

## Prospects of a deflationary theory of self-knowledge

### Abstract

Following a certain transparency claim, evidence concerning the external world can be used in order to ground the attribution of a belief to oneself. This epistemic privilege arguably depends on the applicability of norms of rationality that are manifest in paradoxes of the Moorean kind ('It is raining but I don't believe it'). In this paper I clarify how these norms of rationality apply to the self-attribution of belief and I determine the features by virtue of which they deliver genuine epistemic warrants. A central issue in this debate depends on the way experiences provide the subject with reasons to form a belief. It is useful in this respect to compare the self-attribution of judgements, of acts of judging, to the self-attribution of other kinds of experiences, such as perceptions and desires.

### §1. *Attributing a thought to oneself*

In a full account of self-knowledge (i.e. knowledge in the first person about oneself), the attribution of one's own thoughts to oneself (self-attribution of thoughts) plays an important role. Three considerations are relevant in this respect.

First, it is a domain where the *epistemic authority* of the subject is particularly manifest. I know what I think without having to rely on the kind of evidence you would typically have to rely on in order to know what I think. It has been important to see that this kind of epistemic asymmetry does not require my knowledge about my own thoughts to be infallible. And it has further been noticed that this epistemic asymmetry should not imply any semantic asymmetry, where the latter would involve the idea that there is some fact, proposition or other kind of content involving the subject, such that it would be a unique privilege of the subject to be in a position to grasp or express it.<sup>1</sup>

The self-attribution of one's own thoughts secondly plays an important role in the determination of the concept *self*. If usage of the expression 'I' is governed by a rule that states that its utterer thereby refers to herself, the thought expressed by an utterance containing that expression must be more fundamental. One arguably experiences oneself as the origin, as the (mental) agent and possessor, of the thoughts one produces. If that is so, then the concept *self* might (at least partially) be the concept of the producer of the very thought one is thinking when one makes a self-attribution. In any first person thought about myself I would think of myself as the producer of that

---

<sup>1</sup> More about this in Gianfranco Soldati: *Self-Knowledge, Referential Immunity and the Concept Self*, in: Michele Di Francesco et. al. (eds.): *Filosofia Analitica 1996-1998. Prospettive teoriche e revisioni storiografiche* (Milano: Guerini, 1998), pp 192–206.

very thought. This leaves open the question as to whether there is an even more fundamental, non-conceptual way I may be given to myself.

Finally, the attribution of a thought to oneself plays a central role with respect to the way *reasons* are made available to the subject. The thought that it is raining makes a reason available for my further thinking (I should take an umbrella) and for my action (I take an umbrella) not only by virtue of being my own thought, but also by virtue of me being in a position to attribute that thought to myself. It surely would appear questionable for me to take the umbrella if I did not even have a tendency to judge that I think that it is raining.

In all these cases there is a question about the epistemic basis of the self-attribution of one's own thoughts. And that question often concerns the role, if not the very possibility, of an introspective access to our own thoughts. It has been wondered, for instance, if the epistemic authority mentioned at the beginning is the result of a special, introspective access to our own thoughts. In a similar direction, one may wonder whether the possession conditions of the concept *self* would depend on the subject being able to apply it on the basis of her introspective awareness of her own mental agency. Finally, it might be conjectured that for me to have a reason to judge or act, I need to be introspectively aware of the thought that provided me with the relevant reason.

Introspective access to our own thoughts would play an essential role if all these questions were to be answered in the positive. A number of doubts, however, have been formulated in this respect. It appears, for instance, that the authority one has over one's own thoughts does not depend on the evidential role of a supposedly introspective access. And it might be wondered whether the possession conditions of the concept *self* really depend on one's ability to identify oneself in introspection as the author of the very thought one is thinking. It does not seem right, finally, that for me to use a reason provided by one of my thoughts, I need to have introspective evidence for the claim that the thought is indeed mine rather than someone else's.

Now, whatever the merits of these familiar doubts about the epistemic role of introspection, one would not have to abandon the belief in the centrality of the self-attribution of one's own thoughts with respect to self-knowledge if an alternative to the introspective model could be offered. Deflationary models may indeed be used in order to preserve that centrality without having to adopt the standard model of introspective knowledge. What follows is a critical assessment of the prospects of one such deflationary approach.

Two remarks before I start may be in order. First, I consider *one* deflationary approach and wish to leave the assessment of other deflationary approaches for another occasion. I do not think that the results I shall present in this paper generalize to all other deflationary approaches without substantial refinements. Second, the deflationary approach I am inquiring into is not meant to offer a model for all kinds of self-knowledge. There are important and complex relations between

different kinds of self-knowledge. A proper description of those relations has to be left for another occasion. By accepting that a deflationary analysis applies to some cases of self-knowledge, one does not thereby suggest that it applies to all cases of self-knowledge. More specifically, suggesting an alternative to introspective knowledge for some cases of self-knowledge does not commit one to denying the very possibility of introspective knowledge of one's own experiences. If, however, knowledge of one's own thoughts plays the kind of role mentioned above, then the deflationary approach, if successful, would provide more than a marginal contribution to our general understanding of self-knowledge.

## §2. *The deflationary model introduced*

The starting point of the deflationary account I wish to examine is what I take to be an intuition about a certain method a subject can use in order to settle the question as to what she thinks. The intuition has notoriously been expressed in these words:

...in making a self-ascription of belief, one's eyes are, so to speak, or occasionally literally, directed outward—upon the world. If someone asks me 'Do you think there is going to be a third world war?' I must attend, in answering him, to precisely the same outward phenomena as I would attend to if I were answering the question 'Will there be a third world war?'.<sup>2</sup>

... with respect to the attitude of belief, the claim of transparency tells us that the first-person question 'Do I believe P?' is 'transparent' to, answered in the same way as, the outward-directed question as to the truth of P itself.<sup>3</sup>

The intuition about transparency of belief self-ascription has to be distinguished from other well-known kinds of transparency intuitions, such as the transparency of experience.<sup>4</sup> It is an intuition in the sense that it rationally appears right to claim, as Evans and Moran do, that one can answer questions of the kind 'do I think that *p*?' by simply considering the question '*p*?' The intuition is based on our understanding of the concepts involved in the description of the situation.

---

<sup>2</sup> Gareth Evans: *The Varieties of Reference* (Oxford: Oxford University Press, 1982), p 225.

<sup>3</sup> Richard Moran: *Authority and Estrangement* (Princeton: Princeton University Press, 2001), p 66.

<sup>4</sup> The two notions of transparency are well established in recent literature. Transparency of experience is discussed, with respect to perception by Michael Martin: The Transparency of Experience, In *Mind and Language* 17(4) (2002), pp 276–425; Charles Siewert: Is experience transparent?, In *Philosophical Studies* 117 (2004), pp 15–41; Michael Tye: Representation and the transparency of experience, In *Noûs* 36(1) (2002), pp 137–151; and with respect to emotions by Michael Tye: The Experience of Emotion: An Intentionalist Theory, In *Revue Internationale de Philosophie* 62 (2008), pp 25–50; Gianfranco Soldati: Transparenz der Gefühle, In *Leben mit Gefühlen. Emotionen, Werte und ihre Kritik* hg. von Barbara Merker (Paderborn: Mentis, 2009), p 257-280. Transparency of belief is at issue, for instance, in Alex Byrne: Transparency, Belief, Intention, In *Aristotelian Society Supplementary Volume* 85(1) (2011), pp 201–221 and Brie Gertler: Self-Knowledge and the Transparency of Belief, In *Self-Knowledge* Anthony Hatzimoysis ed. (Oxford: OUP, 2011), p 125-45. Since this paper addresses the latter kind of transparency, I shall use the term 'transparency' in order to speak about the transparency of belief.

It is based neither on a sensory appearance (actual or imaginary) of that very situation, nor on explicit empirical assumptions about the kind of questions the subject would be inquiring into.

A philosophical assessment of the transparency intuition requires a clarification of its theoretical plausibility and of the domain of its application. Here are some of the questions that ought to be addressed in this respect. How can we more explicitly articulate the transparency intuition? What are its basic conceptual ingredients? Is it an intuition about the nature of beliefs, or about the justification of the self-attribution of beliefs, or about something else? Does the transparency intuition apply to other mental states or experiences apart from beliefs? And if yes, by virtue of which feature that those states and experience would have in common with beliefs? Once the transparency intuition has been properly articulated and theoretically evaluated, and once its domain of application has been determined, it should become clear in what sense exactly it is supposed to provide an alternative to the introspective model. Suppose that following the latter model the subject has an introspective warrant for the self-attribution of a belief. The question then is whether the deflationary model is supposed to offer an alternative warrant or whether it is rather supposed to deflate the very need for any such warrant, whether introspective or not. Finally, it would be crucial to understand how the deflationary model is supposed to apply to the three domains I mentioned at the beginning. Can the deflationary model do justice to the epistemic asymmetry without destroying the semantic symmetry? Does it better explain why the concept *self* is the concept of the producer and (or) possessor of the very thought one is thinking? And finally: how does the deflationary model account for the way experiences provide us with reasons?

I shall not be able to address all those questions in this paper.<sup>5</sup> I shall describe a line of thought that looks promising to me. The general issue I am interested in, here as well as in a variety of other contexts,<sup>6</sup> concerns the role experiences play with respect to belief and knowledge. In the introspective model experiences, and the way they are given to us in introspection, play a crucial role. This is supposed to be so especially in the case of the self-attribution of belief. If I know by introspection, on the basis of inner evidence, that I believe or judge that *p*, then there should rather

---

<sup>5</sup> I will not be able to say much, for instance, about the different domains of application of the transparency intuition. It has been argued, for instance, that transparency does not apply to practical intentions (see Jonathan Way: Self-knowledge and the limits of transparency, In *Analysis* 67 (295) (2007), p 223–230). Some have wondered about how to apply it to desires (Jordi Fernandez: Desire and self-knowledge, In *Australasian Journal of Philosophy* 85 (4) (2007), p 517–536; Lauren Ashwell: Deep, dark...or transparent? Knowing our desires, In *Philosophical Studies* 165 (1) (2013), pp 245–256. One may further wonder whether I cannot have an unwarranted belief about the external world and still be warranted to attribute the belief to myself. As I said at the beginning, the approach articulated in this paper does not depend on the assumption that transparency applies to all mental states and experiences.

<sup>6</sup> For instance regarding thought in general Gianfranco Soldati: Begriffliche Qualia. Zur Phänomenologie der Bedeutung, In *Anatomie der Subjektivität. Bewusstsein, Selbstbewusstsein und Selbstgefühl*, Thomas Grundmann et al. (eds.) (Frankfurt: Suhrkamp, 2005), pp 140–168 and in relation to demonstrative thinking more specifically: Gianfranco Soldati: Die Rolle der Wahrnehmung in demonstrativen Gedanken, In *Husserl und die Philosophie des Geistes*, Manfred Frank und Niels Wiedtmann (eds.) (Frankfurt: Suhrkamp, 2010), pp 112–133.

be something it is like for me to believe or judge that  $p$ .<sup>7</sup> The question would then be whether the deflationary model makes such an appeal to experience in the account of self-knowledge superfluous. My view is that it does not. The fact that we do not need an evidential introspective warrant for the self-attribution of a thought does not imply that experiences do not play any role with respect to self-knowledge. My main contention is that they do so by virtue of making reasons manifest in consciousness. Without such reasons being manifest in consciousness the deflationary account would not get off the ground.<sup>8</sup>

### §3. *Articulating transparency*

Let us start by looking at the transparency intuition. It concerns the relation between a belief (the belief that  $p$  -  $Bp$ ) and its self-attribution (the belief that I believe that  $p$  -  $BBp$ ).<sup>9</sup> In what follows I shall generally suppose that we are speaking about consciously occurring beliefs, also called 'acts of judgement'. I do think that some, but not all the considerations that follow apply to dispositional beliefs, understood as dispositions to judge. But for the purpose of this paper, and in the light of the domains of self-knowledge mentioned in §1, it will suffice for the argument to apply to acts of judgement.

The transparency intuition involves a claim that says that on some occasions one can settle the question as to whether one believes that  $p$  by addressing the question as to whether  $p$ . But what does it mean in this context to 'settle the question'? If I want to settle the question as to whether  $p$ , then I am arguably looking for something that would make it appear right for me to answer the question in one way rather than another. Something that makes it appear right to answer a question in a certain way is a reason for answering the question in that way. But not any kind of reason will do. In the context we are considering, the issue does not concern primarily what it would be right to *say*, but rather what it would be right to *believe*, and then to say, with respect to the question as to whether  $p$ . It is generally right to believe what is true. So the reason we are looking for is a reason that speaks for the truth of a belief about  $p$ . It is an epistemic reason for forming a certain belief about  $p$ .

Insisting on the way Evans has described transparency in the quotation above, one may object that transparency really concerns what it is right to say, rather than what it is right to believe: I find out what I should assert when I am asked whether I think that  $p$  by considering the question as to whether  $p$ . But it is easy to find cases where this procedure would be inappropriate with

---

<sup>7</sup> The contention that beliefs are precisely not experiences, that there is nothing it is like to believe something, may of course be mentioned precisely as a motivation for dropping the introspective account.

<sup>8</sup> More is provided on how reasons become manifest in consciousness in Gianfranco Soldati: *Direct Realism and Immediate Justification*, In *Proceedings of the Aristotelian Society* 112 (2012), pp 29–50.

<sup>9</sup> Beliefs, or judgements, are properties of subjects. So when I write 'Bp' this is short for 'the subject believes that  $p$ '. The symbolic abbreviation is not supposed to offer the logical form of the sentence attributing a belief.

respect to what it is right to say and yet appropriate with respect to what it is right to believe. I should rather not address the question whether my boss is incompetent if she asks me whether I think that she is incompetent. I should rather ask myself whether she is going to fire me, if I tell her that I think that she is incompetent. But this does not mean that I consider the danger of being fired as a reason not to believe that she is incompetent and therefore as a reason for believing that I do not think that she is incompetent. The reason one finds for saying something is not always a reason one would use for believing something. Transparency involves primarily a claim about what it is right to believe, and only secondarily a claim about what it is right to say, namely under the assumption that nothing speaks against being sincere.

The transparency intuition relies on the fact that I can find the reason that settles the question as to whether I believe that  $p$  by looking for a reason I can use in addressing the question about the truth of  $p$ . So under some conditions what speaks for the truth of  $Bp$  also speaks for the truth of  $BBp$ . If what speaks for  $Bp$  is  $Bp$ 's epistemic warrant,<sup>10</sup> then, under some condition,  $Bp$ 's epistemic warrant is also  $BBp$ 's warrant. On this understanding the transparency intuition can be articulated as involving the claim that under some conditions  $Bp$ 's epistemic warrant can be transmitted to  $BBp$ . The transparency intuition would contain the contention that under some conditions a warrant for a belief can be transmitted to the self-attribution of that very belief.<sup>11</sup> I consider this contention to be part of what one obtains in providing a more explicit articulation of the transparency intuition.

#### §4. *The theoretical basis of transparency*

Now we need to understand the theoretical basis of this claim and evaluate the original domain of its application. To understand the theoretical basis of an intellectual intuition means, among other things, to determine the concepts the intuition is about, and the relations between the concepts that are relevant for the intuition. To evaluate the original domain of application means to provide independent theoretical considerations that explain the conceptual relations the intuition is responsive to.

---

<sup>10</sup> Warrant and justification are properties of a subject forming a belief, rather than properties of the belief itself. When I speak of a belief's warrant, I thus take this to be short for a subject having a warrant for that belief.

<sup>11</sup> This kind of transmission appears to be presupposed by what Williams calls 'Evans's principle', namely the principle 'that whatever justifies me in believing that  $p$  also justifies me in believing that I believe that  $p$ ' (see John N. Williams: Moore's paradox, Evans's principle, and iterated beliefs, In *Moore's Paradox: New Essays on Belief, Rationality, and the First Person*, Mitchell Green und John Williams eds (Oxford: OUP, 2007), p. 82-99). Evans himself stated the principle in terms of assertion, rather than belief: 'whenever you are in a position to assert that  $p$ , you are *ipso facto* in a position to assert "I believe that  $p$ "' (Gareth Evans: *Varieties of Reference* (Oxford: OUP, 1982) p 226). Whatever the relation between Evans's own view and Williams's interpretation of it, the important point in our context is that Williams suggests using this principle in order to 'solve Moore's Paradoxes'. This reverses the explanatory direction adopted in this paper. I suggest using a certain interpretation of one version of Moore's paradox in order to clarify the theoretical basis of the transparency intuition.

The transparency intuition, in the way I have articulated it above, has something to do with the concepts of belief and warrant. More precisely, it has something to do with the transmission of a warrant from a belief to its self-attribution. The relation between a belief and its self-attribution, as it has often been noticed, is the relation that is also at issue in Moore's Paradox. Let us then look at Moore's Paradox in order to see whether it can be used in order to clarify the theoretical basis of the transparency intuition.

Moore's Paradox, in its original form, is the claim that there is something paradoxical in the fact that an utterance of the form 'it is raining but I don't believe it' is absurd without containing any contradiction.<sup>12</sup> There is a debate about the exact nature of the absurdity.<sup>13</sup> Some think that the utterance after all does contain a contradiction ('p and non-p'), for instance by virtue of rules governing assertion (to assert 'I believe that p' is to assert 'p').<sup>14</sup> Others think that it is, at least in part, a consequence of a norm of rationality that governs our beliefs.<sup>15</sup> In order to be relevant in our context, we have to assume that this last interpretation is acceptable.<sup>16</sup> On this interpretation, the concept of belief is the concept of an attitude that is subject to rational requirements. Among them one finds the requirement that one ought not to believe that *p* and believe that one does not believe that *p*, even if there is nothing contradictory in doing so. Let me call this *Moore's Requirement*. A

---

<sup>12</sup> See George E. Moore: *Selected Writings* (London: Routledge, 1993), p 209 and Mitchell Green & John N. Williams: *Moore's Paradox: New Essays on Belief, Rationality, and the first person* (Oxford: OUP, 2007), p 5. Note that in what follows I shall consider only this version of the paradox.

<sup>13</sup> It has often been noted that there are circumstances where the utterance in question would not be absurd. Among such circumstances one may mention, following Wittgenstein, cases where somebody utters 'He's coming, but I still can't believe it' (Ludwig Wittgenstein: *Bemerkungen über die Philosophie der Psychologie; Letzte Schriften über die Philosophie der Psychologie* (Frankfurt: Suhrkamp, 1984), §485), or where the railway speaker announces the arrival of the train he takes to be delayed and adds: 'personally I don't believe it' (ibid. §486–7); Mitchell Green & John N. Williams: *Moore's Paradox: New Essays on Belief, Rationality, and the first person* (Oxford: OUP, 2007), p 8). With respect to beliefs one can imagine a situation where a subject discovers that she was unconsciously believing that *p* and consciously believing that she did not believe that *p*. I find it more difficult to provide convincing cases of this kind for the sort of conscious acts of judging I am presupposing here. Notice further that the examples mentioned above would not be used in order to dissolve Moore's Paradox. They either present cases of non-standard, e.g. idiomatic uses of an assertion (as in 'He's coming, but I still can't believe it'), or they help to better determine the precise domain of application of the Paradox (as in the case of unconscious beliefs).

<sup>14</sup> See Wittgenstein's famous remark in Ludwig Wittgenstein: *Bemerkungen über die Philosophie der Psychologie; Letzte Schriften über die Philosophie der Psychologie* (Frankfurt: Suhrkamp, 1984), §478.

<sup>15</sup> See for instance Sydney Shoemaker: Moore's paradox and self-knowledge, In *Philosophical Studies* 77 (1995), pp 211–228. Contrary to what I am doing here, Shoemaker does not restrict the argument to the case of conscious acts of judging (which appear to come close to what he calls 'available beliefs'). He nevertheless thinks 'that in a rational person belief that *P* brings with it belief that one believes that *P*' (ibid. p 225). More about this below.

<sup>16</sup> It must be acceptable that, as Moran puts it, 'were someone to think to himself, as he looks out the window, that it's raining outside, and conjoin this with the thought that he doesn't believe that it's raining, his thought would risk incoherence in just the same way as it would if he were to assert the whole thought to someone else' (Richard Moran: *Self-Knowledge: Discovery, Resolution, and Undoing*, In *European Journal of Philosophy* 5 (1997), p 144). I am not sure whether incoherence is exactly what the requirement is about (see my remarks on coherence below) and I doubt that the problem with respect to belief is the same as with respect to assertion. Moore's Paradox, in fact, points to a variety of problems related to an assertion of the form '*p* and I do not believe that *p*'. For the moment it suffices to establish that *some* of those problems are related to the fact that a subject forming the kind of beliefs described by Moran appears to fail to comply with a norm of rationality. The exact nature and force of this norm has to be further determined.

subject who believes that  $p$  and believes that she believes that  $p$ , finds herself in a situation that is not in conflict with Moore's Requirement.

The question I suggest to inquire into is whether the transparency of belief holds by virtue of Moore's Requirement or by virtue of norms that can be derived from it. If, as I have suggested, transparency of belief concerns the transmission of an epistemic warrant from a belief to its self-ascription, then the question concerns the relation between the legitimacy of such a transmission and the fact that by believing that  $p$  the subject puts herself under the obligation not to believe that she does not believe that  $p$ . What features of Moore's Requirement, if any, can help us to understand the transmission of a warrant from a belief to its self-ascription?

In the way I have introduced it, Moore's Requirement would be a constitutive norm of beliefs: in order for something to be a belief it must be governed by the requirement. If a subject can rationally  $\Phi$  that  $p$  and  $\Phi$  that she does not  $\Phi$  that  $p$ , then  $\Phi$ ing is not believing (but, for instance, desiring). This does not mean that it is impossible for a subject to have beliefs that fail to satisfy the requirement.<sup>17</sup> It rather means that in such a case the subject would have to be described as not complying with a requirement under which she stands simply by virtue of judging that  $p$ . The mere fact that one desires, hopes, imagines that  $p$ , instead, would not put one under this kind of requirement.<sup>18</sup>

It is a further question whether Moore's Requirement qualifies a belief in the light of justification and knowledge. A requirement that states how a mental state ought to be, in order for it to be a belief, is not yet a requirement that says what the state ought to be in order to be justified or to qualify for knowledge. So if we wish to use Moore's Requirement in our account of the epistemic basis of the self-attribution of belief, we need more than just a constitutive norm of beliefs.

These considerations have a direct bearing on our intention to use Moore's Requirement in order to understand the kind of warrant transmission from  $Bp$  to  $BBp$  we have taken the transparency intuition to be responsive to. It might well be that by transmitting a warrant from  $Bp$  to  $BBp$  a subject complies with Moore's Requirement. Indeed, if Moore's Requirement articulates a constitutive norm for beliefs, then whatever licences the kind of warrant transmission we are

---

<sup>17</sup> Shoemaker is prepared to claim that a subject's «first-order belief's being available *constitutes* her having the at least tacit belief that she has that first-order belief» (Sydney Shoemaker: Moore's paradox and self-knowledge, In *Philosophical Studies* 77 (1995), S. 211–228). The argument I shall present here is far from aiming at such a strong claim. It is a mistake to conclude from the fact that a norm is constitutive for a certain kind of mental state that it is constitutive for states of that kind to satisfy the norm. It is important to see, on the other side, that the claim that an epistemic warrant is transmitted from a belief to its self-attribution does not presuppose that the two states (or experiences) are psychologically distinct. They might be psychologically unified and epistemically distinct. One and the same belief may have, under varying conditions, more than one warrant.

<sup>18</sup> One does not put oneself under the requirement not to desire that one does not desire that  $p$  just by desiring that  $p$ . This does not mean that there is no rational requirement that applies to desires, which could be used in the elucidation of the self-attribution of desires (thus for the analysis of the relation between desiring that  $p$  and judging that one desires that  $p$ ). I have to leave this issue for another occasion (see also note 5 above).

considering should be compatible with it. But this alone would not suffice to show that the transmission itself holds by virtue of Moore's Requirement, or that Moore's Requirement can be used in order to explain why the transmission obtains. For the latter to be the case, Moore's Requirement itself would have to be related not just to beliefs, but to the role warrants play with respect to them.

There is at least one aspect of Moore's Requirement that contributes to qualifying the self-attribution of beliefs with respect to knowledge. It is the fact that Moore's Requirement is falsehood preventing. A norm that stipulates that one should not believe that one does not believe that  $p$  when one in fact believes that  $p$  is falsehood preventing. Since a belief needs to be true in order to qualify for knowledge, Moore's Requirement contributes to the epistemic qualification of the self-attribution of beliefs.

This qualification still fails to do justice to the transparency intuition we started from. We do not need Moore's Requirement in order to expect the self-attribution of a belief to be (at least) sensitive to falsehood. Falsehood-prevention arguably is a requirement that holds for all beliefs, not specifically for the self-attribution of beliefs. So, the permission to attribute a belief to oneself that is earned by virtue of the fact that one has the belief does not depend specifically on Moore's Requirement. A subject having a belief and using introspective evidence that speaks in favour of the presence of that belief would conform to the falsehood-preventing norm, but she would precisely not be using the resources that are articulated by the transparency intuition. So, if the only epistemic constraint one could obtain from Moore's Requirement is that one should not form false beliefs about one's own beliefs, then Moore's Requirement would be of little use for our understanding of the kind of warrant transmission the transparency intuition is responding to. For Moore's Requirement to be used in that context, it should involve more than a purely doxastic norm of belief (a norm that stipulates what it takes for a mental state to be a belief), and it should generate more than falsehood prevention.

It appears indeed that Moore's Requirement does not concern simply the relation between a state and its self-attribution, but more specifically the relation between a *rational* state, a state that is itself responsive to reasons, and its self-attribution. Let me explain by suggesting the following contrast. Suppose I stand under a practical requirement to do such and such if I have a bodily sensation of a certain type. Say: raise your left hand when you feel cold! Suppose I feel cold. I may still fail to find a reason to raise my hand. The requirement offers me a reason to raise my hand I would not have had in its absence. The simple fact that I feel cold does not put me under the requirement to raise my hand; it is not constitutive for me feeling cold that I stand under the requirement to raise my hand.

The context to which Moore's Requirement applies is different in various respects. One important difference concerns the fact that the antecedent in the practical requirement mentioned above does not involve any responsiveness to reasons. Feeling cold is not a state that is itself responsive to the presence of a reason. The action of bringing oneself into a situation where one feels cold may be responsive to reasons, but the simple fact that one feels cold is not. In opposition to that, believing something is a rational attitude, an attitude that is by nature responsive to reasons. So one may surmise that a requirement that applies on the basis of states that are responsive to reasons, a rational requirement, does not act on the subject complying with it in the same way as a requirement that applies to non-rational states or experiences.<sup>19</sup>

But what kind of rational requirement are we considering in this context? Some clarifications are needed before we inquire into that hypothesis. Let me then describe a number of conditions that have to be satisfied for the notion of rational requirement to apply. Let me call the kind of rational requirement that is exemplified by Moore's Requirement, a *strong rational requirement*. I shall present a number of cases where this notion does not apply and then show in what sense it applies to Moore's Requirement.

Let me first recall that for a belief to be responsive to reasons it is not needed that it is of a kind that is formed only, or even generally, as a response to reasons. As a matter of fact, most of us form beliefs on the basis of all kinds of considerations, many of which fail to provide reasons, let alone epistemic reasons, for the beliefs we form. The kind of responsiveness we are considering concerns the way one reacts to considerations one wholeheartedly takes to speak against the truth of a belief one happens to have formed. One's beliefs are responsive to reasons if, in such cases, one would be willing to revise them, whether one finally succeeds in doing so or not.

A requirement is not a rational requirement simply by virtue of applying on the basis of beliefs and by the fact that beliefs are responsive to reasons. The practical requirement I mentioned above could easily be applied to beliefs: 'Raise your arm if you believe Sam ate the cake!' Such a requirement would not only fail to be constitutive of beliefs; it would not even apply by virtue of beliefs being responsive to reasons. Nothing, in the application of that requirement, depends on the fact that the antecedent belief is responsive to reasons. For a requirement to be genuinely rational it should not only happen to apply on the basis of rational states, it should apply by virtue of them being rational. There may be various conditions a state has to satisfy in order to be rational: in the case of beliefs, responsiveness to reasons is an essential condition. We thus have to distinguish at least three cases in our context: requirements that apply to states that are not responsive to reasons; requirements that apply to states that are responsive to reasons but do not apply by virtue of them

---

<sup>19</sup> Requirements, I suppose, apply *to subjects* on the basis of certain conditions. Having a belief of a certain kind may thus be a condition on the basis of which a specific requirement applies to the subject. I shall then say that the requirement applies *on the basis* of the belief.

being responsive to reasons; and finally requirements that apply to states that are responsive to reasons by virtue of them being so. Only the latter are strong rational requirements in the sense intended in the present argument.

There are standards of rationality, such as coherence, that apply to beliefs by virtue of their representational (truth-conditional) content. It is thus possible to select beliefs in the light of their coherence. A subject's set of beliefs might be said to be more or less rational depending on the amount of beliefs that could be selected with the help of such a procedure. Rationality understood as a standard of coherence does not involve the kind of responsiveness to reasons I have mentioned above. It does not even presuppose that beliefs are responsive to reasons. They only need to have truth-conditional content. A subject's beliefs might be said to be more or less rational in this sense without the subject ever using coherence as a reason to form her beliefs. This is not the notion of rationality that is at stake in the sort of strong rational requirements I am considering in the present argument.

Suppose however that a subject considers coherence as a reason to form certain beliefs. She may aim, for instance, at forming true beliefs and think that coherence increases her chances to do so. So she takes herself to stand under the requirement to form coherent beliefs. Since coherence, as we just saw, is a form of rationality, the subject might again be said to put herself under a kind of rational requirement. Let me call it the *coherence requirement*. The coherence requirement would still fail to be a strong rational requirement in the sense I am considering here. A number of considerations are relevant in this respect. First, although the subject uses the coherence requirement as a reason to form certain beliefs, the requirement itself does not apply to her on the basis of the fact that her antecedent beliefs are responsive to reasons. It applies merely by virtue of them having a kind of content that allows them to stand in a relation of coherence to other beliefs. Secondly, and relatedly, the requirement applies to the belief forming process in a way that bears an important similarity to the way the practical requirement mentioned above applies to the subject's intentions to act. Consider again the case of the requirement to raise one's hand when one believes that Sam ate the cake. By complying with the requirement one contributes to determining Sam's responsibility with respect to the disappearance of the cake. This is, let us assume, a contribution to the establishment of social morality. One may be motivated to comply with the requirement by one's desire to contribute to the establishment of social morality and by one's belief that by raising one's hand under the given circumstances one does so. The fact that the condition for the application of the requirement is a rational state, in our case the belief that Sam ate the cake, does not impinge on the nature of the requirement. It could be any state, whether rational or not. Now, rationality functions with respect to the coherence requirement like social morality with respect to the practical requirement. It is not a part of the conditions under which the requirement applies, but

a qualification of the situation that is brought about by complying with it. That qualification may be part of what motivates the subject to form a certain belief, but it is not a feature the antecedent needs to have for the requirement to apply.

We are now ready to return to the hypothesis formulated above about the fact that Moore's Requirement concerns the relation between a state that is responsive to reasons and its self-attribution. The hypothesis was that Moore's Requirement is a strong rational requirement in the sense I have now specified. Two of the considerations made so far are immediately relevant for this claim. The first consideration concerns the fact that Moore's Requirement is constitutive for beliefs. I wrote above that if a subject can rationally  $\Phi$  that  $p$  and  $\Phi$  that she does not  $\Phi$  that  $p$ , then  $\Phi$ ing is not believing (but, for instance, desiring). This means that one brings oneself under Moore's Requirement simply by having a belief. It would thus be wrong to describe the situation in which one finds oneself when one believes that  $p$  as a situation where one might still lack a reason for believing that one believes that  $p$ . Moore's Requirement does not add a reason for one to believe that one believes that  $p$  one would not already have had by simply believing that  $p$ .<sup>20</sup>

The second consideration concerns the sense in which Moore's Requirement applies to beliefs by virtue of them being responsive to reasons. The point, remember, is not whether beliefs are responsive to reasons. The argument under consideration presupposes that they are. The question is whether them being so plays a role with respect to the fact that Moore's Requirement applies to them, as opposed, for instance, to the coherence requirement mentioned above. That requirement, we said, applies in virtue of the content of the beliefs, not in virtue of their responsiveness to reasons. This would be so even if one were to maintain that the coherence requirement articulates a constitutive requirement for beliefs. Suppose thus that if one can rationally  $\Phi$  that  $p$  and  $\Phi$  that that  $q$ , although  $p$  and  $q$  are incompatible, then  $\Phi$ ing is not believing. If this were so,<sup>21</sup> then one would have a reason not to believe that  $q$  simply by virtue of the fact that one believes that  $p$ . It would still not be necessary for this requirement to apply that beliefs are responsive to reasons. It would still suffice for them to have a kind of content that can generate coherence. This is not what is at stake in Moore's Requirement, if one accepts that there is no contradiction between  $Bp$  and  $BBp$ . The requirement does not apply just by virtue of the content of the beliefs. If Moore's Requirement does not apply on the basis of the content of the beliefs, then it

---

<sup>20</sup> This point bears a number of similarities with Kolodny's influential argument about the nature of rational requirements. Kolodny writes, for instance: '... while the fact that a rational requirement applies to one is not in fact a reason for one to comply with it, it will always seem to one, when one is subject to a rational requirement, that one has a reason of another kind to comply with it: namely, the reason to form (or drop) the attitude that, in so far as one satisfies the antecedent of the requirement, one already believes one has (or lacks). This is what gives the "ought" of rationality its normative force—or, rather, its seeming normative force' (Niko Kolodny: Why be rational?, In *Mind* 114 (2005), p 513). A proper discussion of the relation between Kolodny's point and the argument presented here must be left for another occasion.

<sup>21</sup> The idea of such a constitutive requirement for beliefs is quite implausible.

is natural to think that it applies by virtue of the kind of attitude involved. It should have something to do with the contrast between belief and other attitudes, attitudes that do not involve the same, or the same kind of, responsiveness to reasons. I already mentioned above the case of desires. A similar remark applies to a situation where one imagines that  $p$  and imagines that one does not imagine that  $p$ . It might appear psychologically difficult for one to bring oneself into such a situation, but there is nothing rationally problematic to it. It is simply a case where one has an experience and imagines not having it.

The above considerations have provided an understanding, and a number of reasons in favor of the claim that Moore's requirement is a constitutive rational requirement that applies to beliefs by virtue of the way they are responsive to reasons. The general argument is meant to show that the license one obtains for the self-attribution of a belief by the simple fact that one has the belief is grounded in that requirement. More details should now be provided about the exact mechanism that grounds the license in the requirement. Before providing such details, however, one might wish to look at the possibility of extending the strategy to other mental states apart from beliefs. As I said at the beginning, inquiring into the possibility of such an extension should help to better understand the nature and the roots of the transparency intuition we started from. It thus appears reasonable to venture some hypothesis about such an extension before returning to the details mentioned above. In what follows I shall speculate about one hypothesis in this direction. As we shall see, that hypothesis sheds new light on the relation between transparency and reasons-responsiveness. A proper integration of these insights with the results obtained so far has to be left for another occasion.

### *§ 5. Extensions and modifications*

As I said at the beginning, it should not be expected that transparency applies to all domains of self-knowledge. We should, however, expect to be able to determine criteria that distinguish the domains where transparency applies from domains where it does not. If the considerations above are on the right track, then responsiveness to reasons would be one such criterion. We have seen how that kind of responsiveness operates in the case of beliefs. A first question would thus be whether other kinds of mental states present a form of responsiveness to reasons that could ground transparency.

I suggest considering the case of perceptual experience. By perceiving something, let us suppose, I earn a warrant to form a certain belief about the external world. Can I use that warrant in order to attribute that very perceptual experience to myself? Can we say that if my perceptual experience justifies my belief that  $p$ , then it also justifies my belief that it perceptually appears to

me that *p*? And if yes, can this be grounded in the fact that perception is, in some sense, responsive to reasons?

The idea that transparency applies to perceptual experiences might be motivated by considerations along the following line. Suppose I wonder whether it perceptually appears to me that it is raining. In order to answer that question I must attend to the very same external fact I would attend in trying to answer the question whether it is raining. In order to settle the question whether it perceptually appears to me that it is raining I do not have to ‘look inside’: I have to bring my (perceptual) attention to how the weather is outside.<sup>22</sup>

The purpose of what follows is not to defend the application of transparency to the case of perceptual experiences. The point is to present a modified notion of rational responsiveness that is applicable both to beliefs and perceptual experiences. If this notion is coherent, then we would have obtained an insight into how the account of transparency formulated above could be extended beyond the case of belief. This would be a contribution to the assessment of the prospects of a deflationary account of self-knowledge.

What does it take, in general, for an experience to be *rationally grounded*? A minimal requirement might be formulated in the following terms. An experience is rationally grounded when its occurrence is the manifestation of a faculty that is normally responsive to, but not dependent on and thus not constituted by, the obtaining of an external condition. By external condition I mean a condition that does not depend on the very experience whose responsiveness is at stake, not necessarily a condition that does not depend in some other sense on the mind. A *rational commitment* is an attitude that is not only responsive to the satisfaction of some external conditions, but whose function it is to present those conditions as obtaining (rather than, say, as having to obtain). Rational responsiveness can be obtained either by a direct dependence of the attitude (which I understand as a particular exercise of a mental faculty) on the satisfaction of those conditions, or by some more indirect dependence, for instance through a relation of reliable correspondence. In both cases the attitude is warranted by either the external conditions themselves, or by the fact that there is a reliable relation to them. The first warrant is of the kind *factuality*, and the second warrant is of the kind *reliability*.

I have concentrated above on the example of belief. The belief that *p* is a rational commitment by being (among other things) responsive to the fact that *p*, either by depending on it, or by standing in some reliable relation to it. The latter case is much more plausible for most beliefs I can think of (such as the belief expressed by: ‘It is raining’). The judgment that *p* is warranted

---

<sup>22</sup> Considerations along these lines might appeal to the kind of *transparency of experience* I mentioned at the beginning of this paper. It is common, with respect to that phenomenon, to urge that one has to look at the world in order to discover the properties of one’s own experience. The claim above would need to go one step further: it would require showing that one must look at the world in order to find out *that* one has a perceptual experience.

either by the fact that  $p$  or by the fact that there is a reliable relation between  $p$  and the judgment that  $p$ . The fact that  $p$ , or the fact that there is a reliable relation between the judgment that  $p$  and the fact that  $p$ , speaks in favour of the judgment's truth. It thus constitutes a reason for judging. When the warrant is of the kind reliability, the belief does not lose its warrant by being false. Whatever contributes, in the latter case, to the reliability of the judgment, is part of the warrant. So, if my judgment that  $p$  is based on (e.g. caused by, partly constituted by, etc.) my perceptual experience, then the perceptual experience is part of what warrants my belief. This does not mean that the perceptual experience as such warrants my belief: it is part of the warrant, for instance by standing itself in a reliable relation to  $p$  and in causing the judgment that  $p$ .

Consider now perceptual experiences themselves. One's visual experience, for instance, is rationally grounded if it's occurring is the manifestation of a faculty that is responsive to the way the world visually appears in normal circumstances. And it is a rational commitment when it presents those conditions as obtaining. Suppose a red sphere is presented in my visual field under normal lighting conditions and that I have an experience of a red sphere visually appearing to me. I thereby earn a license both to judge that there is a red sphere in front of me and to judge that I am having the visual experience of a red sphere in front of me. The way the experience of seeing is grounded suffices for me to be allowed to attribute the experience to myself.

Notice that this conception is compatible with the idea that rational grounds are defeasible. Suppose that my perceptual experience is not veridical, for instance as a consequence of the fact that the lighting conditions are not normal. We would still need to make a difference between an experience being the manifestation of a faculty that is responsive to external conditions and an experience that would not be so. And it would still provide a rational ground for my judgment that there is a red sphere in front of me and for judging that I am having an experience of a red sphere visually appearing to me.

The case of perception can be used in order to highlight two central features of the account under consideration. First, the faculty whose responsiveness is accountable for rational grounding is said to manifest itself in experiences, in conscious mental events that can properly be attributed to subjects. Nothing is thus said about the possibility of the same faculty manifesting itself in the production of sub-personal non-conscious representational states. One may reasonably doubt that the same faculty can be at work in both kinds of circumstances. But even if it were, this would not imply that rational grounds would be made available to the subject in both situations. It might indeed be required that for the faculty to provide the subject with a rational ground, its response

must be manifest in a conscious experience. When this is not the case, although responsive, the faculty might be said to fail to provide the subject with a rational ground.<sup>23</sup>

The second feature can be described by attending to the following objection. One might be required to act and judge in the light of reasons, but one can hardly be required to perceive under such conditions. We think that there ought to be reasons to act, and in some sense even to judge, but it does not make sense to suppose that one has or fails to have a reason to perceive. How then can a perceptual experience be rationally grounded? In the account under consideration, the notion of a rational ground encompasses cases where ordinary language appears to prevent usage of the notion of reason. It is not clear that guidance from ordinary language yields a coherent philosophical notion of reason. We sometimes mention reasons, for instance, in relation to purely causal explanations. But the point might not be about ordinary language. One might rather insist on the fact that a perceptual experience cannot be rationally grounded in a way similar to how a judgement can be rationally grounded. The account under consideration can make room for such a distinction. A judgement based on perception *uses* a rational ground *made available* by a perceptual experience. Using a reason and making it available are indeed different relations to reasons. A judgement based on perception could not use a reason if the latter were not made available by the perception. A judgement based on perception can both use the ground offered by the perception in making a judgement about the external world and make that very ground available for a further belief, such as precisely the belief that one has that perceptual experience. The view under consideration offers a notion of rational grounding that is general enough to encompass such a possibility.<sup>24</sup>

---

<sup>23</sup> It might indeed be submitted that for the rational ground to be available, it must be poised for use in the belief. But if the experience is not conscious, it is hard to see how it could.

<sup>24</sup> I would like to thank Davor Bodrozic, Jean Bonhert, Coralie Dorsaz and Wolfgang Freitag for detailed comments on previous versions of this paper.