Thomas Metzinger

Why are identity disorders interesting for philosophers?

1. Identity Disorders and their Relevance for the Philosophy of Mind

"Identity disorders" constitute a large class of psychiatric disturbances that, due to deviant forms of self-modeling, result in dramatic changes in the patients' phenomenal experience of their own personal identity. The phenomenal experience of selfhood and transtemporal identity can vary along an extremely large number of dimensions: There are simple losses of content (for example, complete losses of proprioception, resulting in a "bodiless" state of self-consciousness, see Cole 1995, Gallagher and Cole 1995, Sacks 1998). There are also various typologies of phenomenal disintegration as in schizophrenia, in depersonalization disorders and in *Dissociative Identity Disorder* (DID), sometimes accompanied by multiplications of the phenomenal self within one and the same physical system. It is important to not only analyze these state-classes in terms of functional deficits or phenomenology alone, but as *self-representational* content as well. For instance, in the second type of cases just mentioned, we confront major redistributions of the phenomenal property of "mineness" in representational space, of what is sometimes also called the "sense of ownership". Finally, there are at least four different delusions of misidentification (DM[1]; namely Capgras syndrome, Frégoli syndrome, intermetamorphosis, reverse intermetamorphosis and reduplicative paramnesia). Being a philosopher, I will discuss two particular types of identity disorder

in this contribution - disorders, which are of direct philosophical relevance: A specific form of DM, and the Cotard delusion. Why should philosophers do this? And why should psychiatrists care?

If we are seriously interested in a conceptually coherent and empirically plausible theory of the self-conscious mind, then it is important to test our conceptual tools at least against some examples of the enormous phenomenological richness of our target phenomenon. Generally, the relevance of case-studies from cognitive neuropsychiatry consists in the fact that they allow for "reverse engineering": If we can develop an empirically plausible representationalist and functional analysis of identity disorders, we will automatically arrive at a better understanding of what precisely it means to consciously *be someone* in standard situations. Second, it is important to have an effective cure for what Daniel Dennett importantly has identified as "Philosopher's Syndrome" - mistaking a failure of imagination for an insight into necessity (Dennett 1991, p. 401). But there is much more to be learned.

The first thing psychiatrists have to learn from philosophers is that an identity is not something you can simply have, like a bicycle or the color of your hair. From a purely logical point of view identity is not a thing or a property, but a *relation*: Identity is the most *subtle* of all relations, the relation in which every thing stands to itself. On the neurophenomenological level of description, however, we only see how this relation is

represented, how it is portrayed on the level of an individual brain and on the level of the individual person's conscious self-experience. And in psychiatric disorders we frequently witness how this self-representational process can be damaged in various ways. But even here there is no mysterious thing or property – "a" personal identity – that is damaged or lost. What is changed are certain representational and functional properties in the central nervous system. Of course, we could decide on purely theoretical grounds that a human being with a certain kind of damage to such properties – say, a patient who has irreversibly lost all autobiographical memory and all access to rational modes of cognition – does not fulfill the conditions of personhood any more. But this would not directly change the patient's phenomenology. "Personal identity" can be either a complex theoretical concept or a concrete from of subjective experience, a conscious content – but it never is a thing, neither in the brain nor anywhere else. To put the point differently: Psychiatrists must stop being naïve realists about personal identity.

The first thing philosophers have to learn form psychiatrists, is that "personal identity" is not only the topic of a traditional *theoretical* field of research, which is centered around questions like, "What are the conditions determining the *sameness of persons*, synchronously as well as diachronously?" The identity of persons is something that is also *phenomenally* represented, by the human brain, prereflexively and pretheoretically, on the level of conscious experience itself. Philosophers have to

understand that there is not only an analytical problem of personal identity, but a neurophenomenological one as well – and both are linked in an interesting way. Why is this so? The conscious self, constituted by the content of an organism's *phenomenal self-model* (PSM; for details see Metzinger 2003a), constitutes a concrete kind of standing-in-a-relation-to-oneself, sometimes even *by* representing identity for an organism. But what precisely is it that is being represented? What – given an appropriate cultural context - are the necessary and sufficient conditions for us to first experience ourselves as being *persons*, identical through time? Put very shortly, possessing a globally available and integrated self-representation enables an information-processing system to stand in new kinds of relations to itself (in functionally accessing different kinds of system-related information *as* system-related information), and if a certain portion of this information is highly invariant, the phenomenal experience of *transtemporal* identity can emerge. However, we have to be careful at this point: Conceptually, indiscriminability is certainly not equivalent to identity. Identity is a transitive relation, indiscriminability is not. Indiscriminability in the self-model may just cause certain functional invariances – for example on the level of body image, background emotions, or autobiographical memory. However, a transparent representation[2] of such functional invariances can result in the *phenomenology* of being identical through time, of transtemporal sameness. And this phenomenology is what, first, functionally *enables* us to refer to ourselves as persons and,

second, constitutes the roots of our theoretical discourse on personal identity. Without this phenomenology we could not write a book like this, and you could neither read nor understand it. Conceptual self-representation is anchored in phenomenal self-representation; personal-level properties are to a considerable degree determined by subpersonal properties in the brain. The first lesson philosophers have to learn, then, is how vulnerable and how fragile human beings actually are on the level of those subpersonal properties. More specifically, they have to take notice of the fact that there are strong neurophenomenological conditions of possibility for personhood, that there is no part of human self-consciousness which is immune to epistemic disturbances resulting from subpersonal disintegration, and that; therefore, empirical constraints are relevant and indispensable for anybody who is seriously interested in the philosophy of self-consciousness.

Philosophy and psychiatry share many common epistemic targets. The most central of them may lie in achieving some more substantial progress, a growth of knowledge with regard to human *self-consciousness.* What we need is a conceptually coherent theory of the self-conscious mind, which is phenomenologically and empirically plausible at the same time. Therefore, one important first step is to analyze unusual neurophenomenological configurations of self-consciousness from two directions at the same time: From the point of view of theoretical psychiatry and from the metatheoretical

perspective developed by analytical philosophy of mind.

2. Example 1: Delusional Misidentification

Let us begin with *reverse intermetamorphosis,* usually defined as "the belief that there has been a physical and psychological change of oneself into another person" (Breen et al. 2000, p. 75) I will argue that this kind of disorder rests on a deviant form of phenomenal self-modeling, and that in principle it could also occur in a non-linguistic/non-cognitive creature not able to form anything like "beliefs" in the more narrow philosophical sense. Here is a brief excerpt from a recent case study on a patient Roslyn Z, conducted by Nora Breen and colleagues:

> RZ, a 40-year-old woman, had the delusional belief that she was a man. This had been a stable delusion for two months prior to our assessment with RZ. During most of that two months she believed that she was her father, but occasionally she would state that she was her grandfather. At the time we saw RZ, she had taken on the persona of her father. She would only respond to her father's name, and she signed his name when asked to sign any forms. She consistently gave her father's history when questioned about her personal history. For example, she said she was in her 60s. (…) The following excerpts are from an interview with RZ. Throughout the interview RZ's mother, Lil, was sitting beside her.

Examiner: Could you tell me your name?

RZ: Douglas.

Examiner: And your surname?

RZ: B_____.

Examiner: And how old are you?

RZ: I don't remember.

Examiner: Roughly how old are you?

RZ: Sixty-something.

Examiner: Sixty-something. And are you married?

RZ: No.

Examiner: No. Have you been married?

RZ: Yes.

Examiner: What was your partner's name?

RZ: I don't remember. Lil.

Examiner: Lil. And you have children?

RZ: Four.

Examiner: And what are their names?

RZ: Roslyn, Beverly, Sharon, Greg.

(…)

8

RZ standing in front of a mirror looking at her own reflection.

Examiner: When you look in the mirror there, who do you see?

RZ: Dougie B_____ (*her father's name*)

Examiner: What does the reflection look like?

RZ: His hair is a mess, he has a beard and a moustache and his eyes are all droopy.

Examiner: So is that a man or a woman?

RZ: A man.

Examiner: How old is Dougie?

RZ: Sixty-something.

Examiner: And does that reflection you are looking at now look like a sixty-something person?

RZ: Yes.

Examiner: It looks that old does it?

RZ: Yes.

Examiner: Do you think that a sixty-something year old man would have grey hair?

RZ: Well, I haven't worried a lot over the years so my hair didn't go grey.

Examiner: So it's not grey?

RZ: No. It's brown.

(Breen et al. 2000, p. 94p/98p)

Reverse intermetamorphosis is of philosophical relevance because it challenges the Wittgenstein/Shoemaker-principle of immunity to error through misidentification (Wittgenstein 1953, S. 67, Shoemaker 1968) Very obviously there are cases of phenomenal self-representation, of the phenomenal representation of one's own *personal identity* in particular, which are misrepresentations. What philosophers have to see is that there are subsymbolic, non-criterial, and phenomenally transparent forms of self-representation, which are fundamentally fallible and can lead to an obvious error through misidentification on the level of linguistic, personal-level self-reference. I will briefly come back to this point in section 4.

Delusional misidentification is a symptom rather than a syndrome comprising a stable collection of symptoms, and the particular variety of self-misidentification is closely associated with severe psychotic states (Förstl et al. 1991, p. 908p). I now want to shift my reader's attention to a full-blown syndrome, which, again, is of great relevance for the philosophy of mind.

3. Example 2: Cotard syndrome

Let us briefly look at a second type of identity disorder, which once again is of particular theoretical relevance. This disorder is what I would like to call *existence denial*. In the year 1880 French psychiatrist Jules Cotard introduced the term *délire de négation* to

refer to a specific kind of “nihilistic” delusion, the central defining characteristic of which consists in the fact that patients deny their own existence, and frequently even that of the external world (see Cotard 1880, for a more detailed account see Cotard 1882) From 1879 onwards this condition was referred to as the “Cotard Syndrome” in the scientific literature (Séglas 1879; for a concise review of the literature see Enoch and Trethowan 1991, p. 163pp) Although there still is considerable discussion about the notion of a delusion as such (see, e.g., Young 1999) and about the conceptual status of “pathological” belief systems (see Coltheart and Davies 2000, Marshall and Halligan 1996) in general, most researchers tend to agree that the Cotard syndrome is likely a distinct theoretical entity.[3] I will here simply treat it as a neurophenomenological state-class characterized by a specific form of deviant self-modeling, without entering into any further empirical speculations.

As a delusion, the Cotard syndrome certainly is quite dramatic - for instance, because it violates the global logical coherence of the patients “web of belief” (see Young 1999, p. 582p) and simply *ends* biographical coherence, while exhibiting a more or less modularized damage to the cognitive model of reality to which the patient has conscious access. The Cotard syndrome is a *monothematic* disorder (Davies et al. 2002). As in many delusions, it is the rather isolated nature of a specific belief content that initially raises serious doubts about his status as a rational subject. However, as a closer look at the data

will reveal, a Cotard patient may actually count as a rational subject, because he develops the only possible conclusion from a dramatic shift in his subcognitive PSM. A promising attempt towards a testable and conceptually convincing hypothesis may therefore start from the assumption that the Cotard disorder is simply a modularized, cognitive-level reaction to very uncommon perceptual experience (Young and Leafhead 1996) It must be clearly noted, though, that the phenomenology of firmly believing in one's own non-existence may also turn out to be too intricate and complex to be tractable under classical belief-desire-approaches - for instance, because it decisively involves pathology in *nonpropositional* levels of phenomenal self-representation.

Obviously, any good future philosophical theory of mind should be able to incorporate the "existence denial" exhibited by Cotard subjects as an important phenomenological constraint, one to be satisfied by its own conceptual proposals. There exists a whole range of neurological disorders characterized by an unawareness of specific deficits following brain injuries, all of them falsifying the Cartesian notion of epistemic transparency for phenomenal self-consciousness, i.e. the idea that one cannot be wrong about the contents of one's own mind, that unnoticed errors are impossible because the light of knowledge shines through and through the self. In pure and extreme versions of the Cotard delusion we are confronted with a particularly interesting example of this type of representational configuration: Patients may explicitly state not only that

they are dead, but also that they don't *exist* at all. In other words, something that seems an *a priori* impossibility on logical grounds – a conscious subject truthfully denying its own existence – turns out to be a phenomenological reality. And phenomenology has to be taken serious. In this case the first lesson to be drawn is this: You can be fully conscious and still truthfully describe the content of your own phenomenal self-experience as "non-existence". In other words, there are actual, nomologically possible representational configurations in the human brain, which lead truthful subjects into logically incoherent autophenomenological reports. What is needed is a representationalist analysis of this target phenomenon, which can lay the conceptual foundations for a truly explanatory account on functional and neuroscientific levels of description.

Elsewhere (Metzinger 2003a, section 6.4.4; 2003b), when discussing cognitive subjectivity as a challenge to naturalism, I offered a representationalist analysis of the Cartesian thought:

- **[I am certain, that I* exist.]**[4]

In extreme forms of the Cotard delusion, we are faced with a delusional belief that can be expressed as follows:

- **[I am certain, that I* do not exist.]**

Weaker forms are delusional beliefs that can be expressed as follows:

- **[I am certain, that I* am dead.]**

What is the phenomenological landscape of the Cotard delusion? A recent analysis of 100 cases (Berrios and Luque 1995, see Fig. 2, p. 186) points out that severe depression was reported in 89% of the subjects, with the most frequent forms of “nihilistic delusion” concerning the body (86%) and existence as such (69%) Other very common features of the reported content of the PSM in Cotard patients are anxiety (65%), guilt (63%), hypochondriacal delusions (58%) and, even more surprising, delusions of immortality (55%; see cf. Berrios and Luque 1995, p. 187) In many cases, certain elements of the bodily self-model seem to have disappeared, or at least to be attentionally unavailable. For instance, one 59-year-old patient would say, “I have no blood” (Enoch and Trethowan 1991, case 5, p. 172), or another one would say, “I used to have a heart. I have something which beats in its place” (ibid., p. 173) while a further case study (Ahleid 1968, quoted after Enoch and Trethowan 1991, p. 165) reports a patient asking “to be buried because he said he was ‘a corpse which already stinks’. A month later he said that he had no flesh and no legs or trunk. These ideas were unshakeable, so that the clinical picture remained unchanged for months”. Such early stages will frequently proceed to states in which the body as a whole is denied, and the patient feels like a “living corpse”, e.g., when saying “I am no longer alive”; “I am dead” (Enoch and Trethowan 1991, case 1, p. 168) or stating “ “I have no body, I am dead” (Enoch and Trethowan 1991, case 2, p. 168) or, like Young

and Leafhead's patient WI (Young and Leafhead 1996, 154pp), simply being convinced that he was dead for some months after a motorcycle accident (involving contusions in the right hemisphere temporo-parietal areas and bilateral frontal damage), with this belief then gradually resolving over time.

Interestingly, there are a number of other phenomenological state-classes, in which a person may experience herself as bodiless or disembodied – ranging from cases like Christina (Sacks 1998) or Ian Waterman (Cole 1995) to certain types of OBEs and lucid dreams (see Metzinger 2003a, sections 7.2.3.2 and 7.2.5) What seems unique about the Cotard delusion is the additional belief that one is dead. The Cotard patient, on the level of his *cognitive* self-model acquires a new, consciously available content – and this content is specific in being highly irrational and functionally immune to rational revision. I would propose that the difference in the phenomenal content in the transparent and *subcognitive* layers of his self-model can be aptly described by employing a traditional conceptual distinction, which is only available in the German language, but not in Greek (e.g., Homer only uses "soma" as referring to corpses), Latin (with *corpus* only referring to the body-as-thing, being the etymological root of the English term "corpse") and other important languages like Italian, French, Spanish or English: The Cotard patient only has a bodily self-model as a *Körper*, but not as a *Leib* (see also Alheid 1968, and Enoch an Trethowan 1991, p. 179p) What is the difference? A *Leib* is a *lived body*,

one that is connected to a soul, or, in more modern parlance, the body *as subject*, as locus of an individual first-person perspective. The body *as inanimate object*, on the other hand, is what the PSM in the Cotard configuration depicts. The Cotard-patient only has access to a *Körper*-model, but not a *Leib*-model.

It is interesting to note how one can arrive at a better understanding of the underlying neurophenomenological state-class by simply following the traditional line of thought. What kind of loss could make a *Leib*-model a representation of something inanimate, of something that is not any more tied to the inner logic of life (see Damasio 1994, 1999)? If the logic of survival is, made globally available by conscious emotions, then a complete loss of the *emotional* self-model should have precisely this effect. Philosophically speaking this would mean that what the Cotard patient claiming to be a dead corpse is truthfully referring to is the transparent content of his self-model, predominantly concerning the spatial, proprioceptive and emotional layers. This content portrays a moving *res extensa*, from the inside, closely resembling a living human person, but, as a matter of phenomenal fact, not tied to the logic of survival any more. In traditional philosophical terminology, the patient has a new belief *de se*. As the first-order, non-conceptual self-representational content grounding this belief, in the most literal sense possible, simply does not any more *contain* the information that actual elementary bioregulation is still going on, as emotions are completely flattened out.

Technically speaking there is no subjective representation of the current degree of satisfaction of the adaptivity-constraint (Metzinger 2003a, section 3.2.11) any more, and this leads the patient forms a hypothesis. This hypothesis, given his current internal sources of information, is absolutely coherent: He must be a dead object resembling a human being. The existence of this object, although experienced as the origin of an internal perspective, does not affectively *matter* to the patient in any way. An important author in the field, Philip Gerrans (Gerrans 2000, p. 112), writes: "The Cotard delusion, in its extreme form, is a rationalization of a feeling of disembodiment based on global suppression of affect resulting from extreme depression." Many Cotard patients make utterances like "I have no feelings" (Enoch and Trethowan 1991, case 4, p. 171) or, like Young and Leafhead's patient KH (Young and Leafhead 1996, 160pp), state that they are dead as a result of "feeling nothing inside". If it is the conscious self-model which, as I would claim, mediates embodiment not only on the phenomenal, but also on the functional level, then these patients suffer from functional deficits, because they are, due a severe impairment in their PSM, emotionally disembodied. Computationally speaking it is a specific subset of system-related information, which cannot be made globally available any more. This fact then triggers further changes on the cognitive level of self-modeling.

As Gerrans (1999, p. 590) has pointed out, on the level of representational content

the Cotard delusion may be miscategorized if described as a DM. At least in weaker forms, the Cotard patient truthfully reports about the content of a very unusual PSM. This unusual PSM results from a globalized loss of affect, Gerrans hypothesizes, mirroring the *global* distribution of the neurochemical substrate that causes the actual deficit, which in turn is then cognitively interpreted. The issue is to first explain how such a PSM could actually come about. Classical belief-desire-psychology and a traditional philosophical analysis in terms of propositional attitudes may, however, not be very helpful in bridging the necessary levels of description in order to arrive at a fuller understanding of the target phenomenon. For instance, it seems plausible that a non-linguistic creature like a monkey could suffer from most of the Cotard phenomenology, without being able to utter incoherent autophenomenological reports. An animal could *feel* dead and emotionally disembodied, without possessing the capacity to self-ascribe this very fact to itself, linguistically or on the level of cognitive self-reference. What is special about the human case is that an isolated, and functionally rigid new element on the level of the cognitive self-model is formed. As soon as we arrive at a convincing representational and functional analysis of the human Cotard patient's self-model, we can proceed to the philosophical issue of whether it is possible to exhibit strong first-person phenomena in Lynne Baker's (1998; see also Metzinger 2003a,b) sense, while simultaneously entertaining the belief that one actually doesn't exist at all. I will return

to this issue below.

The Cotard delusion may be analysed as a combination of loss of a whole layer of non-conceptual, transparent content and a corresponding appearance of new, quasi-conceptual and opaque content in the patient's PSM. Is there a more specific way of describing the causal role of the information now not available on the level of the patient's phenomenal model of reality any more? What *triggers* the massive restructurization of his "web of belief"? A second, and more specific hypothesis may be the "emotional disembodiment"-conjecture: The PSM of the Cotard patient is emptied of all emotional content, making it, in particular, impossible to consciously experience *familiarity* with himself. What the Cotard patient loses may be precisely the phenomenal quality of "prereflexive self-intimacy", the experience of always being in an unshakeable, maximally direct contact with himself. The Cotard patient is not infinitely close to herself, but infinitely distant. If, in addition, it is true that emotional content, generally speaking, represents the logic of survival to an organism, then *self-representational* emotional content, in particular, will represent the internal logic of autonomic self-regulation, of its *own* life-process to this organism. For the Cotard patient, the logic of survival has been suspended: His life-process – although functionally unfolding as still continuously realized in the physical body - is not *owned* any more, by being represented under a PSM. His life-process is not only not his own any more, it may not even be part of

his phenomenal *reality* any more – depending on the severity of the degree of psychotic depression. The fact that autonomous self-regulation, a continuous bodily process aiming at self-preservation is going on is not a globally available fact anymore; thereby this fact is not a part of the patient's reality anymore. It is not appropriated under the PSM. It is important to note that the dynamic process of self-modeling does not generate static forms of mental content, how it is a process of *self-containing.* A Cotard patient may therefore be described as a living, self-representing system that cannot self-contain the fact that it actually is a *living* system any more.

However, there is more to the Cotard delusion than emotional disembodiment, leading to a persistent false belief via the principle of phenomenal self-reference (i.e., *all* cognitive self-reference ultimately can only refer to the content of the currently conscious self-model, Metzinger 2003a, 436) Claiming to be dead – in terms of a dead *body* - is not the same as claiming to be *non-existent*. As Enoch and Trethowan write:

> Subsequently the subject may proceed to deny her very existence, even dispensing altogether with the use of the personal pronoun 'I'. One patient even called herself 'Madam Zero' in order to emphasize her non-existence. One of Anderson's patients said, referring to herself, 'It's no use. Wrap it up and throw "it" in the dustbin' [referring to Anderson 1968; TM]. (…)
>
> If the delusion becomes completely *encapsulated,* the subject may even be able to

> assume a jovial mood and to engage in a philosophical discussion about her own existence or non-existence. (Enoch and Trethowan 1991, p. 173; 175)

To my knowledge there is only one other phenomenal state-class in which speakers sometimes consistently refer to themselves without using the pronoun ‘I’, namely during prolonged mystical and/or spiritual experiences. What seems to be common to both classes is that the phenomenal property of selfhood is not instantiated any more, while a coherent model of reality as such is still in existence. However, a Cotard patient may express his dramatic and generalized emotional experience of unfamiliarity by even denying the existence of reality as a whole. The phenomenon of explicit existence denial can not be ignored, because on the level of explicitly negated content, it is the second most common representational feature of the Cotard delusion, to be found in 69% of the cases (Berrios and Luque 1995, p. 187) Therefore, the pivotal question is: What kind of neurophenomenological configuration could lead a human being into a) denying his or her own existence, and b) stop using the personal pronoun “I” (or “I*” for that matter, see Metzinger 2003a, section 6.4.4; 2003b)?

Let us begin by considering the first issue. The conscious representation of existence is coded via phenomenal transparency (Metzinger 2003a, section 3.2.7) Phenomenally, we are beings experiencing the content of a certain active representations as *real*, if and only if, earlier processing stages of this representation are attentionally

unavailable to us. This leads to the prediction that if a human being's self-model became fully opaque, then this person would experience herself as non-existent – the phenomenal property of selfhood would not be instantiated any more. The subject-component of such a being's PMIR[5], the self-component in its consciously experienced first-person perspective, would be fully opaque, and only continue to function as the origin its first-person perspective for *functional* reasons, due to the ongoing source of continuous input making it the *functional* centre of its representational space. As there would be no phenomenal *self as subject* any more, such a system would not have a PSM in the true sense of the term any more, and its PMIR, strictly speaking, would only instantiate a functional, but not a *phenomenal* first-person perspective any more. I take this to be the natural explanation for the prolonged spiritual and mystical experiences mentioned above, but will not pursue this line of thought any further here. The second logical possibility is that in extreme Cotard configurations existence is denied because there is no PSM at all in existence any more. This is empirically implausible, because Cotard patients certainly exhibit a high degree of sensorimotor integration, of coherent speech, etc. – they certainly are not comatose on in deep sleep.

The third possibility, then, is that a transparent, conscious self-model is in place, but it is not a *subject*-model any more, only an *object*-model. Something still exists, something that looks like the model of a person, but something that is utterly unfamiliar,

not alive, and not a phenomenal self in the act of living, perceiving, attending and thinking. The PSM has lost the emotional layer. The consciously experienced first-person perspective in such a case would not be a model of a subject-object-relation, but only one of an object-object-relation. It would not constitute a phenomenal *first-person*-perspective, but rather a *first-object*-perspective. The "first object", for purely functional reasons persists as the invariant centre of reality, because it is tied to an invariant source of internally generated input. Phenomenally, this functional centre is the place at which things happen, but all of them – as well as the centre itself – are not *owned*. The phenomenal property of "mineness" has disappeared from the patient's reality. As Philip Gerrans puts it:

> In this type of case the patient conceives of herself as nothing more than a locus, not of experience – because, due to the complete suppression of affect, her perceptions and cognitions are not annexed to her body – but of the registration of the passage of events. (Gerrans 2000, p. 118)

Let us now proceed to the second issue: Why would such a system stop using the pronoun "I" when referring to itself? The answer to this question would have to explain the phenomenon for two different phenomenal state-classes: Spiritual experiences and the Cotard delusion. In both cases, the system still operates under a functionally centred model of reality. Motor control, attentional processing and cognitive availability are in

place, and in principle well integrated. Sensorimotor integration is successfully achieved. In both cases, the subject-component of the PMIR is not a *subject*-component any more on the level of phenomenal experience. Phenomenologically, both state classes are constituted by *subjectless* models of reality. On the representational level of analysis we find that there is no globally available representation of a *self as subject*. We currently know very little about the minimally sufficient neural correlates for both types of states, but it is plausible to assume that the aetiologies are highly different: In one state-class it is a considerable elevation in the degree of attentional availability for earlier stages of processing contributing to the system-model, in the second state-class it is a loss of a particular layer, namely the emotional layer, which transforms a subject-model into an object-model, turning a first-person state into a "first-object" state. If my short analysis points into the right direction, we can give a new answer to the old philosophical question of what the personal pronoun "refers" to. "I*" inevitably refers to a specific form of phenomenal mental content: to the transparent, *subcognitive* partition of the currently active PSM depicting the speaker *as subject*. If this partition is lost or become opaque, then speakers will stop using "I*".

4. Conclusion

Of course, my brief philosophical interpretation of the empirical material may be false or utterly misguided. But I hope it has served to demonstrate the *relevance* of

identity disorders to the philosophy of mind. Identity disorders, while being diagnosed on the *personal* level of description, result from subpersonal disintegration (see Gerrans 1999, see also Dennett 1998). Especially when operating from a genetic perspective, when investigating the causal history of such dramatic "personality shifts", subpersonal levels of explanation have to be taken into account. The concept of a PSM forms the logical link between subpersonal and personal levels of description: As a neurobiological, functional, and representational entity it is subpersonal, but by satisfying the transparency-constraint (see Metzinger 2003a, sections 3.2.7 and 6.2.6) it generates personal-level properties like phenomenal selfhood for the system as a whole, enabling "strong" first-person phenomena, social cognition and intersubjectivity. Classical, egologic, theories of mind simply fail by not being able to provide any satisfactory account when confronted with phenomenological material like the one presented. But certain standard assumptions underlying modern analytical philosophy of mind create major difficulties in developing a conceptually clear analysis as well.

Example 1, delusional misidentification clearly challenges the Wittgenstein/Shoemaker-principle of immunity to error through misidentification (Wittgenstein 1953, S. 67, Shoemaker 1968). Why? Very obviously there are cases of phenomenal self-representation, of the phenomenal representation of one's own *personal identity* in particular, which are misrepresentations. One can be entirely mistaken about

which person one is, not only in a weak sense of being disoriented about one's own personal identity or of temporarily not knowing *who* one is, but in the strong sense of consistently experiencing oneself as *another* person. And this can happen without a chance to gain higher-order insight into this very fact, under the condition of epistemic opacity. What philosophers have to see is, first, that there are subsymbolic, non-criterial, and phenomenally transparent forms of self-representation and self-identification. Second, these non-linguistic and obviously internal (i.e., locally supervening) forms of self-identification are fundamentally fallible and can therefore lead to an obvious error through misidentification on the level of linguistic, personal-level self-reference. This error would then not be an error based on an object-use of the first-person pronoun "I", but on a subject-use *based on the content of a subpersonal misrepresentation, namely a delusional and phenomenally transparent PSM of oneself as another person*.

I have discussed the philosophical implications of Example 2 at greater length in the previous section. I can therefore be brief in pointing out its metatheoretical relevance: Existence denial is important for *theories of rationality*, because they show how an obviously false belief can be highly modularized, immune to revision, and embedded into a complex network of cognitive self-representation that remains largely coherent. It is also important, because phenomena like the Cotard delusion provide us with a valuable possibility to investigate, first, the relationship between phenomenally transparent and phenomenally

opaque mental content in the constitution of human self consciousness. Second, they permit for a closer description of the relationship between the emotional and the cognitive self, between the prereflexive and the reflexive layers in the human PSM. For reasons of space, I cannot penetrate deeper into these issues here (see Metzinger 2003a for a more detailed account). I nevertheless hope that my short comments could serve to establish the general conclusion that philosophy has a lot to learn from psychiatry.

One last question remains to be answered: Why should psychiatrists care about all of this? Here the answer can be even more simple and short: Theoretical progress is one of the most important factors in the alleviation of human suffering.

References

Ahleid A (1968) Considerazioni sull'esperienza nichilistica e sulla syndrome die Cotard nelle psicosi organiche e sintomatiche. *Il Lavoro neuropsichiatrico* 43: 927-945.

Anderson EW (1964) *Psychiatry*, 1st edition. London, Baillière, Tindall & Cox.

Baker LR (1998) The first-person perspective: A test for naturalism. *Amer Phil Quart* 35: 327-46.

Berrios GE, and Luque R (1995) Cotard's syndrome: analysis of 100 cases. *Acta Psychiatr Scand* 91: 185-188.

Breen N, Caine D, Coltheart M, Hendy J, and Roberts C (2000) Towards an understanding of delusions of misidentification: Four case studies. In Coltheart and Davies 2000.

Cole J (1995) *Pride and a Daily Marathon*. Cambridge, MA, MIT Press.

Cole J, and Paillard J (1995) Living without touch and peripheral information about body position and movement: Studies with deafferented subjects. In Bermúdez, Marcel, and Eilan 1995.

Coltheart M and Davies M (eds) (2000) *Pathologies of Belief*. Oxford and Malden, MA, Blackwell.

Cotard J (1880) Du délire hypocondriaque dans une form grave de la mélancolie anxieuse. *Annales Médico-Psychologiques* 38: 168-170.

Cotard J (1882) Du délire des négations. *Archives de Neurologie* 4: 152-170/282-295.

Damasio AR (1994) *Descartes' Error*. New York, Putnam/Grosset.

Damasio AR (1999) The Feeling of What Happens: Body and Emotion in the Making of Consciousness. New York, NY, Harcourt Brace & Company.

Davies M, Coltheart M, Langdon R, and Breen N (2002) Monothematic delusions: Towards a two-factor account Phil Psychiat Psych (special issue 'On Understanding and Explaining Schizophrenia', ed. C Hoerl) 8 (2/3, June/September 2001), 133-58.

Dennett DC (1991) *Consciousness Explained*. Boston, Toronto and London, Little, Brown and Company.

Dennett DC (1998) Postscript. In DC Dennett, ed., *Brainchildren – Essays on designing minds*. Cambridge MA, MIT Press.

Enoch MD and Trethowan WH ([3]1991) *Uncommon Psychiatric Syndromes*. Oxford, Butterworth-Heinemann.

Förstl H, Almeida OP, Owen A, Burns A, and Howard R (1991) Psychiatric, neurological and medical aspects of misidentifications syndromes: a review of 260 cases. *Psychol Mee* 21: 905-910.

Gallagher S, and Cole J (1995) Body schema and body image in a deafferented subject. *J Mind Behav* 16: 369-90.

Gerrans P (2000) Refining the explanation of Cotard's delusion. In M Coltheart and M Davies (eds) *Pathologies of Belief*. Oxford and Malden, MA, Blackwell.

Metzinger T (2003a) Being No One. The Self-Model Theory of Subjectivity. Cambridge, MA, MIT Press.

Metzinger T (2003b) Phänomenale Transparenz und kognitive Selbstbezugnahme. In U. Haas-Spohn (Hrsg.), *Intentionalität zwischen Subjektivität und Weltbezug*. Paderborn, mentis.

Séglas J (1897) *Le Délire des Négations: Séméiologie et Diagnostic.* Paris: Masson, Gauthier-Villars.

Shoemaker S (1968) Self-reference and self-awareness. *J Phil* 65: 555-567. Reprinted in S Shoemaker (1996), *The First-Person Perspective and other Essays*. Cambridge, Cambridge University Press.

Wittgenstein L (1953) *Philosophical Investigations*, trans. G.E.M. Anscombe. London, Macmillan.

Young AW and Leafhead KM (1996) Betwixt of life and death: Case studies of the Cotard delusion. In PW Halligan and JC Marshall (eds) *Method in Madness*: Case studies in cognitive neuropsychiatry. Hove, UK, Psychology Press.

Young AW (1999) Delusions. *Monist* 82: 571-90.

Young AW, Robertson IH, Hellawell DJ, de Pauw KW, and Pentland B (1992) Cotard delusion after brain injury. *Psychol Med* 22: 799-804.

---

[1] In using the abbreviation "DM" instead of "DMS" (for "delusional misidentification *syndromes*") I follow a recent terminological modification introduced by Nora Breen and colleagues in order to avoid the inaccuracy of assuming a whole cluster of symptoms as defining characteristics for the different categories. Cf. Breen, Caine, Coltheart, Hendy and Roberts 2000, p. 75, n. 1,

[2] A transparent representation is one that cannot be experienced as such. For more on the notion of "transparency", see Metzinger 2003a, section 3.2.7 and 6.4.2; Metzinger 2003b, section 2)

[3] Young and Leafhead (1996, p. 154) argue that there is no specific symptom shown by every one of Cotard's pure cases, whereas Berrios and Luque (1995) offer a statistical analysis of historical clinical usage of the term, concluding that a pure Cotard syndrome (represented by "Cotard type 1" patients) does exist, and that its nosological origin *"is in the delusional and not in the affective disorders;…"* (ibid., p. 187) I will not take a position on this issue here.

[4] The asterisk is here used following the common notation introduced by Hector Neri Canstañeda 1966.

[5] A PMIR is a *phenomenal model of the intentionality relation*, i.e. a conscious representation of the subject as currently related to a specific object or as interacting with another

subject, for instance the conscious experience of a “self in the act of knowing”. A PMIR is a consciously experienced first-person perspective. See Metzinger 2003a, section 6.5 for details.