003a, 429, 431). Even if they had, it is by no means clear what type of conclusions one should draw from pathological cases. Are these cases mere anomalies? Are they the exceptions that prove the rule? Should they, rather, force us to abandon our habitual classification of behavior and experience with the realization that the normality that has been our point of departure has no priority, but is merely one variation among many? Does pathology reveal some hidden fundamental feature of normal experience, or does it rather reflect or manifest an abnormal mode or a compensatory attempt to deal with dysfunction (see Marcel 2003, 56)? Whatever the precise answer to these questions turns out to be, it does seem problematic

to simply draw unqualified conclusions about normal cases on the basis of pathology.

Metzinger spends considerable time discussing pathological cases, and although he repeatedly emphasizes how important it is to listen closely to the patients and take their phenomenology seriously (Metzinger 2003a, 446, 455), I also think he underestimates the difficulty of actually doing the latter. He frequently, and mistakenly, equates it with taking the patients' first-person assertions at face value. The danger of doing this comes to the fore in his analysis of both thought insertion and Cotard's syndrome.

According to Metzinger, the phenomenology of schizophrenia is so well known that it is superfluous to offer any explicit case study of it (Metzinger 2003a, 445). One prominent feature of schizophrenia is that it typically involves forms of alienated self-consciousness. In what is known as thought insertion, for example, the patient may have direct access to his or her own mental states but still experience them not only as being controlled or influenced by others, but as alien, as belonging to another. As one patient complained:

Thoughts are put into my mind like "Kill God." It is just like my mind working, but it isn't. They come from this chap, Chris. They are his thoughts. (Quoted in Frith 1992, 66)

Thus, as Metzinger puts it, schizophrenia confronts us with situations wherein patients experience introspectively alienated conscious thoughts for which they have no sense of agency or ownership. He takes this to demonstrate that the phenomenal quality of mineness is not a necessary precondition for conscious experience (Metzinger 2003a, 334, 382, 445–446).

As Metzinger himself observes, however, "phenomenal mineness is not an all-or-nothing phenomenon" (Metzinger 2003a, 443). It comes in degrees, and perhaps the situation is slightly less clear-cut than Metzinger seems to think. Gallagher has recently argued for a distinction between a *sense of ownership* and a *sense of agency*. Whereas the sense of agency refers to the sense of being the initiator or source of an action or thought, the sense of ownership refers to the sense that it is *my* body that is moving, that the experiences I am living through are given as mine. In normal voluntary action, the sense of agency and ownership coincide. When I reach for a cup, the movement is felt as mine, and I have a sense of initiating or generating the movement. In cases of involuntary action, the two can come apart. If I am pushed or if I am undergoing spasms, I will experience ownership of the

movement—I, rather than somebody else, am the one moving—but I will lack a sense of agency; I will lack an experience of being the agent or initiator of the movement (Gallagher 2000b, 204). The fact that ownership can persist without agency, but not vice versa, might suggest that the former is more fundamental than the latter.¹⁷

It may not be difficult to find first-person statements about thought insertions that, if taken in isolation and at face value, seem to offer ample evidence in support of the claim that some experiential states completely lack a quality of *mineness*. One should, however, not overlook that the subjects of thought insertions clearly recognize that they are the subjects in whom the alien episodes occur. They are not confused about where the alien thoughts occur; they occur in the patients' own minds. That is why they suffer from and complain about it (see Stephens and Graham 2000, 8, 126). To put it another way, there is nothing obviously wrong in thinking that foreign thoughts occur in other minds; it is only the belief that alien thoughts occur in one's *own* mind that is pathological and dreadful. Even if the inserted thoughts are felt as intrusive and strange, they cannot completely lack the quality of mineness and first-personal mode of givenness, since the afflicted subject is quite aware that it is he, himself, rather than somebody else, who is experiencing the alien thoughts. When schizophrenics assert that their thoughts are not their own, they do not mean that they themselves are not having the thoughts, but, rather, that someone else has inserted them and that they, themselves, are not responsible for generating them. Thus, rather than involving a lack of a sense of ownership, passivity phenomena like thought insertions involve a lack of a sense of authorship, or self-agency, and a misattribution of agency to someone or something else.¹⁸

Cotard's syndrome, an extreme kind of nihilistic delusion named for the French neurologist and psychiatrist Jules Cotard, comprises any one of a series of delusions ranging from the fixed and unshakable belief that one has lost money, organs, blood, or body parts to believing that one has died and is a walking corpse. In its most profound form, the delusion takes the form of a professed belief that one does not exist. Thus, patients suffering from Cotard's syndrome might deny their own existence, may explicitly state, not only that they are dead, but also that they do not exist.

According to Metzinger, patients suffering from Cotard's syndrome are *truthfully* denying their own existence (Metzinger 2003a, 455). This choice of term might be slightly surprising, since one may have thought that the

appropriate term would have been “sincerely.” Given Metzinger’s own adherence to a no-self doctrine, perhaps he believes such patients are actually closer to the truth than nonpathological subjects. In any case, according to Metzinger, such delusional statements must be understood literally, and he, therefore, argues that they can function as knock-down arguments against any form of Cartesianism. But does nihilistic delusion really testify to the complete absence of pre-reflective self-intimacy (Metzinger 2003a, 459)? The patients might cease using the first-person pronoun, but does this imply that they lack first-personal access to their own experiences?

In his own description of the syndrome, Metzinger provides a further piece of information that should make us hesitate before accepting any literal interpretation. This is the fact that Cotard patients frequently express a coexisting belief in their own *immortality* (Metzinger 2003a, 456)! These patients will, moreover, typically engage in activities that are quite incongruent with the professed belief. In other words, they frequently demonstrate what is known in the psychiatric literature as “double bookkeeping.” This feature is rather typical of schizophrenia, where patients with paranoid delusions or delusions of grandeur might express the belief that the nursing staff is poisoning their food or that they are the German emperor while unhesitatingly eating their lunch or cleaning the floors, respectively. The fact that the patients frequently fail to act on their delusions in the appropriate way questions any straightforward literal interpretation of the delusions and suggests that it might be wrong to interpret the delusions as if they were simply strongly held ordinary beliefs that happen to be false.

To reject a literal interpretation of delusional statements and argue that such a type of interpretation is unsatisfactory is not intended as an endorsement of the Jaspersian principle of un-understandability. Delusional statements are not meaningless, not simply empty speech acts, or, for that matter, merely extravagant metaphors used to describe otherwise normal situations. Rather, they are attempts to express highly unusual and frequently dreadful experiential situations that inevitably stretch ordinary language to its limit.

I do not, however, intend to offer an alternative positive account or interpretation of delusions since they are a highly complex topic in need of careful analysis.¹⁹ The only point I wish to make is that pathological phenomena, like any other empirical phenomena, are open to interpretation and that their proper elucidation frequently requires long clinical experience with patients. To identify a phenomenological approach to psychopathology with a literal

interpretation of first-person statements is much too facile and belittles the major contributions provided by phenomenological psychiatrists such as Minkowski, Binswanger, Tatossian, Tellenbach, and Blankenburg, among others.

VI A Sense of Self

At the start of this chapter, I posed the question of whether there is an intimate link between self and self-awareness. Does self-awareness involve a reference to a self, or is it possible to speak of self-awareness without assuming the existence of anybody being self-aware? In short, is *self*-awareness to be understood as an awareness of *a self*, or, rather, as the awareness that a specific experience has of *itself*? This phrasing of the question has turned out to be misleading. First, it presents us with a false alternative. Self-awareness is not *either* an awareness of a self *or* the awareness that an experience has of itself. On the contrary, it must be realized that there are different kinds of self-awareness. I can be pre-reflectively aware of my current psychological states, be they perceptions, memories, desires, or bodily sensations, and I can reflect on and, thereby, thematize these individual states. However, I can also reflect on myself as the subject of experience, that is, I can reflect on myself as the one who thinks, deliberates, resolves, acts, and suffers. If I compare that which is given in two different acts of reflection, say, a perception of birds and a recollection of a walk, I can focus on that which has changed, namely the intentional acts, but also on that which remains identical, the subject(ivity) of experience. Second, the formulation suggests that if self-awareness were merely a matter of the awareness that an experience had of itself, we would be dealing with a non-egological or subjectless type of self-awareness. As I hope to have made clear this suggestion is mistaken since it overlooks the *ipseity* of the experiential dimension. Ultimately, this is why Gurwitsch's distinction between an egological and a non-egological theory turns out to be too crude a distinction.

Thus, my conclusion is that there is a minimal sense of self present whenever there is self-awareness. Self-awareness is there not only when I realize that *I* am perceiving a candle, but whenever I am acquainted with an experience in its first-personal mode of givenness, that is, whenever there is something it is like for me have the experience. In other words, pre-reflective self-awareness and a minimal sense of self are integral parts of our experiential life.²⁰