

Gianfranco Soldati
Mineness, Deflation, and Transparency

§1

Gilbert Ryle famously argued that there is no «difference in kind between a person's knowledge about himself and his knowledge about other people» (Ryle, 1949, p. 181). He especially opposed the «theory that minds must know what they are about, because mental happenings are by definition conscious, or metaphorically self-luminous» (Ryle, 1949, p. 161). Ryle did not accept the idea that there is something in our experiences being conscious that plays a role in the explanation of the distinctiveness of self-knowledge.¹ He had qualms both about consciousness being luminous and about self-knowledge being *sui generis*.

Less than thirty years later Donald Davidson proudly announced that «Ryle was wrong» (Davidson, 1987, p. 441). Davidson argued that Ryle «stoutly maintained that we know our own minds in exactly the same way we know the minds of others, by observing what we say, do, and paint» (Davidson, 1987, p. 441). Contrary to this, Davidson suggested that «it is seldom the case that I need or appeal to evidence or observation in order to find out what I believe; normally I know what I think before I speak or act» (Davidson, 1987, p. 441).

Davidson went on developing his own account of self-knowledge, concentrating on one of its central aspects, namely epistemic authority. He did not really wonder whether there was something about the conscious features of experience that would play a role in the explanation of self-knowledge. On the contrary, Davidson emphasized that the authority of self-knowledge can be fully accounted for without relying on anything as experiences, understood as mental objects that are directly given to the mind. In spite of the declared disagreement, Davidson was disregarding the conscious features of experience in his account of self-knowledge just as plainly as Ryle did. They both opposed the Cartesian tradition, that

¹ Self-knowledge is knowledge of a subject about herself as a subject. It is typically expressed by utterances containing the first-person pronoun. As such, self-knowledge can be about any fact concerning the subject. In this paper, I shall restrict myself to self-knowledge concerning experiences, i.e. knowledge in the first person about one's own experiences. Unless explicitly stated, the expression 'self-knowledge' will be used with such a restriction.

was said to attribute a special role to the conscious and subjective character of experience.²

But what exactly does it mean to attribute a special role to the conscious character of experience in one's account of self-knowledge? There are various answers to that question. In this paper, I am interested in one such answer. It is the answer that says that a proper account of self-knowledge needs to consider a special feature of conscious experience, namely the fact that each experience comes with a feeling of first personal ownership, sometimes called phenomenal mineness. I shall try to understand what exactly is meant by this sort of mineness and then elucidate its potential in playing a role in an account of self-knowledge. I shall argue that phenomenal mineness is not required for a proper account of first personal knowledge of one's own experiences.

In *section two* I shall introduce the notion of mineness as it is used in some recent writings by Kriegel and Zahavi. It will appear that it is related to the idea that an experience *is given* in the first person in a special way, different from the way it is presented in the third person. I shall assume that the expression 'to be given in a certain way' can be interpreted epistemologically, thus referring to a way one can come to know something. To say that an experience is given to the subject in a special way is thus to say that there is a special way a subject can come to know her own experience. I shall focus on the claim that there is a direct relation between phenomenal mineness and first personal knowledge of one's own experiences. The claim under scrutiny will be that one cannot accept the idea that there is a special way a subject can come to know her own experience and yet reject the presence of phenomenal mineness. I shall label this contention the *mineness link*. In the following sections of the paper I shall argue against it and propose an alternative. An argument against the mineness link is not as such an argument against phenomenal mineness. In this perspective, the present paper pursues a different goal than Howell's and McClelland's contributions in this volume. They both aim at confronting the advocates of phenomenal mineness with arguments challenging its reality. The present project might thus be understood as arguing for the epistemological inefficiency of

² See also (Davidson, 1988).

mineness, rather than for its psychological unreality.

In *section three* I discuss deflationist accounts of self-knowledge and mineness. These are different forms of deflationism, but they are related. Deflationism about self-knowledge (*epistemological deflationism*) says that an account of the specificity of self-knowledge does not rest on introspective evidence. Deflationism about mineness (*phenomenal deflationism*) says that there is nothing more to the mineness of an experience than the fact that it occurs in me, a subject. Kriegel and Zahavi reject the latter form of deflationism and argue that we must accept a phenomenological fact involving mineness in addition to the mere fact that an experience occurs in a subject. I show that their argument fails to speak for the link between mineness and self-knowledge. We do not need to accept a phenomenological fact involving mineness in order to explain the specificity of self-knowledge. This being so, epistemological deflationism about self-knowledge is still in the run. It might still be the case that a proper explanation of the specificity of self-knowledge does not need to appeal to any phenomenal mineness.

Section four introduces the notion of transparency that is often used by epistemological deflationism. Kriegel and Zahavi argue that mineness is compatible with transparency. There are various forms of transparency. I start by introducing a notion of transparency (as characterized by M. Boyle) that is meant to be general enough to cover various variants. According to such a general notion of transparency, there is a route to first personal knowledge about one's own experiences that does not rely on any inner or introspective warrant. Instead of saying that my experience is given to me in a different way than it is given to you, we should say that the world is given to me – when I want to find out something about my own experience – in a different way than it is given to you – when you want to find out something about my experience.

Before returning to the issue about compatibility with mineness, we need to enquire into the roots of transparency. *Section five* focuses on the case of judgement and argues for a rationalistic account of the transparency of judgement: when one judges that p one has a reason to judge that one judges that p because the reason to which one is responding by the first order judgement rationally commits one to the

second-order belief. This result is presented as a consequence of the fact that rationality does not only require one to avoid having contradicting beliefs, or to form the beliefs that are appropriately commanded by the reasons one has, but that it also demands a cogent sensitivity to the reasons one has.

Section six summarises the argument and presents the conclusion that if there is a sense of phenomenal mines that is compatible with transparency, it is far from involving the sort of phenomenological fact Kriegel and Zahavi are arguing for.

Section seven sketches the main lines of an extension of the rationalistic account presented in section five to conative states, such as desires.

§2

The idea that conscious experiences come with a sense of ownership which plays a role in the explanation of the way we come to know about them in the first person is not new. It was articulated by a number of authors belonging to the tradition of German Idealism³. I shall, however, concentrate on a more recent proposal, made by Dan Zahavi and Uriah Kriegel, in part separately and in part in a common publication.⁴ In this section I shall quote some passages in order to extract the central claims. I shall later look more closely at some of the arguments.

Let me start with a passage from Dan Zahavi:

Imagine a situation where you first see a green apple and then see a yellow lemon. Then imagine that your visual perception of the yellow lemon is succeeded by a recollection of the yellow lemon. [...] If we compare the initial situation [...] with the final situation [...], there has been a change of both the object and the intentional type. Does such a change leave nothing unchanged in the experiential flow? Is the difference between the first experience and the last experience as radical as the difference between my current experience and the current experience of someone else? We should deny this. Whatever their type, whatever their object, there is something that the different

³ See (Frank, 2002) and the influential (Henrich, 1967).

⁴ I shall assume there there is one view at stake, although I suppose that each of them might distance himself from some of the claims accepted by the other.

experiences have in common. [...] The different experiences are all characterized by the same fundamental first-personal character. They are all characterized by what might be called a dimension of for-me-ness or mineness. (Zahavi, 2010, p. 58)

This passage contains a claim and a terminological suggestion. The claim is that there is something common to all the experiences of one and the same subject, independently of their content and mode (or «intentional type», as Zahavi calls it). It remains constant in the flow of experience and it is something none of those experiences shares with experiences of another person. The terminological suggestion is to call this fundamental first-personal character, *for-me-ness* or *mineness*. I shall use the second term.

Most people who think that we have conscious experiences might be prepared, pace Ryle, to endorse the claim that those experiences are given to their owner in a way that is very different from the way somebody else's experiences are given. But Zahavi has an opponent in mind who goes further, maintaining that «there is no property common to all my experiences, no stamp or label that clearly identifies them as mine» ((Zahavi, 2010, p. 59)). To such a contender he responds as follows:

[...] this objection is misplaced in that it looks for the commonality in the wrong place. The [...] mineness in question is not a quality like scarlet, sour, or soft. It doesn't refer to a specific experiential content, to a specific *what* [...]. Rather, it refers to the distinct *givenness* or *how* of experience. [...] It refers to the fact that the experiences I am living through *are given* differently [...] to me than to anybody else. (Zahavi, 2010, p. 59, my emphasis)

Which is to say, as Kriegel and Zahavi put it:

What-it-is-like-ness is properly speaking what-it-is-like-for-me-ness. (Kriegel and Zahavi, 2015, p. 36)

The issue then does not concern so much the asymmetry between first and third personal access to one's experience, but the status of the feature common to all experience of one and the same person. We are told that it «does not refer» to an experiential content, a *what*, but to a givenness or *how*, which is now called «what-it-is-like-for-me-ness».

These are technical terms that are meant to point to something

that, I suppose, cannot easily be described with a more mundane terminology. One important point is that one should not think of mineness along other common phenomenal qualities, such as those related to colours, tastes or smells. It is rather something all experiences of one and the same subject have in common. But what does it mean to say that it is a *how* rather than a *what*? The distinction between what (or that) and how is often used in order to oppose theoretical to practical knowledge. To know how to do something is not the same as to know what is the case, or to know that such and such is the case. But mineness is not announced as characterizing some sort of knowledge. The distinction might rather concern different aspects of experiences. There is something one experiences, this is *what* one experiences, and there is the way one experiences it, this is *how* one experiences. But when an apple appears round, red and sweet, is it not right to say that sweet is the *way* the apple tastes and that red is the *way* it looks? We certainly say things like this: “*How* does the apple taste? It tastes sweet”. The distinction between how and what is not straightforwardly helpful for setting mineness apart from other phenomenal features of experience. We should probably wait for more information in order to settle the point. Let us agree for the moment that mineness is a feature common to all experiences of one and the same subject and that no experience of another subject can share it.

Such a claim obviously needs clarification in order to be properly evaluated. If it is read as meaning that the common phenomenal feature A’s experiences share with nobody else’s experiences is constituted by the simple fact they are A’s experiences and not anybody else’s experiences, then it will be difficult to find any contender. Zahavi is indeed happy to make a stronger point. He declares that «anybody who denies the [...] mineness of experience simply fails to recognize an essential constitutive aspect of experience. Such a denial» he insists, «would entail the view that my own mind is either not given to me at all—I would be mind- or self-blind—or present to me in exactly the same way as the minds of others» (Zahavi, 2010, p. 59). This is the contention that one cannot endorse the idea that my own mind is present to me, or *is given to me*, in a different way than it is present or given to you, and yet reject the

claim that experiences all have the common feature of phenomenal mineness.

Let us consider the supposedly Rylean *reductionist claim*: my experience is given to me in the same way as your experience is given to me. Much depends on what is meant by the locution 'given to'. A first guess would be that the locution is used in this context in order to characterize *ways of acquiring knowledge*. Among the different ways one can obtain knowledge, some are distinguished by the fact that they offer an epistemic warrant. So the reductionist claim under this interpretation denies that there is a fundamental distinction between first personal and third personal warrant for knowledge of one's experiences. This comes indeed close to what Ryle seems to have had in mind, or at least to what Davidson took him to have said. Zahavi's contention would then be that one cannot endorse the idea that there is any asymmetry between first personal and third personal warrant for knowledge of one's experiences, and yet reject the claim that a subject's experiences have the common feature of mineness. This establishes an explicit relation between mineness and self-knowledge.

If this is the right interpretation, then Zahavi's claim would reject not only reductionist positions like Ryle's, but also more conciliatory positions, like Davidson's, who accept an asymmetry between first and third personal access to one's own mind without supposing any experiential mineness. The issue, in this sense, is not whether one should accept mineness, but whether one must accept it in order to make a distinction between first personal and third personal access to the mind. To maintain that there is such an obligation means to maintain that there is a strict link between self-knowledge about one's own experiences and mineness. Let me call this the *mineness link*.

In what follows I shall inquire with more details into the precise relation between mineness and self-knowledge about one's experiences, and into some arguments that might be taken to speak in favour of the mineness link.

§3

Deflationist accounts of self-knowledge may generally be characterised by the fact that they aim at providing an account of the

specificity of self-knowledge that does not rest on any introspective evidence.⁵ As such they do not need to deny the mineness link. They might simply reject the idea that mineness plays a role in a proper account of self-knowledge by virtue of being introspectively accessible. If at all, the mineness link would have to obtain by virtue of some other feature.

Zahavi and Kriegel think that deflationist accounts ought to be rejected. They write:

The for-me-ness of experience still admits of two crucially different interpretations. According to a deflationary interpretation, it consists simply in the experience *occurring* in someone (a ‘me’). On this view, for-me-ness is a non-experiential aspect of mental life—a merely metaphysical fact, so to speak, not a phenomenological fact. In contrast, a non-deflationary interpretation construes for-me-ness as an experiential aspect of mental life, a bona fide *phenomenal* dimension of consciousness. On this view, to say that an experience is *for me* is precisely to say something more than that it is *in me*. It is to state not only a metaphysical fact, but also a phenomenological fact. (Kriegel and Zahavi, 2015, p. 36)

According to a deflationist following Kriegel and Zahavi’s interpretation, mineness of experience can be reduced to the fact that the experience is ‘in’ a subject. They call this a metaphysical fact (I find this terminology awkward, but I shall adopt it for the sake of the argument) and they oppose it to an phenomenological fact, that would contain more than just the fact that an experience belongs to, or is the property of a subject.

We have two forms of deflationism here. Deflationism as I defined it above is silent about phenomenal mineness and about the mineness link. It simply insists on the fact that mineness cannot play any evidential role in self-knowledge. Let me call this *epistemological deflationism*. Deflationism in Kriegel and Zahavi’s conception goes further. It reduces the phenomenological fact involving mineness of experience to a metaphysical fact about ownership of experience. Call this *phenomenal deflationism*. The latter implies the former. If there is no genuine phenomenological fact involving mineness, then mineness

⁵ Cf. (Soldati, 2014)

cannot play any role, let alone any evidential role, with respect to self-knowledge. By rejecting phenomenal deflationism one has not yet rejected epistemological phenomenalism. One must reject phenomenal deflationism, however, in order to question epistemological deflationism. One must establish that experiences have phenomenal mineness if one intends to show that mineness plays a role, whether evidential or other, in self-knowledge.

We saw above that on Kriegel and Zahavi's view there is a link between self-knowledge and phenomenal mineness. If they think that the relation is not evidential, then they owe us an alternative account. I shall simply assume that they would minimally agree that mineness plays a role in determining the specific warrant for self-knowledge.⁶ In this section I intend to show that the phenomenological fact is not better suited to explain the epistemological peculiarity of self-knowledge than the metaphysical fact. From this it will be possible to conclude that a rejection of phenomenal deflationism cannot be used as a starting point for an argument against epistemological deflationism.

Let us assume that the metaphysical fact we are considering is the fact that, for any singular experience, it necessarily belongs to (occurs in the flow of consciousness of) one and only one subject. No two subjects can share one and the same experience. Does this fact contribute to an understanding of self-knowledge? Much depends, of course, on what one generally requires for an attribution of knowledge. But the necessity of *p* does not generally suffice for one's belief that *p* to qualify for knowledge, let alone self-knowledge. Even if I believe that I want an ice-cream, and if this present desire is necessarily mine, this does not suffice for me to know that I want an ice-cream. One would typically require my desire to offer me a reason for believing that I want an ice-cream. So, the metaphysical fact that my experiences are necessarily mine might contribute to self-knowledge, but it does not suffice to explain its specificity.

Does the phenomenological fact fare any better in this respect? Unfortunately, it does not. The phenomenological fact, we are told, contains mineness as a *bona fide* phenomenal dimension. Whatever

⁶ As I said above, I take this to be implied by the idea that there is a specific way one's own experiences are given to oneself, which I understand as meaning that there is a specific way one comes to know one's own experiences.

that may be, it seems safe to say that there is a difference between the metaphysical fact that this experience is necessarily mine and the intended phenomenological fact involving mineness. Let us call this phenomenological fact an *f*-fact (*f* may thus be short for ‘experienced mineness’). Consider now the following peculiar argument. (i) We suppose that we all have our own, personal *f*-facts. I have my own *f*-facts and you have your own *f*-facts. (ii) My *f*-facts are given to me in a different way than they are given to you. (iii) So, I know that I have an *f*-fact in a way you cannot come to know. But why is this so? (iv) It can’t be so just because my *f*-facts are necessarily mine (see argument above). (v) So, there must be something about *f*-facts that accounts for the fact that I come to know them in a way you cannot come to know them. (vi) Call this feature ‘hyper mineness’, a super-phenomenological feature that explains the asymmetry. (vii) And so on.

The argument is odd, but its peculiarity is not easily removed. For instance, it is not a matter of first-order *vs* second-order beliefs, or of pre-reflexive against reflexive consciousness. The mineness of my experience may be given pre-reflexively, or in a first-order belief (the two claims are not equivalent): the argument would still go through. There is simply a problem with the idea that the obtaining of a special phenomenological fact, in addition to what Kriegel and Zahavi call the metaphysical fact, would help to make any progress.

It is rather common at this point to appeal to perspectival, subjective facts. Again, much depends on what such facts are supposed to be. One rather common interpretation is that perspectival facts are facts whose nature prevents them from being known from any perspective but one. Suppose that *f*-facts are such facts. They can be known only from the perspective of the subject they are about. Only I can know my *f*-fact, and only you can know your *f*-fact. One may arguably wonder why such an asymmetry should apply to *f*-facts at all. If I can know that you are having a pain, why should I not be able to know that you are experiencing your pain as yours and not mine? But even apart from that, it is unclear how the perspectival character of experiences is related to the alleged mineness which constitutes the *f*-fact. The perspectival character of experiences is often explained by emphasizing the way the first person would

attribute them to herself – namely by using the first person pronoun. No other person can do so in order to state the same fact – indeed the *same* fact: the fact that one is having a certain experience. There is no obvious need to introduce mineness in order to explain that kind of perspectival difference. It certainly is a difference that applies to the self-attribution of experience just as much as to the self-attribution of any other property. You, but not me, can express the fact that you are alive by using the first person pronoun. But this does not require the fact that you are alive to contain some sort of mineness. So *f*-facts being perspectival does not show that they are so in virtue of involving some phenomenal mineness.

The above argument is certainly not meant to show that a subject having an experience cannot be a perspectival fact. Quite to the contrary. It is rather meant to show that if we do not understand what makes the fact I am having an experience perspectival, then adding the phenomenal property of mineness to the very same fact won't bring any substantial progress. Special phenomenological facts are not suited to explain the specificity of self-knowledge. If the introduction of those special facts was supposed to offer an alternative to a deflationist conception of self-knowledge that would reject phenomenal mineness, then the point was misguided. The upshot is not that the deflationist must be credited with a convincing argument against phenomenal mineness, but that phenomenal mineness does not deliver a contribution to a better understanding of the specificity of self-knowledge.

§4

One central argument in deflationist accounts of self-knowledge uses an introspective feature often called 'transparency'.

Transparency has many facets, but there seems to be a core idea that has suitably been characterized by Matthew Boyle as follows:

«I can *know* various aspects of the nature, content and character of my own mental states by attending in the right way, not to anything 'inner' or psychological, but to aspects of the world at large. Indeed, it seems that [...] all there is for me to contemplate in my sensation of blue is the (apparent) blueness of some worldly thing, and all there is for me to attend to in my

belief that P is the (apparent) fact that P. Various questions about my own present mental state are thus normally ‘transparent’ for me to questions about the world at large» (Boyle, 2011, pp. 225–26, my emphasis).

Kriegel and Zahavi wonder whether transparency speaks against phenomenal mineness. They argue that:

[...] phenomenal consciousness does not only represent but also *presents* something (to someone). Compare a conscious perceptual experience of the color and shape of a yellow lemon and a subliminal or blindsighted representation of the same color and shape. Both represent the same distal features [...]. But only the experience *presents* those features, in the sense of making *someone* phenomenally aware of them. To that extent, although all the presented items are worldly items, the presenting itself—presenting to someone—is an aspect of phenomenal consciousness as well. There is thus a minimal dimension of for-me-ness without which we cannot distinguish consciousness from unconscious representations of the same environmental features. This minimal for-me-ness is fully consistent with the contention that *once* a state of a subject presents something to the subject, it *is necessarily* [*my emphasis*] some putative environmental feature that it presents [...]. If we interpret the transparency claim as exhausted by this contention, we can appreciate that *transparency is compatible with for-me-ness* [*my emphasis*]. (Kriegel and Zahavi, 2015, pp. 40–41)

There are four central steps in this argument. First, it is assumed that both a conscious and an unconscious (perceptual) experience *represent* some features of the external world. It is then argued that in a conscious representation those features are further *presented to someone*. This expression is taken to mean that one is phenomenally aware of those features. In a conscious experience one is phenomenally aware of the represented features. But now, third step in the argument, this very phenomenal awareness is taken to be equivalent to for-me-ness. So, conclusion of the argument, what distinguishes conscious from unconscious representations of the world is something (phenomenal mineness) that cannot be reduced to

a represented feature of the world. Transparency is compatible with the acceptance of phenomenal mineness, if it is understood as implying that one does not find mineness among what is represented, or consciously presented, neither by looking inside nor by looking outside – simply because mineness itself is an aspect that is never presented *to* the mind.

What is the relation between Kriegel and Zahavi's notion of transparency and the notion Boyle has been using in the passage quoted above? In what follows I shall first inquire into the mechanics of the sort of transparency Boyle has described above. I shall then come back to the compatibility claim made by Kriegel and Zahavi.

If mineness is an aspect of experience that is supposedly compatible with transparency, then transparency in Boyle's sense would entail that I can come to know that aspect by attending to aspects of the world at large. What does that mean? Remember that we were wondering about how mineness impinges on an account of the specificity of self-knowledge. We were asking what role mineness plays in the explanation of the asymmetry between first personal and third personal knowledge of one's own mental states. If we apply Boyle's notion of transparency, the difference between my way of knowing my own experience and your way of knowing my experience should not be accounted for in terms of me attending to something other than to aspects of the world at large. The asymmetry between first personal and third personal knowledge should then be accounted for in terms of which aspects of the world we would be attending. In order to find out something about my experience I should attend to aspects of the world that are somehow different from the aspects you would attend to in order to find out something about my experience. Instead of saying that my experience is given to me in a different way than it is given to you, we should say that the world is given to me (when I want to find out something about my own experience) in a different way than it is given to you (when you want to find out something about my experience). This is the recipe for the application of Boyle's transparency. We intend to establish whether it is compatible with phenomenal mineness. But before doing so, we must understand how its application can be supported. What kind of considerations speak in favour of transparency *à la* Boyle?

Boyle's transparency is typically applied to acts of judging. Suppose you deliberate about what to do tomorrow. You look at the weather forecast and judge on its basis that it will snow tomorrow. By applying transparency in Boyle's sense to this act of judging, we obtain the claim that you can come to know aspects of your own judging by attending in the right way, not to anything 'inner' or psychological, but to aspects of the world at large. You come to know that you judge that it will be snowing not by looking inside, but by looking into the world. If mineness is supposed to be an aspect of your act of judging, and if transparency in our sense is supposed to be applicable to it, then you ought to attain knowledge about it by attending to some aspect of your environment. Indeed, it has often been argued that a subject may come to know that she judges that p by attending to the world in order to find out whether p .⁷ Why is that so?

§5

Here is a diagnosis of this form of transparency. The issue concerns the relation between the judgement that p (Bp), and the judgement that one judges that p (BBp)⁸. It appears that the reason one has for judging that p suffices, in the first person, for the self-attribution of the very same judgement. In other words: if I wonder whether I am judging that p the reason I find for judging that p settles the question as to whether I am judging that p . By virtue of what does this transmission of reason obtain? In this section I shall contemplate some options and argue for an account that paves the way for an alternative understanding of the relation between mineness and self-knowledge.

One might first take the relation between Bp and BBp to be simply inferential.⁹ But that would indeed be peculiar. There is no logical relation between p and Bp . The fact that it is raining does not entail that I judge that it is raining. We should rather look for an

⁷ (Evans, 1982, p. 225) is often quoted for this idea. More recent versions of the notion can be found in (Moran, 2001, p. 66), (Moran, 2012) and (Byrne, 2011). There are differences between the different usages of the notion of transparency. I have discussed the differences, and presented the epistemological version of transparency I am using in this paper in (Soldati, 2014).

⁸ I use the symbol ' B ' although I speak of acts of judging for simplicity. I shall occasionally use the term 'belief' for stylistic reasons in what follows, but I mean acts of judging, not the disposition or tendency to do so.

⁹ See (Byrne, 2008).

alternative. One such alternative uses the notion of rationality as a relation between different attitudes of one and the same subject. It is irrational for instance to believe that p and to believe that p is not the case. If one finds oneself in such a situation, one should aim at resolving the conflict. The rationality requirement offers a reason to drop one of the two beliefs. But it does not suffice to decide which of the two beliefs one ought to drop. One needs to look for independent reasons speaking in favour of one of the two beliefs. The situation in our case is of a similar kind, with important differences.

There is something incoherent in me believing that p and believing that I do not believe that p , although the two beliefs do not contradict each other. The irrationality of the situation becomes manifest when a subject expresses that doxastic state by declaring: ‘ p , but I do not believe that p ’.¹⁰ We surely expect a subject caught in such a predicament to adopt an alternative set of beliefs. Why is that so although there is no contradiction between Bp and $B\neg Bp$?

One might point out that Bp speaks in favour of BBp and against $B\neg Bp$. This is certainly right, but it can’t be the whole story. It is not *just* a question of Bp speaking against $B\neg Bp$. A belief is not irrational simply by virtue of there being a fact speaking against it. The belief might simply be mistaken. So even if the fact that one believes that p speaks in favour of attributing the belief to oneself, it does not suffice as such to establish the irrationality of doing the opposite. One might naturally respond that Bp and BBp do not stand in the same relation as p and Bp . After all: by judging that p one is not simply creating a reason for attributing the belief to oneself. One *holds* that very reason. It might indeed be argued that if I *have* a reason to judge, then I am rationally required to judge (or for that matter: to act) on its basis. To have a reason to judge that p means to see (to be aware of the fact; to have reflective access to the fact) that one ought to judge that p . One should thus feel the obligation to do so.

This might be accepted. But the kind of incoherence that becomes manifest in the assertion of ‘ p , but I do not believe that p ’ cannot be fully explained in those terms. Contrast our case with the case of one perceiving that p and believing that p on its basis. If one’s

¹⁰ This of course is related to what is often called Moore’s paradox (see (Moore, 1993, p. 209)). The idea that the paradox has something to do with rationality is expressed for instance in (Shoemaker, 1995, p. 225.) and (Moran, 1997, p. 144). More about my view on this in (Soldati, 2014, p. 175 ff).

perceptual experience provides a reason for judging that p , then by perceiving that p one is rationally required to judge that p . That reason can nevertheless be trumped by considerations speaking against it. This alone does not modify the nature of the perceptual experience.¹¹ But precisely this happens in our case. It is not possible for one to wholeheartedly judge that p and yet find a reason to doubt that one judges that p . If one finds a reason to doubt that one is judging that p , then it cannot further seem to one that one is judging that p . Consider an example.

Suppose I am standing on the peak of a mountain and I feel the pouring rain sliding down my shoulders. I see and feel the falling rain. In the light of this experience I judge that it is raining. My belief is formed in response to my full appreciation of the fact that it is raining. Can I come to judge that I do not judge that it is raining while fully appreciating the fact that it is raining? It seems that the only way for me to form that second order belief is that I somehow stop to fully appreciate the fact I am experiencing. But if I do that, then my first order judgement will sag as well. I simply cannot wholeheartedly embrace a reason for judging that p and find a reason to doubt that I judge that p . The conjunction of Bp and $B\neg Bp$ is unstable in face of fully appreciated evidence.

The root of the incoherence under consideration rests neither in the simple fact that the belief that p speaks in favour of the belief that one believes that p , nor in that fact that by believing that p one acquires a reason to form the belief that one believes that p . The alternative hypothesis is that by believing that one believes that p one ought to be responding to the very same reason to which the first order belief is sensitive. In the conjunction of Bp and $B\neg Bp$ one reacts to one and the same reason by taking incompatible commitments. This is a specific form of irrationality. It is not the case of having beliefs with contradicting contents. Nor is it the case of having a reason and not forming the belief it commands. It is rather the case of having conflicting attitudes towards one and same reason. Rationality does not only require one to avoid having contradicting beliefs, or to

¹¹ This is sometimes glossed by saying that the fact that one can find grounds for doubting that p does not remove the possibility of it perceptually seeming to one that p . I think that this is not the proper way of describing the situation, since I do not think that perceptual experiences have correctness conditions as usually understood.

form the beliefs that are appropriately commanded by the reasons one has, it also demands a cogent sensitivity to one's reasons.

The difference between these various rational requirements becomes manifest when one considers the routes of their fulfilment. Consider first the rational instability generated by the conjunction of Bp and $B\neg p$. We saw that rationality demands to revise one of the two beliefs, but rationality alone does not determine which of the two contradicting beliefs one ought to drop. One *needs* to look for further reasons. This is different in the case of one having a reason speaking in favour of a belief. If one is offered a ground for forming a belief one would rather not have, one cannot reason from the undesirability of the belief to the absence of the ground. The fact that I'd rather not believe that I just lost my front tooth does not discredit the fact that it appears to be missing when I look at myself in the mirror. One can, however, flood the perceptual experience with evidence speaking against its authority. This undermines the reason provided by the experience without modifying the experience itself. The relation between Bp and BBp is like the latter in the sense that simply by judging that p one has all one needs to be rationally required to judge that one judges that p . One cannot reason from the absence or undesirability of BBp to the rejection of Bp . Contrary to the case above however, one cannot discredit the reason provided by the first order judgement without corroding that very judgement. If one has enough reasons to doubt that one is judging that p , then one is not only questioning the authority of the judgement, one is spoiling the judgement itself.

When one judges that p one has a reason to judge that one judges that p not simply because one judges that p , but because the reason one is thereby responding to rationally commits one to the second-order belief. This is the source of the transmission of the reason from one's judgement that p to the judgement that one judges that p .

§6

We were wondering what conclusions one can draw about the nature of phenomenal mineness from the application of transparency. I said that if mineness is an aspect of experience, then transparency

would entail that I can come to know that aspect by attending to aspects of the world at large. I can thus come to know that my act of judging is mine, i.e. that I am judging, by attending to aspects of the world at large. The asymmetry between first personal and third personal knowledge should not be accounted for in terms of the first person attending to something other than to aspects of the world, but rather in terms of which different aspects of the world she would need to be attending.¹²

In the light of the considerations made above about the sources of transparency, we may now conclude that the difference between first personal and third personal knowledge about one's own experiences lies in the fact that the experiences themselves belong to the aspects of the world one must take into consideration in order to attain third personal knowledge, but not so in the case of first personal knowledge. The same point can be expressed by saying that the experience itself constitutes evidence for the attribution of the experience in the third person, but not so in the first person. The experience delivers a reason for the self-attribution, but it is not itself the reason one responds to in self-knowledge.

On this understanding of transparency, phenomenal mineness might simply be understood as the kind of rational commitment one acquires through an experience. In this sense Kriegel and Zahavi's compatibility claim would be confirmed. For this to be the case, however, we do not need a phenomenological fact in addition to the metaphysical fact Kriegel and Zahavi were rejecting. On the contrary. The idea that we should accept an additional phenomenological fact in order to explain the asymmetry between first and third personal knowledge about one's own experiences is a symptom of the fact that one looks for further evidence after having realised that the simple presence of the experience did not suffice. The mistake was there from the beginning: it is the consequence of not having properly appreciated the rational roots of transparency.

¹² In his paper in this volume Howell writes that «while it seems to make sense to project the phenomenal character of redness onto a car—so that the subjective features of redness don't appear subjective but appear as a modification of the car—it doesn't seem plausible that we project the subjectivity of sensations. What would it be, after all, to project the mineness or for meness of a sensation? Is it that something out there—like the redness of the car—presents itself as belonging to me, or even some random conscious subject? This is implausible». In the case of sensations, the point may seem plausible. It is less plausible in the case of beliefs. The judgement that *p* presents *p* as *my* reason to attribute the judgement to myself.

§7

The rationalistic account sketched so far is generally confronted with an objection concerning its applicability to experiences that appear not to respond to reasons. We seem to be able to attribute such experiences to ourselves, but the rational mechanics described above would not apply to them. Should we then concede that at least for those cases the mineness link applies?

A proper reply to such an objection would obviously need an argument that goes far beyond the scope of the present paper. Let me however indicate the two main directions such an argument might go. One ought to consider first the fact that the realm of consciousness that is responsive to reason is wider than one might think. Second, one might arguably distinguish between experiences that constitute the core self, which are sensitive to reasons, and experiences that are attributed to the self in the light of their relation to the core self. I shall leave the discussion of the second point for another occasion. In what follows I shall sketch the main lines of the first one.

Let us thus consider the extent to which consciousness is responsive to reasons. We dealt above with the case of judgement which might be taken as a prototype of an attitude that is responsive to reasons. What about intentions to act, desires and other conative attitudes? In judging that p one naturally responds to something in the light of which p appears to be true. The reason for judging that p is provided by one's awareness of something speaking in favour of p 's truth. Similar considerations apply, *mutatis mutandis*, to intentions to act and to desires. In forming the intention to act, one naturally responds to something in the light of which the action ought to be performed. The reason for intending to do something is provided by one's awareness of something speaking for the importance, the utility or simply the value of the action. A desire is not simply a brute urge for an action. It involves a sense for the desirability of the latter's outcome. In desiring p one forms an attitude in the light of p 's desirability. There is something about p that constitutes one's reason to desire p .¹³

¹³ For reasons of simplicity, I shall assume that one typically desires p , where the latter is some fact or state of affairs. Nothing in what follows should depend on this specific assumption, as opposed to the view that one can (also) desire a particular.

These are some initial characterisations of the way desires and intentions to act respond to reasons. Many more details would be needed in order to provide the beginning of a proper account. One might urge, however, that even granting these initial characterisations, there is a fundamental difficulty that prevents the extension of the rationalistic account to conative states. The point concerns the fact that conative states, contrary to cognitive states, cannot be true or false. Transparency, it might thus be expected, cannot be applied to them in the same way as it applies to judgements.

Let me concentrate on the case of desire. P 's desirability speaks in favour of the *belief* that p is desirable by letting the belief that p is desirable appear true. The relation of p 's desirability to the desire itself cannot be of the same kind. There is another relation instead. P 's desirability speaks for the desire for p not by letting the desire appear true, but by letting it appear appropriate, or fitting. In desiring p one experiences p as having the kind of features that make it seem right for one to desire it. To desire p is to experience p under the guise of the good, however bad one may judge p to be under a different light. To experience p under the guise of the good is not to experience p 's being good under the guise of the true. If this is right, if the reason of one's desire doesn't speak for its truth, then one may wonder how it could speak for the truth of its self-attribution. The point can be made in more general terms. Consider a situation where one desires p (Dp) and judges that one desires p (BDp). For the rationalistic account of transparency to apply, the reason provided by Dp should somehow be transmitted to BDp . But if the reason for Dp doesn't speak for the truth of Dp , how could it possibly speak for the truth of BDp ? If at all, one might surmise, it could speak for the appropriateness of DDp .

A further but related problem appears to threaten the application of transparency to conative states. Remember the point we made above about belief. I argued that by maintaining Bp and $B\neg Bp$ one reacts to one and the same reason by taking incompatible commitments that manifest the absence of a cogent sensitivity to reasons. If one has enough reasons to doubt that one is judging that p , then one is not simply questioning the authority of the judgement, one is spoiling the judgement itself. We contrasted the case of belief with the case of perception, by showing that one can undermine the

perceptual experience with evidence speaking against its authority without modifying the experience itself. Desire resembles perception in this perspective. One certainly feels the pressure to adjust one's desire when one finds it unfitting. But one may be sceptical about the reasons that seem to make one's desire appropriate without thereby destroying it. One does not need to suppose an instance of amorality in order to acknowledge the phenomenon.

There is one central point where the analogy appears to limp. When a desire presents p under a positive light, it typically does so with respect to some, but not all its features. These are *evaluative features*, features that are responsible for the evaluative response, whether conative or cognitive. The desire to smoke a pipe presents smoking under a positive light with respect to the taste of tobacco, or with respect to the stimulating effect of nicotine. The belief that smoking is not desirable is typically formed in the light of considerations concerning health hazards. The positive light the desire sheds on smoking bears on features that are not at stake when one judges that smoking is undesirable. In such cases the tension between desire and belief is generated by the fact that the features in the light of which p is desired are not identical to those in the light of which it is judged to be contemptible. Let us call this a situation of *external tension* between desire and belief. One is forced to deliberate between conative evaluative features of one kind and cognitive evaluative features of another kind.

When we find a desire inappropriate, we may wonder how it can be sustained. The desire may indeed respond to a feature we are not able to articulate. One may desire p without being able to say what one finds desirable about p . The apparent incoherence might here be rooted in the fact that we simply do not know the feature in the light of which p is desired. We may know that F makes p undesirable, but we do not know in the light of what we desire p . Confronted with a person who desires some p we take it to be disagreeable, we naturally wonder which features of p she experiences as valuable. This is only confirmed by the fact that the discovery of the features that make p desirable is generally prompted by the desire, rather than the reverse. We experience the desire as an indication of there being something desirable in p , even if we don't know what it is.

The question can now be raised as to the proper description of a situation where one judges that p is repulsing in the light of the very same evaluative features under which it is desired. Can one both desire to smoke in order to enjoy the taste of tobacco and sincerely judge that the taste is repulsing?¹⁴ One may naturally judge that being F makes p desirable and yet fail to desire p even in the light of its F -ness. The issue concerns the possibility of one wholeheartedly judging that F makes p undesirable, and yet experiencing one's desire as appropriate with respects to p 's F -ness.

This situation should again be distinguished from the case of one desiring p for the very sake of its badness. Perversion may be the condition that is required for such a desire to be possible. It is the satanic case of one desiring p in the light of the evil rather than the good.¹⁵ Even if conceptually possible, we have a clear sense of the deep disturbance of such a conative frame of mind. I shall ignore this case in what follows. It needs a separate treatment.

Let us then concentrate on the situation where one's desire presents p 's F -ness in the light of the good and one yet judges it to be bad in the light of the very same features. It is a situation of *internal tension* between belief and desire. The situation ought to be distinguished from the simple presence of contradictory evaluative judgements, such as the judgements that F -ness makes p good and the judgement that it makes it bad. The central difference concerns the available procedures one may adopt in order to solve the tension. In the case of contradicting evaluative beliefs, the standard assumption would be that one of the two beliefs must be false. One would thus look for further reasons speaking for one of the beliefs and against the other. (Or alternatively look for different standards of evaluation). The evaluative judgements respond to reasons by making them available for deliberation. Suppose for instance that one ends up believing both that p is good and p is bad. This is a flat contradiction. Unless one of the beliefs is simply false, the default assumption is that the contradiction can be solved by finding distinct features that speak for both beliefs.

This is not what happens in the case of desire. One does not

¹⁴ As mentioned above, I am assuming that the desire to smoke is not a sheer pull and that it is not fundamentally concealed to one's conscious awareness.

¹⁵ Cf. (Velleman, 1992).

typically look for a reason speaking for or against one's desire. One rather experiences one's desire as providing one with a reason for the corresponding evaluative judgement. The case is particularly vivid in situations of internal tension. If I desire p I may wonder about the feature under which p appears desirable. I may wonder why I want to smoke and find that it is the taste of tobacco which makes smoking desirable.¹⁶ But to experience one's desire for p as fitting with respect to p 's *F-ness* is simply to have a reason to believe that there is something good about p - its *F-ness*. To desire p and to judge that p is bad, with respect to one and the same evaluative feature, would be to deliberately form a belief in flat opposition to the only available reason.¹⁷

Can a desire be questioned in the light of considerations speaking against its authority without spoiling the desire itself? The considerations above have shown that this is possible when there is an external tension between desire and evaluative belief. It is much less obvious that this can happen in case of an internal tension. In order to question the authority of one's desire for p , one must find a feature in the light of which p is judged to be bad. This would inevitably transform the internal tension into an external one.

The considerations above offer a first basis for a rationalistic account of the self-attribution of desire. The self-attribution of a desire and of the corresponding evaluative judgement respond to the same evaluative features. The same reason is used in both cases. The relevant difference between the evaluative judgement and the corresponding desire lies in the way they respond to evaluative features. The judgement makes them available for deliberation, the desire does not. This has consequences for their self-attribution. According to the rationalistic account, in judging that I judge that p is good I respond to the same reasons to which I respond in judging that p is good. Reasons against attributing the judgement to myself jeopardise the judgement itself. The application of the rationalistic account to desires starts from the assumption that the desire responds to the same reason as the corresponding evaluative judgement. The

¹⁶ I might thus describe the taste of the tobacco as being the reason of my desire.

¹⁷ One may doubt that such a belief can be formed at all. If, all things considered, the desire for p offers the only available reason for the belief that p is good, then one would have to *decide* to believe that p is bad. It is doubtful that this corresponds to a psychological possibility.

distinctive feature of desire is that it presents that reason in a way that leaves no room for deliberation. I thus attribute a desire to myself when I am presented with a conclusive evaluative reason. There is transmission of reason in the sense the very reason to which my desire responds is used in the self-attribution of the desire. The desire itself does not constitute additional evidence. To experience a conclusive evaluative reason in a desire is enough for one to attribute the desire to oneself. It should be clear that the claim that a desire presents a conclusive evaluative reason does not imply that one cannot deliberate about one's desires. The point is that such a deliberative process involves a move from an internal to an external tension. It must concern different evaluative features.

The suggested account faces various objections. Some are internal to the account. One may wonder, for instance, whether the present account leads to a view that treats desires as if they were perceptions of evaluative facts. I think that there is something to this idea. Desires and perceptions are both instances of one being directly aware of a fact. They should not be analysed in representational terms. In both cases one should distinguish the fact one is aware of from the way the fact is articulated in belief. Errors in the conceptual articulation must not be attributed to the nature of the awareness. Both perception and desire offer conclusive reasons. Desires, in opposition to perceptions, deliver evaluative reasons. Both perceptions and desires are attributed to oneself not on the basis of their phenomenal mineness, but on the basis of the way they make reasons available to the subject.

A different sort of objection would insist on the idea that some desires are experienced as sheer pull. One fully acknowledges that it is bad to φ , but one still feels a pull to φ . I am supposing that this is not a case of experiencing φ ing under some positive light. As such, I would rather not call it a desire. It is more like an urge to scratch one's head. If this is what is meant, then I would suggest treating such experiences along with other experiences, such as itches, that are only derivatively attributed to oneself. I attribute them to myself by virtue of their relation to experiences I can directly attribute to myself. But this, as I said, is a topic for another occasion.¹⁸

¹⁸ I should like to thank two anonymous referees. Their comments have been greatly helpful.

References

- Boyle, M. (2011) 'Transparent Self-Knowledge', *Aristotelian Society Supplementary Volume*, 85(1), pp. 223–241.
- Byrne, A. (2008) 'Knowing that I am thinking', in Hatzimoysis, A. E. (ed.) *Self-Knowledge*. Oxford University Press.
- Byrne, A. (2011) 'Transparency, Belief, Intention', *Aristotelian Society Supplementary Volume*, 85(1), pp. 201–221.
- Davidson, D. (1987) 'Knowing One's Own Mind', *Proceedings and Addresses of the American Philosophical Association*, 60, pp. 443–458.
- Davidson, D. (1988) 'The Myth of the Subjective', in Benedikt, M., R., B., and (eds.) (eds) *Bewußtsein, Sprache und die Kunst*. Wien: Verlag der österreichischen Staatsdruckerei, pp. 45–54.
- Evans, G. (1982) *The Varieties of Reference*. Oxford: Oxford University Press.
- Frank, M. (2002) *Selbstgefühl: Eine historisch-systematische Erkundung*. 2nd edn. Frankfurt am Main: Suhrkamp Verlag.
- Henrich, D. (1967) *Fichtes ursprüngliche Einsicht*. Frankfurt am Main: Klostermann.
- Kriegel, U. and Zahavi, D. (2015) 'For-Me-Ness: What It is and What It is Not', in Hopp, D. D. and A. E. and W. (ed.) *Philosophy of Mind and Phenomenology*. Routledge, pp. 36–53.
- Moore, G. E. (1993) *Selected Writings*. Edited by T. Baldiwn. London: Routledge.
- Moran, R. (1997) 'Self-Knowledge: Discovery, Resolution, and Undoing', *European Journal of Philosophy*, 5, pp. 141–161.
- Moran, R. (2001) *Authority and Estrangement*. Princeton: Princeton University Press.

Moran, R. (2012) 'Self-Knowledge, "Transparency", and the Forms of Activity', in Smithies, D. and Stoljar, D., *Introspection and Consciousness*. OUP Oxford, pp. 211–238.

Ryle, G. (1949) *The Concept of Mind*. London: Hutchinson & Company.

Shoemaker, S. (1995) 'Moore's paradox and self-knowledge', *Philosophical Studies*, 77(2–3), pp. 211–28.

Soldati, G. (2014) 'Prospects of a Deflationary Theory of Self-Knowledge', in Hügli, A. and Horn, Anita, *Die anthropologische Wende*. Basel: Schwabe Verlag (Studia philosophica), pp. 169–188.

Velleman, J. D. (1992) 'The Guise of the Good', *Noûs*, 26(1), pp. 3–26.

Zahavi, D. (2010) 'The Experiential Self: objections and clarifications', in Siderits, M., Thompson, E., and Zahavi, D. (eds) *Self, No Self?: Perspectives From Analytical, Phenomenological, and Indian Traditions*. Oup Oxford, pp. 56–78.