

**Transparency, Epistemic Impartiality, and Personhood**  
**(A Commentary on Simon Evnine's *Epistemic Dimensions of Personhood*)<sup>1</sup>**

Dorit Bar-On, UNC-Chapel Hill [dbar@email.unc.edu](mailto:dbar@email.unc.edu)

Persons, in everyone's book, form a special category of things. It is common to outline the category of persons in terms of some general features or capacities they possess – consciousness, as well as perhaps *self*-consciousness, agency, higher cognition, rationality. Like other authors, Simon Evnine, in his book *Epistemic Dimensions of Personhood*, distinguishes between the category of persons and the category of human beings. Although normal, adult human beings can be presumed to be persons, he sees no reason to rule out the possibility of nonhuman persons. Persons, as Evnine conceives of them, satisfy several necessary conditions: they are finite beings with a particular spatio-temporal location within the universe (“Finitude”); they are subjects of beliefs (“Belief”), capable of performing intentional actions and engaging in planning (“Agency”), and possessing second-order beliefs about their own and others' beliefs (“Second-Ordinality”).

In the course of his investigation into the nature of persons, Evnine teases out several surprising consequences of this conception. Of special interest is his attempt to trace out some *epistemic* dimensions of personhood that emerge. These are essential features of being a person that have specifically to do with the substantive character and structure of persons' concepts and beliefs. Very briefly, Evnine puts forth the following three theses:

- (1) To be a person, one must have certain logical concepts, such as conjunction, conditionality, disjunction, negation, and quantification;
- (2) To be a person, one must be subject to certain norms governing belief (specifically, one ought to believe the conjunction of everything one believes, and one ought to defer to one's future beliefs, unless one has special reason not to do so); and
- (3) To be a person, one must be *unable* fully to satisfy the demands of epistemic rationality

I shall not here take issue with thesis (1). At least if we accept Finitude, Belief, Agency, and Second-Ordinality as necessary conditions on being a person, the idea that persons must possess certain basic logical concepts doesn't seem too controversial. Thesis (2) is much more controversial. In connection with thesis

(2), Evidine spells out and defends two normative principles. “Reflection” is the principle that: “if one were to come to know one’s future beliefs and not learn of any reason why one should not, one should make those beliefs one’s current beliefs” (p. 109). And “Self-Knowledge” is the principle that one must have knowledge of “Ameliorism”, which is the claim that “in general, people’s beliefs get better over time (p. 111).<sup>2</sup> Though I have misgivings about these principles (even while having some sympathy to thesis (2)), I shall not discuss them in what follows.

My main focus will be thesis (3). More specifically, I am interested in Evidine’s *defense* of this thesis (offered in Chapter 6), which appeals to a phenomenon much discussed in recent years: the so-called ‘transparency of beliefs’. I think transparency is a very interesting phenomenon, but I want to question Evidine’s use of it in arguing for thesis (3). Nothing about the phenomenon of transparency, properly understood, can help support this thesis of Evidine’s. Or so I will argue. If this is right, then Evidine fails to establish one of three key theses of his book.

### **Epistemic Rationality and Personhood**

Consider two of the aforementioned ‘background conditions’ on being a person cited by Evidine: Belief (“persons necessarily are subjects of belief,” p.11) and Second Ordinality (a person must have “the ability to have beliefs about beliefs, both one’s own and other people’s,” p.15). There are all sorts of beliefs a person can have about her own beliefs, including a belief to the effect that one *has* a certain specific belief, or a more general sort of belief, that at least some of one’s current beliefs are false, or that one’s future beliefs are likely to be better than one’s present beliefs, and so on. Are there any epistemic constraints on the sorts of beliefs one can have about one’s own beliefs? Naïvely, one would expect at the very least that it should be epistemically open to a person to have any belief – including any belief about her own beliefs – that she rationally ought to have. Evidine, however, maintains that: “... belief itself – and, with belief, being something, such as a person, that essentially has beliefs – places a limit on how far a person can follow the dictates of epistemic rationality...” (p.160).

Indeed, Evidine wants to show that being a person requires being *unable* to obey at least one specific principle, the so-called Principle of Epistemic Impartiality (defended by David Christensen<sup>3</sup>):

**PEI:** The considerations determining which belief it would be epistemically rational for an agent to adopt do not give special status to any of the agent's present opinions on the basis of their belonging to the agent. (p. 141)

Evidine thinks that “PEI, and the demands of epistemic rationality, in general, are in principle unfeasible and inappropriate as an epistemic ideal for persons, or for believers in general” (p.160). This is because “PEI requires us not to make distinctions of persons in letting our beliefs be shaped by evidence” (*op. cit.*). Yet Evidine thinks that “there are limits on how far a person can see her own beliefs as the beliefs of no one in particular”; and this sets “limits to how far one can follow PEI without running up against the conceptual requirements of being a person” (*op. cit.*).<sup>4</sup>

As I understand Evidine, he thinks the limits are set by considerations surrounding the so-called transparency of belief and the related Moore’s Paradox. Paying attention to these phenomena, he thinks, supports the claim that “the verdicts of PEI... represent ways in which ... personhood is in danger of being compromised by unattenuated adherence to what is epistemically rational” (p. 153). It is this claim that I wish to question. For, as I explain below, it is far from clear to generate any straightforward conflict between PEI and transparency or Moore’s paradox.

### **‘Transparency-to-the World’ as a Distinctive Epistemic Method – Evans**

Consider first transparency as discussed by Gareth Evans.<sup>5</sup> Part of Evans’ aim in that work is to undermine Cartesianism about the mind by showing how we can abandon the idea that ordinary mental self-knowledge “always involves an *inward* glance at the states and doings of something to which only the person himself has access” (1982: 225). In a much-cited passage, he says:

[I]n 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?’ (1982: 225)

As I ascribe to myself a present belief, or thought, I direct my attention outward, not inward. I typically ‘look outside,’ at the objects my belief or thought are about, rather than inspecting goings-on ‘inside’ me. My belief self-ascription is transparent, not in the sense that I can “peer directly into” the character and content of the belief it ascribes, but in the sense that I look *through* it to the world at which it’s directed. Along similar lines, Evnine observes that we get transparency when “we can know what our beliefs are, not by looking in at the mind (...), but by looking out at the world” (p.148). This is what we may call, then, “*transparency-to-the-world*.”<sup>6</sup>

Given this characterization, I think we should go along with Evans and others and speak of transparency-to-the-world as a feature of (some) *self-ascriptions of beliefs* (that is, second-order judgments on our first-order beliefs), rather than speaking – as does Evnine – of our first-order *beliefs themselves* as “transparent to us” (p. 148). A belief may be directly *about* the world, but, unless it’s a belief *about* a belief, it can’t be said to be transparent-*to-the-world*. Evnine agrees with other authors (including Moran, whom I’ll discuss shortly) that not all beliefs about beliefs, and not even all self-ascriptive beliefs, are transparent in the relevant sense. Sometimes we ascribe beliefs as part of an attempt to explain or predict our own behavior. Thus, I might come to ascribe to myself certain beliefs on the basis of therapy, or research in cognitive science, or by observing and interpreting my own behavior. Such self-ascriptions of beliefs are *not* transparent-to-the-world. Arguably, the epistemic status of such evidential self-ascriptions of beliefs is similar to that of reports made on the basis of perception, observation, evidence, inference, conjecture, etc. For example, they do not enjoy so-called first-person authority. But even when dealing with transparent self-beliefs, I think it’s misleading to speak of the *ascribed* first-order beliefs themselves as transparent. Rather, what are transparent are the *self-ascriptions* of those beliefs.

To see why this may matter, consider the first use to which Evnine puts transparency, namely, to address Moore’s Paradox. Evnine points out that, since “I might judge that I do believe that *p* on the basis of

behavioral evidence (...) and yet judge that *not-p*”, I may be “in a position to assert a Moore-paradoxical proposition of the form ‘I believe that *p* but *not-p*’” and it will be “non-paradoxical in these circumstances precisely because the belief reported on is non-transparent to the believer” (p. 149). (Evnine makes similar remarks about the complementary case in which someone says ‘*p* but I don’t believe that *p*’.) Absurdity is presumably avoided if we assume that, while the belief self-attributed in the first conjunct (that *p*) is non-transparent, the belief directly expressed in the second conjunct (that *not-p*) is transparent. But if we take transparency to be a feature of first-order beliefs, it’s not clear what justifies this assumption. All that Evnine tells us about what renders a belief transparent for a person is that the person (and supposedly no one else) *can* know of that belief by ‘looking out to the world’. By that criterion, however, the belief self-attributed in the *first* conjunct of the Moore-sentence (i.e., the belief that *p*) is *also* transparent, since presumably I *can* know it in this way; it’s just that, as stipulated by the description of the circumstances, I *don’t* in fact know it that way. Moreover, the description doesn’t tell us how I know that I have the belief expressed in the second conjunct (*not-p*) – in fact, nothing tells us that I *know* that I have that belief at all (unless, of course, simply *having* the belief entails knowing one has it)! As regards the belief reported in the first conjunct of the Moore sentence (the belief that *p*), Evnine may suppose that what renders it transparent is the fact that the belief report is produced on the basis of evidence. But this suggests that transparency is after all first and foremost a feature of self-ascriptions of beliefs. (And again, we lack justification for taking the belief expressed in the *second* conjunct (i.e., that *not-p*) to be transparent, which means that it’s not clear how absurdity is avoided.)

In support of the idea that belief states themselves, and not just ascriptions of such states, can have transparency as a characteristic, Evnine offers the metaphorical characterization, according to which a belief is transparent insofar as it is ‘optical instrument’ or a “lens through which [one] sees the world” (p. 138).

Transparent beliefs ‘reveal the world to us’ and constitute our ‘epistemic perspective’ on the world.<sup>7</sup>

However, it isn’t at all clear how the optical metaphor is to be cashed out in a way that helps separate beliefs themselves into those that are transparent and those that are not. Assuming we’re thinking of beliefs as psychological states, rather than as the contents of such states – as *believings*, rather than *believeds* (as Sellars

would put it) – isn't it the case that *all* the beliefs one has 'reveal the world' to one? What could constitute an example of a state of belief that one is in which is nevertheless not transparent in this sense? Evnine hints at an example by mentioning a belief that is 'visible' to one, in the sense that one is able to attribute it to oneself, but where the attribution is entirely dependent on one's observation of one's own behavior, actions, etc. In such a case, one will be in a position to affirm that one believes that *p*, even though one may sincerely affirm 'not-*p*' – precisely the sort predicament one who asserts a Moore proposition may be in.<sup>8</sup> Presumably, this possibility illustrates how a belief can be non-transparent because, *assuming* the self-attribution of the belief that *p* is *true*, the self-attributer will *have* the belief that *p*, yet her insistence that *not-p* reveals that she doesn't 'see the world through' the former belief. But there is a clear sense in which, in the case as described, the person *does* 'see the world through' the belief that *p*. After all, that belief presumably guides and structures her behavior and informs her actions (otherwise we wouldn't be inclined to take her evidence-based self-ascription to be true). What else (or what more) is there to 'seeing the world through' a belief? Of course, the person is lacking *immediate* awareness of her belief, which is why she can only come to attribute it to herself on the more indirect basis of her own behavior. But it's not clear how this amounts to identify transparency as a feature of the attributed belief state itself.

Evnine does suggest that perhaps the distinction between transparent and nontransparent beliefs can be identified with the distinction between conscious and unconscious beliefs. But, as he hastens to add, "the distinction between consciousness and unconsciousness is itself contested and difficult to make out" (p. 148). And there is a difficulty in invoking the distinction in this particular case. Since beliefs are distinctively cognitive, non-phenomenal states (they do not have characteristic 'feels'), it is very unclear how to characterize the difference between a conscious and an unconscious belief state other than in terms of a person's own awareness *of* the state. But this will again lead us to draw the non/transparent distinction in terms of the presence/absence, as well as features of, *self-ascriptions* of beliefs.

Back to Evans. As I read him, Evans takes transparency-to-the-world to be a feature characteristic of those self-ascriptions of beliefs that are made on *a different epistemic basis* from evidence-based belief

ascriptions (to others or to ourselves). When issuing transparent ascriptions, we employ a *distinctive epistemic procedure*, which does not depend on recognizing or identifying (or hypothesizing about) a state inside us.

Here is how Evans describes this procedure:

I get myself in a position to answer the question whether I believe that *p* by putting into operation whatever procedure I have for answering the question whether *p*. ... We can encapsulate this procedure for answering questions about what one believes in the following simple rule: whenever you are in a position to assert that *p*, you are *ipso facto* in a position to assert 'I believe that *p*.' (1982: 225f.)

As Evans remarks, "the procedure only involves a direct consideration of the ascribed belief's *content* and requires that the subject exercise her normal abilities and dispositions for forming beliefs about the world."

Now although Evans introduces the transparency procedure in connection with self-ascriptions of beliefs, I think his observation can be generalized to other mental self-ascriptions. If asked whether I am hoping or wishing that *p*, whether I prefer *x* to *y*, whether I am angry at or afraid of *z*, and so on, my attention would be typically directed at *p*, *x* and *y*, *z*, etc. For example, to say how I feel about the upcoming holiday, I will consider whether the holiday is likely to be fun. Asked whether I find my neighbor annoying, I will ponder her behavior or character. To decide whether I am scared of the dog, I will think of the dog, and whether it is scary. To put it in terms Evnine uses, in all these cases, I typically 'look through' my mental states – feelings, emotions, propositional attitudes – to their intentional objects, as if I were to say (or think): "This (going on a holiday) is a great idea" or "She (the neighbor) is so annoying" or "It (the dog) is so scary."

Interestingly, the transparency method can allow us not only to pull apart mental *self*-ascriptions from mental ascriptions to *others*, but also to separate *mental* self-ascriptions from a special class of *bodily* self-ascriptions, namely proprioceptive reports. These latter kinds of reports do utilize a distinctive set of abilities that, we can employ only in our own case. However, the relevant abilities are different in kind from the distinctive ability we employ when issuing transparent self-ascriptions. Proprioception is the ability for own-body perception, which fits very well the materialist model of introspection as inner perception or internal self-scanning (as Evans notes, 1982: 230, fn. 42). By contrast, the procedure of "re-using the conceptual skills

which one uses in order to make judgments about the world” does not require that we postulate any “special faculty of inner sense or internal self-scanning.” (1982: 230, fn. 42) This means that, in the case of mental self-ascriptions, we have no need to exercise any special ability that is directed *either* at states of our immaterial Egos, *or* at a subset of our internal bodily states. Thus, whereas ascriptions of mental states to others (and occasionally also mental *self*-ascriptions) are based on observation of their behavior, and proprioceptive reports are arrived at by a method of internal own-body scanning, ordinary mental self-ascriptions are arrived at through the use of the transparency procedure.

Now recall that what makes for Moore’s *Paradox* is the fact that, even though Moore-sentences do *not* express logical contradictions, their sincere assertion appears *absurd*. We have now seen that there are two different ways for me to come by attributing to myself a belief that *p*: by considering *directly* (“transparently”) whether *p* is the case, or by examining on the basis of evidence (“non-transparently”) whether I *believe* that *p*. It may well be that this has something to do with the seeming absurdity of Moore sentences, as well as with the resolution of the paradox (we’ll see what in a minute). However, Evnine’s agenda is not simply to remove the seeming absurdity or dissolve the paradox. Rather, he wants to show that the *sort* of conflict illustrated by Moore’s Paradox is endemic to the position we would regularly – and unacceptably – find ourselves in, were we to follow the dictates of epistemic rationality (specifically, the Principle of Epistemic Impartiality) in an ‘unattenuated way’. Yet, as far as I can see, nothing in Evans’ way of understanding transparency – namely, as a distinctive epistemic method we have for obtaining knowledge about our own states of mind – tells us anything about why it should be less than perfectly possible for me to follow the dictates of PEI without compromise. For nothing in this way of understanding transparency tells us that I should treat my own beliefs – *even those I learn about through the use of the transparency method* – as in any way privileged as compared with the beliefs of others. Indeed, nothing in this understanding connects transparency especially to *belief*, for as we saw, the method of transparency is applicable wherever we are asked to pronounce on a first-order state that has intentional content – to say whether we want *x*, fear *y*, prefer *z*. (In fact, the method is applicable in the case of self-ascriptions of states that are not strictly speaking mental. When asked whether



we're *seeing* a tree, or *hearing* a loud noise, or even whether we remember whether such&such happened, we typically direct our attention 'transparently' to the world, thereby engaging the same cognitive mechanisms that are causally responsible for our first-order states.)

### Transparency and the 'First-Person Perspective' – Moran

In his discussion of transparency, Evnine draws on another way of understanding transparency, due to Richard Moran. Moran maintains that, in general, ascriptions of intentional states can be made from two different *perspectives*, or *points of view*, or *stances*: either *third-person* or *first-person*.<sup>10</sup> Ascriptions made from the third-person perspective are typically made on the basis of inference or hypothesis about, observation or perception of the subject's mental condition. Ordinary self-ascriptions of intentional states, however, are typically made from the first-person perspective. But like Evans, Moran does not think that what is distinctive about the first-person perspective is adequately captured by appeal to a form of highly reliable (and direct) inner-perceptual access we have to our own mental states.

To bring out the distinctive character of the first-person perspective, Moran calls upon Moore's Paradox. On Moran's proposal, we could understand *the self-ascriptive* conjunct of a Moore-sentence ('I believe that *p*') from two different perspectives. Understanding it from the third-person perspective, we *could* regard it as an empirical claim about someone's beliefs—someone who happens to be oneself. However, it is natural to understand self-ascriptions of beliefs as offered from the first-person perspective. In its ordinary 'transparent' use, "I believe" serves "to declare one's view of how things are, out there in the world beyond oneself." If so, then we may understand the absurdity of an utterance of a Moore sentence as due to the fact that we naturally read the self-ascriptive part of the Moore sentence as declaring one's belief that *p*, yet take its other part directly to express a rejection of *p*'s truth.<sup>11</sup>

Appreciating what makes Moore sentences appear absurd can thus serve to confirm Evans' observation that ordinary self-ascriptions of belief exhibit transparency-to-the-world. Following Evans, Moran takes it to be characteristic of beliefs ascribed from the first-person perspective that they meet the Transparency Condition, which says that "I can report on my *belief* about X by considering (nothing but) X

itself' (2001: 84), and that "in ordinary circumstances a claim concerning one's attitudes counts as a claim about their *objects*, about the world one's attitudes are directed upon" (2001: 92).

Unlike Evans, however, Moran does not think of conformity to the Transparency Condition in terms of employing an epistemically superior method for gaining knowledge about what intentional state one is in. Rather, "the claim of transparency is grounded in the deferral of theoretical reflection on one's state to deliberative reflection about it"; it represents a "normative demand" that reflects "the primacy of the deliberative stance within the first-person" both philosophically, and 'in the life of the person'. (See 2001: xvi.) Transparency, as Moran sees it, is essentially tied to the fact that an ordinary first-person belief self-ascription – what he calls a belief *avowal* "carries with it a *commitment to its truth*" (2001:91), so that when I ordinarily say: "I believe that *p*", I *commit* myself to *p's truth*. By contrast, when I ascribe a belief to another (even if I base the ascription on a transparent consideration of the world, which Moran thinks I *can*), my ascription only offers what I take the *other's* commitment to be.

On Moran's view, this commissive aspect is present in the case of avowals of other intentional states as well. When I avow a desire for *x*, or feeling mad at *y*, I commit myself to *x's* being desirable, and to *y's* being maddening, respectively. But when ascribing desires, etc. to someone else, I can fully distance myself from such evaluative judgments. Unlike ascriptions that simply cite the objective facts about someone's present mental states, avowals of intentional attitudes are declarations or expressions of one's commitments.<sup>12</sup>

The crucial contrast Moran sees between avowals and other ascriptions is not to be captured *simply* through the contrast between ascriptions that are transparent-to-the-world and those that are not. Rather, the crucial contrast for Moran is between self-ascriptions of mental states (beliefs included) that represent the employment of the theoretical (or empirical) perspective on one's beliefs (and other mental states), on the one hand, and those that are expressive of one's *practical* or *deliberative* commitment and of rational authority and agency. (See 2001: p. 92 & p.124f.) We can see Moran as claiming that it as an essential part of being a person that one should be able to occupy a first-person 'unalienated' perspective on one's own mental states, a perspective from which one's self-attributions imply genuine endorsement. (See 2001: xxx.) Indeed,

Evnine explicitly invokes Moran's characterization of transparency's link to practical reason when introducing the claim that "[t]hose of one's beliefs that are transparent form a kind of epistemic perspective occupied by the person whose beliefs they are... the totality of one's transparent beliefs are that through which the world is accessible to a person... Thus, ... transparent beliefs locate one in the world ... they determine one's identity, who one is" (p. 150). But I think we are still some distance from any conflict with the demands of rationality as encapsulated by PEI. To remind, PEI says

The considerations determining which belief it would be epistemically rational for an agent to adopt do not give special status to any of the agent's present opinions on the basis of their belonging to the agent. (p. 141)

The principle requires a certain 'epistemic indifference' as between one's own beliefs and those of others: in determining whether it would be epistemically rational for me to believe that  $p$ , I cannot use as a relevant consideration the fact that I do *in fact* believe that  $p$  (that  $p$  specifies the content of one of *my* present beliefs). Compare: in determining whether it would be practically rational for me to  $\varphi$ , I cannot use as a relevant consideration the fact that I am in fact now  $\varphi$ 'ing, or am presently inclined to  $\varphi$ . Or, when deciding whether I ought to prefer outcome A over outcome B, I cannot use as a relevant consideration the fact that I now in fact prefer A over B. As far as I can see, nothing in this requirement conflicts with the idea that, as a person, I must be able to know *what* my beliefs are in a 'first-person' way, where by that we mean (with Moran) that I must be able to ascribe beliefs to myself in a 'deliberative mode': to avow "I believe that  $p$ " as a way of committing myself to, or endorsing  $p$ . After all, PEI is a normative principle constraining the sort of considerations one is allowed rationally to bring to bear on the formation, or adoption, of beliefs.

Transparency in Moran's hands, on the other hand, represents a feature of a distinctive way we have, as persons, of attributing beliefs (and other states) to ourselves in the course of exercising our rational and deliberative agency. So it is not easy to see how the latter could place necessary limits on the former.

Come to think of it, a natural way of recasting PEI itself involves *using* transparency! To wit: in deciding what would be rational to believe, we ought to direct attention solely and directly to the objective

*contents* of the relevant beliefs, disregarding any considerations pertaining to their being contents of this or that subject's belief states. Far from generating a conflict, transparency can be wielded into a normative principle that is very much in keeping with PEI.

Evnine himself notes that "acceptance of PEI does not threaten any kind of conflict with the transparency of belief via Moore's paradox" (p. 152). This is because, as he points out, "[a]lthough one can recognize that a Moore-paradoxical proposition can be true of one, one could never be in a position rationally to accept or assert it if the belief that *p* were transparent to one." (p. 160). And no principle of epistemic rationality, PEI included, will prescribe acceptance of a belief that no one can be in a position rationally to accept. Evnine also does not want to argue that we should reject the conception of epistemic rationality embodied in principles such as PEI. Nevertheless, he thinks that this conception "presents to us an ideal that is at odds with what it is to be a believer, and [thus] with what it is to be a person" (p. 160f.) Later on, he summarizes the conflict in question as follows:

Thoroughgoing rational self-assessment would have us step outside ourselves in a way that a person necessarily cannot do. Being a person is essentially something perspectival: it involves having an epistemic perspective, a set of beliefs that are transparent to their bearer and in an important sense determine the identity of that person. (p.160).

But wherein the conflict? Suppose that the principles of epistemic rationality, and specifically PEI, indeed require us to 'step outside ourselves' in the relevant sense. That is, suppose that (contrary to my earlier suggestion) when we engage in a rational assessment of our own beliefs, we inevitably must attribute to ourselves beliefs 'non-transparently', or in a 'third-person' way. How would that conflict with the idea that, as persons, we must *also be able to* have a different, distinctively 'first-person' way of attributing beliefs to ourselves? There may well be forms of rational self-assessment that do not deploy our capacity for transparent self-ascription. But even when engaged in such forms of self-assessment, we may remain *capable of* transparent self-ascription. To put it differently, my being a person may require that I have the capacity trans-

parently to ascribe belief to myself, but unless it *disallows* non-transparent self-ascription, it should not conflict with thoroughgoing adherence to the demands of epistemic rationality.

Indeed, Evidine himself is a bit ambivalent about there being an out-and-out conflict here. Soon after spelling out the alleged conflict, Evidine notes that one of the necessary conditions on being a person – the requirement Second-Ordinality – “demands that persons be able to ... assess their beliefs in abstraction from the fact that they are their beliefs” (p. 160). So, as it turns out, it’s not that being a person requires that one be *unable* to follow the dictates of epistemic rationality, as claimed by thesis (3). Rather, being a person requires being subject to a certain conflict: “epistemic perspective, transparency, and identity on the one hand, and visibility, Second-Ordinality, and rational self-assessment, on the other, are always in potential conflict. Since both terms of this conflict are necessary to being a person, persons are necessarily subject to this conflict. The conflict itself is an inescapable part of personal life.” (p.160) Hence, what Evidine dubs ‘aspectual dualism’.

To sum up, PEI tells us what considerations we *may not* bring to bear in determining what is epistemically rational for us to believe, whereas transparency (even if taken to be a relational feature of beliefs themselves) at most tells us that our beliefs are *knowable* in a 'first person way'. So *prima facie* there seems to be no way the two can conflict. As we consider what would be rational for us to believe, we are constrained by PEI *not* to take into account whether or not the beliefs happen to be ones we in fact have.<sup>13</sup> Our consideration of whether to hold any particular belief may (and naturally would) involve directly considering the world, as per transparency. If it does, then whatever beliefs we end up embracing will constitute our (revamped) epistemic perspective. But there seems to be no tension here – and no direct support for Evidine’s thesis that to be a person, one must be *unable* fully to satisfy the demands of epistemic rationality.

### **Moore’s Paradox Revisited**

Before returning to Evidine’s misgivings about PEI, I want to take a brief detour via another possible take on how we should understand transparency and the absurdity of Moore sentences.<sup>14</sup> Let’s note first something that is rarely noted in discussions of Moore’s paradox – that there are natural analogues of Moore

sentences that do not involve beliefs. Consider utterances of the following conjunctions: “I hope it doesn’t rain, but please let it rain!” or “I like this painting a lot, but it’s horrible,” or “I am finding this meeting really exciting, but it’s very boring”. I submit that, in the absence of special context, such utterances will seem as anomalous as an utterance of the original Moore sentence. In fact, Moore-style absurdities can be generated even without producing any explicit self-ascriptions, as in “Brrr! It feels hot in here” or “[Yawn]; How interesting!” “[Laughter]; Not funny at all!”)

Here’s a suggestion about what is going on here. The so-called transparency of belief is a special case of a broader phenomenon: the *expressive* character of certain self-ascriptions of mental states – of so-called avowals. Avowals, so the story goes, are acts in which speakers (and thinkers) use self-ascriptive vehicles *in lieu* of simply giving direct expression to a mental state. Thus I can express my desire for ice cream by reaching for the treat, by saying: “Ice cream please!” *or* by self-ascribing the desire, saying “I’d like some ice cream”. Similarly, I can express my present hope that it stops raining by uttering “it’s going to stop raining” hopefully, by saying “Let it stop raining” or by saying “I hope it stops raining”. And the same for other propositional attitudes, belief included. On the present proposal, avowals such as “I believe that *p*” or “I hope that *q*” or “I like *x*” are acts in which speakers directly express their self-ascribed states *rather than* – or in addition to – their expressing *second-order belief* that they are in the relevant state.

When asked whether I believe that *p*, we ordinarily take this as an invitation to produce an avowal, rather than to render a verdict based on self-study of our mental states. This is why we directly consider whether *p*. We put ourselves in a position to give direct voice to our (first-order) belief by simply considering whether things are as *p* says. If we then say “I believe that *p*” our self-ascriptive utterance will be understood as an avowal: as though we were simply to affirm *p*, thereby expressing the first-order belief. And our audience, in turn, will typically respond by confirming or taking issue with the first-order belief, rather than by addressing the self-judgment.

Now suppose we understand utterances of Moore-sentences as combining an avowal of a belief with a non- self-ascriptive expression of a contrary belief. We can then see them as involving, like all Moore-style

anomalies, an *expressive conflict*. For example, a speaker who says “I hope it doesn’t rain but let it rain” can be seen as self-ascriptively expressing (i.e. avowing) her hope that it doesn’t rain in the first half of the utterance at the same time as she non- self-ascriptively expresses a contrary hope in the second half; just as one who rubs her arms but adds “It feels hot in here” can be seen as non-verbally expressing her feeling cold while verbally expressing her feeling hot.

### **Downloading Beliefs?**

If one accepts the above expressivist reading of transparency and Moore’s paradox, there is no reason to think that there are necessary limits on the ability of persons to adhere to the dictates of epistemic rationality of the sort that Evidentiary suggests. However, following his discussion of transparency, Evidentiary considers the Belief Downloader case imagined by Christensen, in which we have someone whose beliefs on some subject matter are of the same level of expertise as my own. According to PEI, “it should be a matter of epistemic indifference as to whether I activate the Belief Downloader that will sear our respective minds and, where we differ... will replace my beliefs with the beliefs of this other person... since failure to recognize the indifference would indicate a partiality to my beliefs merely because they were mine” (p.153). Evidentiary, however, thinks that “such indifference seems almost pathological in the degree of alienation from one’s own beliefs it recommends” (p.154). In the remainder of this commentary I’d like to consider whether the expressivist take on avowals canvassed above can provide some independent support for Evidentiary’s intuition about the limits of PEI as revealed by considering the Belief Downloader case.

What is crucial about the imagined Downloader case is that it introduces the possibility of a wholesale replacement of one’s beliefs that is in a sense ‘blind’: I am supposed to embrace the other’s beliefs, wherever they are at least as likely to be true as mine, but *not* via rational and transparent consideration of each belief. Such replacement, as Evidentiary observes, will have as a consequence that “I will simply have wrenched myself from one epistemic point of view to another, without having traced a path through epistemic space to get there” (p. 155). Generalizing the case to allow for the downloading of other mental

states – hopes, desires, intentions, preferences, etc. – there may be an expressivist way of articulating Evnine’s worry that downloading will result in my ‘becoming a different person’.

Suppose we think of an *authentic* subject of beliefs and other mental states as someone who is regularly in a position to give direct expression to her mental states, at times by self-ascribing them to herself via avowals. And suppose we think of downloading as a matter of entering contents  $p, q, r, \dots$  into one’s ‘belief-box’ or ‘desire-box’ or ‘hope-box’, etc. (and crowding out existing contents wherever conflict arises). Insofar as PEI and other principles of epistemic rationality indeed make room for the uncontrolled downloading of mental states, authenticity would seem to be at risk. If so, we may come closer (albeit from a different direction) to Evnine’s desired conclusion that there’s “a tension between epistemically rational indifference and the ways in which our beliefs place us in the world in a way that is deeply determinative of who we are.” (p. 155) For, consider: in an authentic subject, there is a regular match between what we may call ‘first-order’ expressions (both verbal and non-verbal) and ‘self-ascriptive’ expressions (or avowals) of her mental states. Such a subject is in a position to, e.g., utter (or think) “ $p$ ” hopefully *or* avow “I hope that  $p$ ”; she can say (or think) “Ice cream!” eagerly, *or* “I’d really like some ice cream”. But in a subject who has undergone uncontrolled downloading, there may be massive mismatches between the subject’s avowals and her first-order expressions. After downloading, our subject may suffer, e.g., an ‘alien hope syndrome’: while she would readily volunteer “ $p$ ” hopefully, she may be unable to avow “I am hoping that  $p$ ” (or she may even avow “I’m not hoping that  $p$ ”). If she undergoes a thoroughgoing downloading of mental states, our subject may end up being regularly surprised by, fail to identify with, and disown her own beliefs, desires, hopes, etc.; and this may well be regarded as undermining her identity as a person.<sup>15</sup>



## Endnotes

---

<sup>1</sup> An earlier version of this commentary was presented at an Author Meets Critics session at the Pacific APA, April 2009. The present version benefited from Simon Evnine's response to my comments (as well as subsequent correspondence), from audience discussion at the session, and from comments by Ram Neta and Gurpreet Rattan.

<sup>2</sup> See pp. 114-116, 120, 124, 126, where these norms are defended.

<sup>3</sup> See his "Diachronic Coherence versus Epistemic Impartiality", *Philosophical Review*, 1009: 349-71.

<sup>4</sup> Note that this is a stronger claim than the claim of 'tension' and the thesis of 'aspectual dualism' which Evnine defends in the same chapter. (See footnote 13 below.)

<sup>5</sup> see his *Varieties of Reference*, Oxford University Press, 1982

<sup>6</sup> For discussion of what I am calling 'transparency-to-the-world' and for relevant references, see Dorit Bar-On *Speaking My Mind: Expression and Self-Knowledge*, Oxford: Clarendon Press, 2004, especially Ch. IV.

<sup>7</sup> See pages 138, 146-152.

<sup>8</sup> See p. 149.

<sup>9</sup> Evans notes, however, that "mastery of this procedure cannot constitute a full understanding of the content of the judgment 'I believe that *p*'." (1982: 226) A subject who ascribes the belief that *p* to himself is different from the subject who merely forms the belief that *p*; the former must possess "the psychological concept expressed by 'ξ believes that *p*,' which the subject must conceive as capable of being instantiated otherwise than by himself." (*op. cit.*)

<sup>10</sup> For some relevant passages, see *Authority and Estrangement*, Princeton University Press 2001, p.xxxi. See also "Self-Knowledge: Discovery, Resolution, and Undoing." *European Journal of Philosophy* 5, no.2, pp. 143, 154 and "Interpretation Theory and the First-Person." *The Philosophical Quarterly* 44, No. 175, 156f.

<sup>11</sup> Since the analogous Moore-sentences in thought also seem absurd, to apply Moran's line to such cases, we'll presumably need to suppose that one can 'declare one's view' in thought, and not only in speech.

<sup>12</sup> Moran emphasizes that not all self-ascriptions of intentional states are commissive. We do sometimes ascribe intentional states to ourselves by way of pure explanation of our own behavior. However, Moran thinks that such self-ascriptions are in an important sense 'alienated,' and the kind of self-knowledge they represent is 'merely attributional'.

<sup>13</sup> Or, if we do take this into account, it can only be one more piece of factual information to consider in figuring out what to believe (e.g., it might be evidence for the truth or falsity of some hypothesis – especially if it is a hypothesis about me – that I happen to believe that hypothesis).

<sup>14</sup> I develop this take in Bar-On 2004, esp. Chapter VI and VII.

<sup>15</sup> Since we're here working with the idea that the relevant self-beliefs are not formed on the basis of, say, the subject's own epistemic access to the contents of his various 'mental boxes', the only way I can see to rule out such massive mismatches is to stipulate that, whenever a new content is downloaded to the relevant mental 'box' of a subject S (when, say, 'p' goes into the 'hope-box'), a matching *self*-ascription is downloaded to S's 'belief-box'. Note that this means stipulating that downloading must be sensitive to the identity of the subject of the mental states. It renders downloading subject to a strong requirement of 'automatic' *self*-knowledge. One may wonder what could rationally justify such a requirement. In any event, this means that the rationality of the Belief Downloader will depend on there being a rational justification of a this strong requirement.