



Distinctive substantial self-knowledge and the possibility of self-improvement

Josep E. Corbi¹

Received: 30 September 2022 / Accepted: 18 May 2023 / Published online: 5 June 2023
© The Author(s) 2023

Abstract

Quassim Cassam distinguishes between trivial and substantial cases of self-knowledge. At first sight, trivial cases are epistemically distinctive insofar as the agent needn't provide any sort of evidence to ground her claim to knowledge. Substantial cases of self-knowledge such as 'I know I want to have a second child' do not seem to bear this distinctive relation to evidence. I will argue, however, that substantial cases of self-knowledge are often epistemically distinctive and, to this end, I will challenge a crucial assumption in the current debate about self-knowledge, namely: if a piece of self-knowledge is based on evidence, it must have been delivered by a detached, theoretical attitude toward oneself (*The Detachment Assumption*).

My case against the Detachment Assumption combines a negative and a positive programme. Regarding the negative aspect, I will first present Cassam's case for Inferentialism; second, I will argue that this view about self-knowledge is at odds with the sort of self-improvement that he vindicates in his analysis of epistemic vices and, third, I will conclude that only by allowing for an engaged relation to evidence, can we make sense of that sort of self-improvement. Regarding the positive programme, I will first examine the sensitivity to the music that is specific of a graceful dancer and, on this basis, outline an attitude toward oneself that, despite involving evidence, is not detached or theoretical but engaged, so that it gives rise to a kind of substantial self-knowledge that is both transformative and epistemically distinctive.

Keywords Self-Knowledge · Distinctiveness · Inferentialism · Rationalism · Self-Transformation · Self-Transformation · Passivity

✉ Josep E. Corbi
josep.corbi@uv.es

¹ Department of Philosophy, University of Valencia, Valencia, Spain

Unlike what is customary in the current debate about self-knowledge, in *Self-knowledge for Humans* (Cassam, 2014), Quassim Cassam distinguishes between trivial and substantial cases of self-knowledge and explores why the current debate is almost exclusively concerned with trivial cases such as ‘I believe there is yoghurt in the fridge’ or ‘I intend to go for a run at 9 am’. At first sight, such cases are epistemically distinctive insofar as the agent needn’t provide any sort of evidence in order to ground her claim to knowledge; in fact, her self-ascriptions are all the more authoritative precisely because no evidence seems to be required. This certainly goes against standard expectations concerning claims to knowledge, where one’s authority is proportional to the available evidence. Substantial cases of self-knowledge such as ‘I know I want to have a second child’ or ‘I know I want to migrate to Europe’ do not seem to bear this distinctive relation to evidence. There is plenty of room in such cases for self-ignorance and self-deception and, consequently, the agent’s authority is surely proportional to the available evidence. My fundamental purpose in this paper is to argue that, contrary to what is customarily assumed, some central cases of substantial self-knowledge are epistemically distinctive and, therefore, that they involve an epistemic asymmetry between the first-person and the third-person perspectives. To this end, I grant that such cases of self-knowledge are based on evidence, but I will challenge what I regard as a crucial assumption in the current debate about self-knowledge, namely: if a piece of self-knowledge is based on evidence, it must have been delivered by a detached, theoretical attitude towards oneself (*The Detachment Assumption*). By this means, I will make room for a sort of attitude towards oneself that is based on evidence and still distinctively first-personal insofar as it is also engaged.

My case against the Detachment Assumption combines a negative and a positive programme, but, more specifically, the structure of the paper is as follows. In Sect. 1, I will present two common approaches to self-knowledge that, as we will see, are ultimately committed to the Detachment Assumption, namely, Rationalism and Inferentialism. Rationalist views associate the distinctiveness of self-knowledge with our capacity to deliberate about the world.¹ They stress, for instance, that one’s reasons to claim that one believes that *p* are transparent to one’s reasons to claim that *p*: that is, whatever reason one may have to claim that *p* will also count as a reason to claim that one believes that *p*. To use Cassam’s phrase, we can refer to this means to self-knowledge as the Transparency Method (TM). Rationalists expect TM to go beyond the case of knowing one’s own beliefs and include several other mental states, such as desires and intentions. Even though TM may not apply to all mental states, Rationalists will stress that TM delivers a kind of self-knowledge that is normal, and even indispensable. Cassam objects to TM and defends instead an Inferentialist view, that is, a view according to which normal, basic cases of self-knowledge are based on — or justified by — inferences concerning evidence about oneself. Inferentialism does not exclude the existence of cases of self-knowledge that are non-inferential, but regards such cases as exceptional.

¹ See Cassam (2014, ch. 6–8) and Gertler (2011, ch. 6; 2021) for an overview of Rationalist approaches to self-knowledge.

Cassam's case for Inferentialism relies not only on the implausibility of the available alternatives, especially Rationalism, but on his defence of Inferentialism from some crucial objections. In this paper, I will examine two. I will take advantage of Cassam's response to the first of these objections, which doubts the distinctiveness of one's inferential access to one's own mental states, to highlight his commitment to the Detachment Assumption, namely, to the idea that an inferential access to oneself is as detached as that of a third party. In connection with the second objection, concerning alienation from one's own mental state as result of having an inferential access to them, in Sect. 2 I will present Cassam's defence of the possibility of self-improvement regarding one's epistemic vices. He is convinced that, in some recalcitrant cases, one's ability to detect one's own epistemic vices will not suffice to eliminate them and that, in such cases, only exposure to certain situations or engagement with some specific practices will help. From an Inferentialist perspective, it is rather mysterious how this benefit could be obtained, that is, what kind of awareness or sensitivity may come with such exposures and practices that could eventually have a transformative effect. More specifically, I will argue that the Detachment Assumption stands in the way of our ability to make sense of the kind of self-improvement that Cassam commends or, in other words, that we must allow for a sort of relation to evidence that is not detached but engaged if we are to make sense of the fact that transformative experiences are associated, as Cassam himself suggests, with the idea of insight or revelation. Here is where the negative programme ends.

Regarding the positive programme, I will examine the sort of sensitivity to the music and to one's own body that is specific of a graceful dancer as opposed to that of an unimaginative one. Based on this contrast, I will outline an attitude towards oneself that, despite involving evidence, is not detached or theoretical but engaged; moreover, I will suggest that this attitude, which I call 'receptive passivity', serves to identify the kind of self-awareness that renders epistemic self-improvement intelligible when mere inferential awareness of one's vice does not suffice. I will finally argue that receptive passivity provides a kind of substantial self-knowledge that is both distinctively first-personal and transformative.

1 Cassam's case for inferentialism

Cassam is perplexed that the current debate about self-knowledge is almost exclusively concerned with trivial cases, leaving aside substantial cases that are so important to our lives. The reason seems to be that trivial cases are epistemically distinctive while substantial ones are not. Cassam will argue, though, that trivial cases of self-knowledge are not so distinctive while I will proceed the other way around, namely: I will defend the view that some central cases of substantial self-knowledge are distinctively first-personal, although, for this purpose, we would have to renounce the Detachment Assumption. But, prior to this, let us briefly consider how Cassam identifies substantial cases of self-knowledge.

1.1 Trivial and substantial cases of self-knowledge

Cassam lists a few features of substantiality that are almost exclusively epistemic. They have to do with fallibility, evidence, cognitive effort and so on. Only one feature on his list refers directly to the idea of practical significance, namely, the Value Condition: “substantial self-knowledge *matters* in a practical or even in a moral sense.” (Cassam, 2014, p. 31). A further feature he mentions is connected with practical or existential concerns through the idea of identity or self-conception: “... this kind of [substantial] self-knowledge often ‘tangles with’ a person’s self-conception.” (Cassam, 2014, p. 30). We can then distinguish two aspects of substantiality regarding self-knowledge: an *epistemic* and an *existential* aspect. The epistemic aspect has to do with the idea that such pieces of self-knowledge are epistemically demanding in various respects, especially concerning the gathering of evidence and the ability to let one’s deliberations be guided by it; the existential aspect refers, by contrast, to situations that are recognised - in a way still to be elaborated - as important to us, that is, as connected to things that matter. What I intend to defend is the view that existentially substantial cases of self-knowledge are not only epistemically substantial but, in some crucial cases, distinctively first-personal too. In what follows I will qualify my use of ‘substantial’ only exceptionally, since the emphasis on the epistemic or existential aspect of substantiality will usually be inferred from the context.

According to Cassam, Inferentialism derives a significant portion of its strength from the implausibility of the available alternatives (Cassam, 2014, pp. 141ff). Regarding these alternatives, he focuses mainly on Rationalist approaches to self-knowledge. Let me briefly motivate these approaches and then present Cassam’s case against them.

1.2 Rationalist approaches to self-knowledge

There are two ways, according to Richard Moran, in which the declaration ‘I don’t know *how* I feel about that’ can be approached. It can be construed either as raising a theoretical question regarding one’s actual feelings (‘I don’t know what it is that I do feel’) or as expressing a practical concern (‘I don’t know what *to* feel about that’) (Moran, 2001, p. 58). The theoretical question presents one’s feelings as mere inner happenings that one must discover, whereas the practical concern involves a preoccupation for the appropriateness of one’s feelings and the need to transform them if required. This practical concern involves, as we see, a deliberative attitude that does not reduce to the formation of a normative judgement but includes the corresponding commitment to transform one’s feelings or, in general, one’s psychological condition (Moran, 2001, p. 59). Two attitudes towards oneself are thus discerned, namely, one theoretical and the other deliberative:

In characterizing two sorts of questions, one may direct toward one’s state of mind, the term ‘deliberative’ is best seen at this point in contrast to ‘theoretical,’ the primary point being to mark the difference between that inquiry which terminates in a true description of my state, and one which terminates in the formation or endorsement of an attitude. (Moran, 2001, p. 63)

The theoretical attitude involves, according to Moran, a detached view about oneself. One regards one's feelings and actions as a passing show and, like a third-party, one must observe one's inner events and external behaviour to know them. The deliberative attitude comes instead with the idea of a commitment as a result of one's sensitivity to reasons.

Moran articulates the contrast between these two attitudes in terms of the Transparency Condition. Even though this condition can be applied to various sorts of psychological states and attitudes, belief is the simplest case (Moran, 2001, p. 65). He thus begins with the question:

(B) 'Do I believe that P?'

and examines how it relates to a question about the world itself:

(T) 'Is P true?'

Despite the disparity of these two questions, it would sound quite weird that an agent might answer them differently, namely, that a yes to one of them would come with a no to the other. But such harmony in the answers to these questions seems to derive from the fact that one's answer to (B) is the product of a deliberation about the appropriate answer to (T), rather than the outcome of a detailed observation of one's inner states and events. In other words, to answer (B) one must inspect those aspects of the world that might be relevant to determine whether P is true. It follows that (B), thus construed, is *transparent* to (T) insofar as the agent must answer B ". . . by reference to (or consideration of) the same reasons that would justify an answer to the corresponding question about the world" (Moran, 2001, p. 62; see Dunn, 2006, pp. 38–43; Edgley, 1969, pp. 89ff; Evans, 1982, pp. 225; Gallois, 2008, ch. 3; and Hampshire, 1957, ch. 3–4). Rationalist approaches tend to generalise over the belief case and present the Transparency Method (TM), to use Cassam's phrase, as the fundamental source of strictly first-personal self-knowledge. The most plausible formulation of TM goes, according to Cassam, as follows:

... the question of whether I believe that P is, for me, transparent to the question of *what I ought rationally to believe*. (Finkelstein, 2012, p. 103)²

Cassam (2014) raises three main objections against this view, namely: the Generality Problem, the Substitution Problem, and the Matching Problem.³ Let us explore them in some detail given that they are not only central to Cassam's case for Inferentialism but serve to qualify Rationalism as well as to highlight how both approaches rely on the Detachment Assumption.

² There are, however, reasons to doubt that that this formulation of TM is the most favourable to Rationalist's views (Boyle 2015, pp. 340–2). We can safely leave aside this issue insofar as my argument against the Detached Assumption is not genuinely affected by it.

³ Cassam (2017) elaborates a more fundamental kind of objection, namely: that TM is only plausible if it involves some inferences (Gertler, 2021). From this perspective, TM would not be an alternative to Inferentialism, but just a particular version of it.

1.3 Cassam's case against rationalism

The Generality Problem has to do with the fact that TM may apply to beliefs but hardly to other mental states such as desires and hopes (Cassam, 2014, p. 103). Moran could certainly reply that desires or feelings might also be known by means of TM insofar as we might accept, like he himself, that there is a matter of fact as to what is worth desiring or what is appropriate to feel on a certain occasion. Regarding hopes, it is harder to defend TM, though, for it is unclear why there should always be an appropriate answer to the question as to whether one ought rationally to hope that P. But, as Cassam highlights, this concern could also apply to beliefs, desires and feelings, since it is quite often undetermined whether one ought rationally to have a certain belief, desire or feeling: "It's not just that it can be very hard to *know* which attitude one ought rationally to have but that in many cases there is no such thing as *the* attitude 'the reasons' require one to have." (Cassam, 2014, p. 104) In other words, we may say that, even though Rationalism might account for one's knowledge of those mental states that one ought rationally to have, there are many other cases of self-knowledge, both trivial and substantial, that hardly meet this constraint and, as a result, the Generality Problem seems to be confirmed.

The Substitution Problem has to do with the fact that, on many occasions, it is easier for an agent to know, say, what she actually desires than whether she ought rationally to have a particular desire: "I have no idea what I ought rationally to want to drink, and if that is the case then why would I think that figuring out whether I ought to want to have a vodka martini is a good way of figuring out whether a vodka martini is what I want?" (Cassam, 2014, p. 104) Hence, TM—and, in general, any transparency condition that may involve some sort of deliberation (Boyle, 2015)—deprives us of the immediacy of trivial cases of self-knowledge that Rationalism is supposed to account for.

Finally, we come to *the Matching Problem*, which hinges on the fact that quite often what an agent is convinced she ought rationally to believe fails to match her actual beliefs, since beliefs do have a dispositional component that may resist the agent's judgment to the contrary (Cassam, 2014, pp. 112–8). A certain agent may be afraid of flying and, nevertheless, acknowledge that it is the safest means to travel, or be frightened at the sight of a certain insect she judges entirely harmless. As we see, the transition from judging to believing is far from trivial, since the latter includes a dispositional pattern that goes beyond the fact that an agent has made a certain judgment. As Richard Moran puts it:

... It is hard to see how anything remotely like our concept of belief could fail to play a sort of dual role: as explanatory of behavior and as bearer of truth values. (Moran, 2001, 129)

This dual role makes room for situations where there is a mismatch between what the agent sincerely claims to believe because of her deliberation about what she has most reason to believe and what her psychological dispositions (both linguistic and otherwise) may reveal she is believing. In such cases, (B) could be raised as a theoretical question:

(B*) Is ‘believing P’ among my psychological dispositions?

Transparency of (B*) to (T) would require that an agent’s psychological dispositions should be sensitive to the same reasons that would favour one or another answer to (T). But quite often such transparency is not a trivial matter for the agent but a remarkable achievement (Moran, 2001, p. 67).

1.4 Normal cases of self-knowledge

Rationalist approaches could certainly reply to all three objections by reducing the scope of their approach to some specific cases of self-knowledge. This strategy might preserve the relevance of TM if combined with the idea that such cases are in a relevant sense fundamental. They might thus argue that knowledge of one’s mental states by TM constitutes the basic or normal case while other cases should be regarded as deviant or exceptional —*the normality claim*; alternatively, one could say that this kind of self-knowledge, even though it is not basic or normal, is at least indispensable insofar as it is presupposed by all other sorts of self-knowledge —*the indispensability claim*. Thus, Boyle (2009) claims that, even though TM cannot plausibly account for all sorts of self-knowledge, this kind of self-knowledge is fundamental insofar as it grasps a capability that any other variety of self-knowledge presupposes:

If this is right, then we are in a position to say why the kind of self-knowledge that Moran characterizes is fundamental. It is fundamental because the ability to say what one believes in the way that Moran specifies is intimately connected with the kinds of representational abilities that must be possessed by a subject who can make comprehending assertions, and a subject who lacks these sorts of abilities cannot be a self-representer, in the sense we specified, at all. (Boyle, 2009, p. 151, see p. 156)

It is unclear to me whether Cassam rejects the indispensability claim, but his inferentialist view surely conflicts with the normality claim. In fact, the three problems he raises point in the same direction, namely: TM cannot be the basic or normal means to self-knowledge, so that all other sources of knowledge should be regarded as exceptional or deviant, since, even in trivial cases, evidence is often involved. Inferentialism does not exclude the existence of non-inferential sources of self-knowledge, although it stresses that self-knowledge based on evidence is not deviant or aberrant, but the most common case: “Inferentialism says that inference is a basic source of intentional self-knowledge for humans. This could be true even if some of our self-knowledge, including intentional self-knowledge, is non-inferential.” (Cassam, 2014, p. 145; see Cassam, 2017, pp. 734-5)⁴ And this applies both to substantial and trivial cases, as the Substitution Problem suggests. So Cassam can provisionally conclude that:

⁴ See Boyle (2011), Byrne (2011), Fernández (2013) and Gertler (2018, 2021) for further discussion of this point.

... knowledge of our own beliefs, desires, hopes, and other ‘intentional’ states is *first* and *foremost* a form of inferential self-knowledge’ (Cassam, 2014: 137; my italics).

It may be relevant to stress, as Cassam (2017) does, that Inferentialism is a view in epistemology and, therefore, that the claim for inferential knowledge is a normative one, namely, that normal cases of self-knowledge are inferential just because they *must* be based on, or justified by, some inferential process. This view must be distinguished from an empirical claim according to which normal cases of self-knowledge actually involve a psychological process of inference (Cassam, 2017, pp. 727-8; Cassam, 2014, pp. 145-6, see p. 124). Hence, Inferentialism may be true even though the subject has no experience of drawing an inference or no psychological mechanism can be individuated to this effect. We can then state the central inferentialist claim about self-knowledge like this: *normal cases of self-knowledge must be based on — or justified by— inferences concerning evidence about oneself.*

1.5 Two objections against inferentialism rebuked

To elaborate his case for Inferentialism, Cassam examines in some detail four objections that could be raised against it. I will confine myself, however, to the only two that are relevant to my line of argument, namely: *the distinctiveness objection*, that is, Inferentialism can hardly honour the distinctiveness of self-knowledge, namely, the essential epistemic asymmetry between the first- and the third-person perspectives; and *the alienation objection*, that is, Inferentialism involves an alienated view about one’s psychological condition that is incompatible with our agency.

Regarding *the distinctiveness objection*, Cassam develops a strategy that seeks to combine accommodation and denial (Cassam, 2014, p 149; see Cassam, 2017). On the accommodation side, he argues that, after all, a third-party does not have access to the same kinds of evidence about one’s own mental states as oneself and, in this respect, there is an asymmetry between the first- and the third-person perspectives (Cassam, 2014, p. 149). Rationalists might reply, though, that this is not the sort of asymmetry they are interested in. The fact that there may be a difference between the kinds of evidence that can respectively be accessed from each perspective, does not render self-knowledge epistemically distinctive.

In response to this, Cassam turns to denial and calls into question the distinctiveness of the first-person perspective as a datum. There is no distinctive access to one’s own mental states. The agent herself and a third-party bear the same relation R to the available evidence, only that the kinds of evidence available may differ in each case. Yet, this relation R is precisely that of the detached observer and, therefore, Cassam’s line of reply to the first objection seems to take the Detachment Assumption for granted. In Sect. 3, I will challenge this assumption and defend the distinctiveness of existentially substantial self-knowledge in terms of a relation R* to evidence, but let us move on now to the issue of alienation.

The Alienation Problem runs as follows: “... the final objection to Inferentialism says that inferential self-knowledge is alienated rather than ordinary self-knowledge, and that Inferentialism doesn’t account for ordinary, ‘unalienated’ self-knowledge.”

(Cassam, 2014, p. 157) Cassam replies that, for any given attitude that an agent may come to know inferentially, it does not follow that she is not identified with it:

Just because inferential or attributional self-knowledge *can* be of attitudes you don't identify with and can't endorse, it doesn't follow that this kind of self-knowledge has to be alienated in these ways. The mere fact that you self-ascribe an attitude on inferential grounds doesn't make the attitude alienated or imperious to reason. (Cassam, 2014, p. 157)

Moreover, Cassam insists that one's identification with a certain desire or attitude may survive one's deliberation to the contrary (Cassam, 2014, p. 157) and, consequently, that one may feel alienated from the conclusions of one's deliberation (Cassam, 2014, p. 158). This suggests that Inferentialism is at least on a par with Rationalism regarding alienation.

I doubt, however, that Inferentialism could properly account for the experience of identification. I agree with Cassam that identification takes more than just reaching a conclusion considering the kind of deliberative attitude that Rationalists contemplate, since an agent's dispositions may not be sufficiently permeable to her decisions or intentions. I am thus happy to grant that knowledge of what projects and attitudes one is identified with requires the contribution of evidence and, in this respect, I agree with Inferentialism. What I will dispute is that an agent's relation to those attitudes and projects she is identified with could be reduced to the deliverances of a detached, theoretical attitude (Boyle, 2015). Of course, a third-party might detachedly discover what attitudes or desires a certain agent may feel identified with and even the agent herself may begin to discern a given identification of hers by this inferential means. The question I want to raise is, however, whether this sort of discovery is all there is from the first-person perspective or, in other words, whether we should rather regard as weird or insane someone who were not able to access those attitudes of hers she identifies with from a different perspective.

In the next section, I will consider this question while examining the sort of self-improvement concerning one's epistemic vices that, in *Vices of the Mind* (Cassam, 2017), Cassam vindicates as commendable and, therefore, as possible. I will suggest that the Detachment Assumption prevents us from understanding the kind of sensitivity that is involved in this sort of self-improvement. I will thus conclude that Cassam must make room for a kind of attitude that is both engaged and sensitive to evidence if we are to make sense not only of the idea of identification but of the sort of self-transformation that he promotes.

2 Epistemic vices and self-improvement

2.1 Three kinds of control: voluntary, evaluative, and managerial

In *Vices of the Mind*, Cassam explores several epistemic vices concerning our attitudes and traits of character. In the final chapters of the book, he addresses the issue of one's moral responsibility for such vices and also that of the means for self-improvement.

ment. He associates moral responsibility with the capacity to control and, in this respect, he distinguishes three kinds of control: voluntary, evaluative and managerial. Obviously, I cannot change my beliefs at will, that is, in the way I may decide to raise my arm and, as a result, raise it. In general, we could say that our epistemic attitudes - and, therefore, our epistemic vices - are not under direct voluntary control (Cassam 201, p. 126). We usually possess, though, evaluative control over our beliefs: “[We] control our beliefs by evaluating (and reevaluating) what is true. Call this ‘evaluative’ control.” (Hieronymi, 2006: 53). This kind of control lies behind TM and seems to be connected with the fact that beliefs - and, in general, epistemic attitudes - must aim at tracking the world for them to count as beliefs at all.

Besides these two types of control, Cassam, following Pamela Hieronymi, introduces a third kind of control, namely, managerial control:

One might combat one’s arrogance by exposing oneself to superior intellects or one’s closed-mindedness by forcing oneself to think long and hard about the merits of opinions with which one disagrees. Such self-improvement strategies are not guaranteed to succeed but nor are they guaranteed to fail. If they succeed they are examples of a person exercising managerial control over their vices. They see their character as something that can be reshaped, at least to some extent. Unlike managerial control over ordinary objects, managerial control over our own character vices is indirect. The layout of the furniture in my office can be changed by moving bits of furniture. Changing one’s character or a particular character trait is a matter of doing other things that hopefully have the desired effect. (Cassam, 2019, p. 130; see p. 127)

According to Cassam, a first step in the attempt to get rid of one’s epistemic vices, or at least to attenuate them, is to become aware of their existence, which is not always an easy task, especially with regard to stealthy vices, such as arrogance or closed-mindedness insofar as they themselves diminish or undermine one’s capacity to become aware of them; after all, some degree of open-mindedness is required to acknowledge one’s closed-mindedness and, similarly, for one’s ability to spot one’s own arrogance.

In any event, Cassam assumes that detecting one’s epistemic vices is a *first step* towards overcoming them, but often just a first. For, even though one may become theoretically or inferentially aware of some epistemically vicious attitudes or character traits, they will typically keep on distorting the way one may assess the available evidence. Some further step is needed if one is to get rid - or at least to attenuate - the impact of one’s epistemic vices. What might this further step consist of?

2.2 Traumatic experiences and self-transformation

Cassam insists that one may try first to be more careful in the process of gathering evidence and make every effort to compensate what one may have identified as one’s vicious tendency to gullibility, to arrogance, and so on. This is, indeed, an effort of the will that could produce some immediate results on a particular occasion and eventually soften the corresponding epistemic vice. This softening would amount to

a transformation whose grounds are, nevertheless, still to be elucidated, since it is unclear how mere inference, interpreted on the basis of the Detachment Assumption, could make sense of this transformative effect. Secondly, one might try to improve one's epistemic attitude by exposing oneself to certain situations in the hope that by this means one's sensitivity will gradually be altered. Traumatic experiences are a particular case of transformation by exposure that Cassam examines in some detail.

By 'traumatic experience' Cassam understands "a sudden, unexpected potentially painful event' that 'ruptures part of our way of being or deeply held understanding of the world" (Rushmer & Davies, 2004, p. ii14). He agrees that "talk of *seeing* oneself in a new light is appropriate because what the traumatic experience produces is *a certain kind of insight*" (Cassam, 2019, p. 161, my emphasis). Still, there is no guarantee that traumatic experiences are genuinely transformative, for "impressions require interpretation, and interpreting one's impressions or traumatic experiences is the work of the intellect. So, it seems that there is no getting away from the potential impact of stealthy vices: they *can* get in the way of self-knowledge by traumatic experience, by causing us to misinterpret these experiences or misunderstand their significance." (Cassam, 2019, p. 164).⁵ We must reflect, however, on the conditions under which a particular traumatic experience may be individuated as transformative.

It seems, to begin with, that this traumatic experience must carry within it some epistemic element, since it is claimed to produce or favour a transformation in one's epistemic attitudes and character as a result of a certain kind of insight. But can we make sense of this transformation if our sensitivity to evidence is necessarily the product of a detached attitude and, therefore, motivationally inert? Didn't we conclude that the deliverances of such an attitude are at most a first step in the process of transformation and that, therefore, a further step is required?

All this suggests that the sort of relation R^* that an agent bears to evidence whenever a traumatic experience has a transformative effect must differ from the kind of relation R that she had in the first place, since the deliverances of R were assumed to be motivationally inert. To put it another way, if there is room for traumatic experiences to be transformative—even though quite often they are not because epistemic vices, like all vices, die hard—, then R^* must be conceived of both as *sensitive to evidence* (otherwise the fact that such experiences are insightful or illuminating would be either mysterious or accidental) and as *engaged* (as the motivational inertness of the first step must be overcome).

Acknowledging the indispensability of R^* to account for the possibility of self-transformation implies the rejection of the Detachment Assumption, since R^* provides a kind of access to evidence that is constitutively engaged. Moreover, the kind

⁵ See Paul (2014) for some epistemic paradoxes that decisions concerning transformative experience may generate. Such paradoxes derive from the fact that the preferences and values of the self that inspire a certain choice may significantly differ from those of the transformed self, that is, the self that emerges from undergoing a particular transformative experience. There is, first, the difficulty to ascertain what it would be like to have such an experience for the transformed self and, second, the problem of determining whose set of preferences and values should prevail, that is, those of the pre-choice self or those of the transformed self. These are most pressing questions because, as Paul highlights, they reveal the epistemic situation we confront when dealing with the most significant decisions in our life, such as the decision to have a second child.

of self-knowledge that R^* provides is not only manifestly substantial but distinctively first-personal too, since R^* is crucially distinct to the sort of relation R that a third-party is assumed to have. After all, R combines gathering of evidence with the sort of detachment philosophers tend to associate with a third-party perspective, while R^* points to an engaged sensitivity to the situation.⁶ In the next section, I will further motivate R^* by exploring some fundamental experiences where this kind of sensitivity seems to be required.

3 Receptive passivity, self-knowledge, and self-transformation

The epistemically vicious agent may sincerely accept, in light of evidence, that her deliberative capacities are hampered by some epistemic vices and still be unable to prevent such vices from distorting the way she perceives a situation or how she deliberates about it. Hence, the issue must be addressed as to what else she can do to improve her epistemic capabilities or, more specifically, what alternative sort of self-awareness may contribute to self-transformation. This is the issue I will approach in the present section.

More specifically, in Sect. 3.1 I will introduce the notion of *receptivity* or ‘being in tune with’ in light of the experience of the graceful dancer as opposed to that of

⁶ Some may doubt the intelligibility of R^* , partly because one should then admit that some psychological states or attitudes have a dual direction of fit, something which sounds absurd to many. This is a vexed issue that cannot be discussed here at length. See Dunn (2006), Frost (2014), Little (1997), and Zangwill (2008) for a challenge to the claim that the very idea of a mental state with dual direction of fit is incoherent. For further discussion, see Anscombe (1963), Gregory (2012), Humberstone (1992), Schueler (1995), and Smith (1994). Let me highlight, however, how the claim of unintelligibility may be partly motivated by some metaphysical assumptions closely connected to the Detachment Assumption, and therefore my case against the latter will also pose a problem to such assumptions. To begin with, I should mention Humean accounts of motivation according to which every action is to be explained by a proper combination of beliefs, which have a world-to-mind direction of fit, and desires, whose direction of fit is mind-to-world. No room is left then for a mental state or attitude that might contribute to explaining an action as the result of both providing a certain view of the world, as beliefs are meant to do, and having some motivational import, which desires constitutively possess. Humean accounts presuppose a divided conception of the self that Kantian approaches to morality come to confirm and, if I am right, the current debate about self-knowledge takes for granted as well. According to this conception, the self is divided into two parts: a rational, deliberative part that is concerned with how the world is and how one should respond to it; and a dispositional part that motivates one to act in a certain way. The rational part is thus assumed to have only a mind-to-world direction of fit and the dispositional part a world-to-mind direction of fit and, consequently, no room is left for mental states or attitudes with a dual direction of fit. To sum up, I could say that the reluctance to accept the intelligibility of mental states with a dual direction of fit may partly derive from a commitment to the divided conception of the self. My line of argument in this paper can be regarded, however, as a challenge to this conception of the self, since the Detachment Assumption is just the epistemic correlate of this metaphysics of the self; to put it another way, if the structure of the self is as the divided conception claims it to be, then the Detachment Assumption must be granted. Hence, a case against the Detachment Assumption is, by *modus tollens*, also a case against such conception of the self. More specifically, I have argued that the Detachment Assumption—and, therefore, the divided conception of the self—cannot account for a subject’s capacity for self-transformation. It follows that only by renouncing the divided conception of the self, and therefore accepting that some mental states and attitudes could have a dual direction of fit, can we make sense of this capacity for self-transformation. See Sect. 3.2 for further discussion.

an unimaginative one.⁷ This receptivity will, in turn, be elucidated in terms of a kind of *imposition* that has to do with the way the conclusion of a mathematical proof imposes itself upon the agent who understands it. Imposition, in turn, comes hand in hand with the idea of passivity. In Sect. 3.2, I will distinguish two sorts of passivity, namely: *base* and *receptive passivity*. The former relies on a *divided conception of the self* and regards an agent's passions as essentially base insofar as they are viewed as constitutively alienated from her true self, whereas the sort of passivity I will associate with imposition departs from this conception of the self and is anchored to a sense of one's agency that favours a certain kind of integration. Once the concept of receptive passivity has been thus elucidated, I will use it in Sect. 3.3 and 3.4 to shed some light on recalcitrant cases of epistemic vice and to account for the epistemic distinctiveness of some central cases of substantial self-knowledge.

3.1 Receptivity and imposition

The dancer perceives an order in a piece of music which she seeks to express in her dancing, in the way her body moves, but how does this transition from the music to the dancing body take place? We may first consider the case of an unimaginative dancer. There are more ways than one in which a dancer may be unimaginative, but I will confine myself to a way of being dull and unimaginative that is closely connected to the Rationalist view about self-knowledge, namely: the dancer whose bodily movements are guided by a set of principles or rules. We may say, by contrast, that the graceful dancer has a certain experience of her body as she pays attention to the music and, as a result, she moves her body in a particular way. Attention to the music, but also to the emotions and bodily experiences that she senses as deriving from the music, are essential to her gracefulness, to her ability to dance the music beautifully.⁸ Her experiences and movements will thus be *finely and creatively* in tune with the music. The unimaginative dancer, on her side, will also have some bodily experiences and some emotions will surely accompany her performance; the worry is rather that she will be connected to the music in a rather *stereotypical and rigid* manner.⁹

⁷ This contrast is inspired in Ruskin (2004, pp. 13–5)'s distinction between the unimaginative painter, who "never works without a principle" (Ruskin, 2004, p. 13), and the imaginative painter, who "is precisely the reverse of this. He owns no laws. He defies all restraint, and cuts down all hedges... He saw his tree, trunk, boughs, foliage and all, from the first moment." (Ruskin, 2004, p. 14). I must say that, even though I will defend the importance of a certain kind of seeing for the possibility of gracefulness, it should not be confused with the romantic idea of seeing at first sight that Ruskin vindicates.

⁸ Rules can surely play a role in graceful dancing, but a limited one. They should rather intervene as a scaffold to the sort of attention I am trying to elucidate (Wiggins, 1987).

⁹ Some might object that the opposite of 'unimaginative' is 'imaginative' rather than 'graceful' and, therefore, that the proper contrast should be between the imaginative and the unimaginative dancers. My motivation to avoid this contrast is that 'imaginative' can be easily interpreted in consonance with the Detachment Assumption insofar that one might imagine is typically approached as part of what one may detachedly observe about oneself. I am certainly happy to grant that the graceful dancer is imaginative but only insofar as her imaginativeness is construed as embodied, since it is this embodiment that the idea of gracefulness intends to stress. My choice of words sounds certainly strange from the viewpoint of some standard dichotomies, but this effect is almost welcome insofar as the terms I propose are meant to articulate a view that calls some such dichotomies into question. Of course, I could instead have

To be receptive to the music, the graceful dancer must also be receptive to her own emotional and bodily responses as being (or failing to be) finely in tune with it. This receptivity will ultimately show in her capacity to let her bodily movements be inspired by such experiences and the order thereby recognised in the music. In general, we can say that the notion of gracefulness involves, as we see, a certain sort of *receptivity*. To elaborate on this kind of receptivity, we may go back to the idea of evaluative control, as Cassam presents it, and the concept of imposition that comes with it.

We do not have voluntary control over our views; there is no way in which one could change one's own beliefs at will. They tend instead to vary if some new evidence becomes available.¹⁰ We should rather say that one *is forced* to change one's beliefs to track the newly available evidence or, in other words, that the new available evidence imposes a certain change in one's beliefs. I am, for instance, forced to believe that my fingertips are on my keyboard as I type this sentence. I cannot intelligibly choose not to believe it because I see my fingertips in contact with the keys. This is what evaluative control amounts to: one's views are controlled by the evidence one may eventually access. Something similar happens with the conclusion of a mathematical proof when the proof is understood. The conclusion imposes itself, but it is not a kind of imposition that degrades the self; on the contrary, one's agency is enhanced or enriched by accepting a theorem as the outcome of a mathematical proof.

The same line of argument applies to several other social practices and institutions such as engaging in a meaningful conversation, cultivating a friendship, or some central experiences of parenthood. This reveals to what extent this sort of imposition plays a central role in our lives and cannot be discarded as marginal or ancillary. In all these cases, *an order*, a sort of necessity, *is imposed upon the agent*, but that imposition, far from oppressing or enslaving her, contributes to her expansion and flourishing. This sort of imposition involves, as we see, a sensitivity or receptivity to the way things are arranged out there in the world. We can thus conclude that the imposition of an order upon the agent and her receptivity to this order are two sides of the same coin. But imposition and receptivity point, in turn, to the idea of passivity. What sort of passivity is this?

contrasted 'graceful' with 'graceless' or 'clumsy' but I wanted (a) to preserve some continuity with John Ruskin's remark and (b) to use a less derogatory word than 'graceless', since the unimaginative dancer can still be a competent dancer.

¹⁰ There is, indeed, room for recalcitrant beliefs, that is, beliefs whose dispositional component persists despite the agent's acceptance in light of evidence of the contrary belief. There are also cases where one's epistemic vices prevent a proper assessment of the available evidence. My point is, however, that such cases are to be construed as exceptional or deviant and, therefore, that in paradigmatic cases agents have evaluative control over their beliefs.

3.2 Base vs. receptive passivity

We must distinguish between *base* and *receptive* passivity.¹¹ Base passivity is the standard notion of passivity, namely, the one concerned with an agent's yielding to the power of passions (and, therefore, that of her vices, epistemic or otherwise) to the detriment of reason. This kind of passivity is base insofar as indulging in it is assumed to degrade the self. The *locus* of this kind of passivity is a conception of the self as divided between those aspects of it which the self truly identifies with and those others that are alienated from it. Within this approach, all passions may be regarded as essentially base inasmuch they constitutively belong to the alienated parts of the self.¹² No matter whether any such passion may eventually coincide with the dictates of reason, this will be only accidentally so, for there is always the chance that any given passion might lead us away from reason. For my purposes, I can just say that, within this framework, passions constitute a system of forces that the true self must make every effort to keep under control, and base passivity describes the eventual incapacity of the self to resist such forces.

Receptive passivity, on the contrary, is at odds with the divided conception of the self. To substantiate this claim, we may turn again to the graceful dancer and consider the following question: *what is the graceful dancer's true self?* It is certainly hard to imagine what her true self might consist of, what parts of herself should be detached or alienated in her dancing. It seems instead that dancing gracefully has to do with the articulation of music, bodily experiences, emotions, decisions, and actions. As we see, the point is not detachment but what we may coin as *agential articulation*, where 'agential' alludes both to what is articulated and to one's own contribution to this process of articulation; and, if I am right, this contribution has more to do with a certain kind of passivity, of letting oneself go once a particular kind of *gestalt* has been formed, than with the notion of activity associated with the effort of the will or with the formation of an intention.¹³ After all, agential articulation could hardly

¹¹ 'Base' is primarily an evaluative term whereas 'receptive passivity' comes also with a descriptive element, and this asymmetry may sound problematic. The distinction between thin and thick concepts may be of some use here. I would say that 'base' is a thin concept that comes with a negative evaluative import, whereas 'receptive passivity' is a thick one, insofar as it involves both a descriptive and an evaluative component. It is the thickness of the latter term that makes it suitable to account for a kind of attitude towards oneself that departs from the Detachment Assumption, which, as we have seen, presupposes that mental states can have only one or another direction of fit.

¹² Some may find the contrast between reason and passions inadequate insofar as it does not consider the subtleties of the current debate on the nature of emotions and, in general, of affective states. Still, my purpose here is not to elucidate the nature of affective states, which is clearly beyond the scope of this paper, but only to sketch a notion of passivity associated with the divided conception of the self—and, therefore, with the Detachment Assumption—that may be interfering with a proper understanding of the notion of receptive passivity that, in turn, I regard as indispensable to make sense of the kind of self-knowledge that might have a transformative effect. See Korsgaard (1996, 2009), Blackburn (1998) and Dunn (1996) for the combat between reason and passion that I have in mind.

¹³ I therefore depart from the Kantian idea that agency has mainly to do with being active while passivity leads one's agency astray. Boyle (2009) presents the Kantian sort of activity as constitutive of a fundamental kind of self-knowledge: "The Kantian contrast between an active and a passive form of self-knowledge has been a source of puzzlement to commentators. Our discussion, however, has equipped us to see a point in the distinction. For on the one hand, we have seen that it is attractive to understand our knowledge of

be reached if the dancer adopted a detached attitude towards her own emotions and bodily experiences, for such an attitude presupposes a divided conception of the self, so that one's emotions and bodily experiences will merely be observed as a passing show.¹⁴

To put it another way, the point of receptive passivity is not to vindicate that our emotions and bodily experiences belong to our true self but rather to call into question the divided conception of the self, since this notion of passivity only makes sense within a framework where the clear-cut contrast between the true self and its alienated parts has been abandoned. Once we give up the divided conception of the self, we may still regard *some passions as base*, but we can no longer dismiss all passions as such. The fact that we may identify a certain passion as base will partly depend on its role in the agent's outlook about what is worth pursuing. Moreover, receptive passivity may favour not so much a pure repression of that base passion, but the search for a more appropriate expression of the needs lying behind it (Williams, 2002, ch. 10; Corbí, 2012, ch. 6, 2017). We can thus understand how receptive passivity may contribute to agential articulation and, more specifically, to attenuate one's epistemic vices. Recalcitrant cases of epistemic vice have to do, as we have seen, with the mismatch between the agent's epistemic dispositions and the kind of epistemic attitude that the agent may judge appropriate. Receptive passivity is a kind of awareness that goes precisely in the direction of unifying these conflicting elements by, among other things, endowing one's bodily responses and experiences (including, hence, one's dispositions) with *prima facie* authority. Each experience, even those that have to do with one's epistemic vices, must be approached as expressing a need to be discerned and met rather than thoroughly dismissed.

Some might reply, however, that Inferentialism can make sense of the kind of self-examination that the graceful dancer illustrates, so that receptive passivity may ultimately be accommodated within an inferentialist framework. From this perspective, Krista Lawlor stresses, in dispute with Moran, "... that *inference from internal promptings* is a routine means by which we know what we want" (Lawlor, 2009, p. 48; see Cassam, 2014), which ought to be distinguished from knowing what *to* want. Among such internal promptings, Lawlor includes simple sensations but also imaged natural language sentences and visual images that may come up inadvertently or be deliberately prompted by the agent herself. Such promptings would require a causal interpretation to determine what desires lie behind them. Lawlor considers the case of a woman, Katherine, that catches herself imagining, remembering, and feeling a range of things connected to the idea of having a second child. She may thus be interested in answering the question 'Do I want a second child?' Lawlor argues that the way to proceed is to infer the cause of such promptings, that is, to infer whether their

what we believe as reflecting our capacity for a kind of agency --the capacity to make up our minds on the basis of grounds for belief." (Boyle, 2009, p.158; see pp. 133-4). I could grant Boyle's general point and still make room for a fundamental kind of self-knowledge derived from the sense of agency that receptive passivity enhances.

¹⁴ It follows that Inferentialism, insofar as it is committed to the Detachment Assumption, can hardly solve the alienation problem. It is only by renouncing this assumption and shifting to an alternative view of how a self may be sensitive to the nuances of her own experience of the world, that we can make sense of the idea of identification.

cause is or is not a desire to have a second child. If I mention this sort of self-examination it is because one might think that what I call receptive passivity could reduce to this and, therefore, that my line of reasoning poses no problem to Inferentialism.

Lawlor's focus on internal promptings points to a crucial phenomenon (Boyle, 2015, p. 344) whose proper significance I seek to elucidate in terms of the notion of receptive passivity; after all, such internal promptings should relevantly figure among the objects of the kind of attention I have identified as receptive passivity. But this attitude crucially departs from the Detachment Assumption, which is, nevertheless, central to the notion of causal interpretation as Lawlor presents it. For this notion presupposes that (a) the agent bears a relation *R* to her internal promptings and to their cause and (b) that one's desires are determined regardless of any of the normative constraints and forms of imposition that are constitutive of receptive passivity. So, it seems that, even though receptive passivity and causal interpretation point to the same phenomena, they provide crucially disparate accounts of them. Let me now examine two metaphysical concerns that some might have regarding receptive passivity. The first has to do with the sort of normativity involved in such an attitude and the second with how it is to be individuated.

3.3 Normativity and individuation

The graceful dancer, like the sensitive agent, is subject to some normative constraints. After all, there are ways of responding to the music that do not count as graceful. This is not to say that there is just one perfect or ideal way of dancing a particular piece of music in detriment to the rest. A dancer's actual performance can be assessed in various ways, but not in terms of its distance from an ideal. We must then dispense with the notion of ideal performance to account for the sort of normativity that is involved in graceful dancing and, in general, in the attitude that I have identified as receptive passivity.¹⁵ We may appeal instead to the idea of proportionality between a piece of music and the dancer's response to it (Brewer, 2009; Wiggins, 1987). This notion admits - and even invites - a plurality of approaches to any particular piece of music (Berlin, 2000). Each such approach will highlight one or another aspect of it. Still, some performances will not be recognised as authorised by the work (Walton, 1990, pp. 58–61).

Regarding the second question, that is, how receptive passivity is to be individuated, some progress has already been made: for an agent *A* to be receptively passive to some aspects of a certain situation some normative constraints must be met. These constraints are to be identified not by reference to what an ideal agent ought to do, but in connection with the idea of the proportionality of *A*'s response to the relevant aspects of the situation. This notion of proportionality implies somewhat that the aspects of the situation and those of *A*'s response (including her self-examination) form a unit of intelligibility, that is, that they are individuated with regard to each

¹⁵ The agent, unlike the standard situation for a dancer on the stage, might have to take care of an array of disparate, and even conflicting, social institutions, as if she had to dance various pieces of music at the same time. In such a case, part of what the situation demands from her is to articulate a response to such a musical complexity that might make sense and, if suitable, excel in gracefulness.

other or, in other words, that they form a network of features each of which must be identified in the context of the rest. It follows that circularity is constitutively present in this process of individuation, even though this circularity may end up being epistemically virtuous insofar as it sheds light on some aspects of the situation that had so far been distorted or even remained in the dark. In any event, I borrow the notion of a unit of intelligibility from David Finkelstein's analysis of Wittgenstein's contextualist views and, more specifically, of the sense in which an agent's experiences, her gestures, her self-ascription, and the situation they are a response to are conceptually interdependent (Finkelstein, 2003, ch. 5). Barry Stroud elaborates a transcendental argument regarding evaluative judgements that leads to a similar conclusion (Stroud, 2011, ch. 4). So much for how to individuate receptive passivity. Let us now turn to how this attitude may be shed some light on the epistemically vicious agent.

3.4 Distinctiveness and self-transformation

Like the neurotic agent, there is a point at which the epistemically vicious agent may find herself divided. She may accept in light of evidence that her deliberative capacities are hampered by some epistemic vices, and still be unable to put these vices aside while reasoning or inquiring. Receptive passivity is a kind of attention that permits us to *make sense* of how exposure to certain situations, or even traumatic experiences, may have a transformative impact in the direction of epistemic virtue, for it certainly makes room for the combination of sensitivity and engagement that such a transformation requires. In other words, we can say that receptive passivity meets the conditions established for R^* and, therefore, that such a kind of passivity may intelligibly account for the sort of self-improvement that Cassam postulates.

Some may object however that, insofar as we are dealing with epistemic vices, they will still contaminate and distort our ability to focus our attention on one or another aspect of the situation; consequently, receptive passivity by itself could hardly be a way out of the trap the epistemically vicious agent is caught in. There is, of course, no safe route to epistemic self-improvement. Nevertheless, there are ways to practice the kind of attention that I have characterised as receptive passivity that may reduce our vulnerability to self-deception. Such ways have to do, for instance, with our ability to focus on rather formal aspects of our own experience and with the way some images and words may be sensed as anchored to certain parts of one's body. In general, I would say that such means are not alien to the way a dancer learns to be graceful, that is, by seeking to passively discern nuances of expression in the piece of music as well as in herself and then let herself go. But I should leave the details for some other occasion because my purpose in this section is just to outline a kind of attention (i.e., receptive passivity) that may put some flesh on the relation R^* that I have been vindicating as indispensable in accounting for a certain kind of self-transformation and, consequently, for the kind of self-knowledge that it requires.

Such processes of self-knowledge are, indeed, existentially substantial. The question is whether they are also distinctively first-personal; apparently, they are insofar as a third-party access to someone else's experience is confined to the deliverances of a detached, theoretical attitude. Some could then argue that an agent's ability to perceive someone else's psychological attitudes often requires some sort of empathetic

response that, in turn, I would have to conceptualise as a case of receptive passivity. After all, if this kind of passivity is involved in a graceful appreciation of a piece of music, it must also be present in our ability to have an intimate conversation with a friend, where a nuanced sensitivity to each other's psychological attitudes plays a crucial role. Seeing that one's friend is sad, angry, or kind might then require to be passively receptive to how her gestures and behaviour resonate within oneself. We could finally generalise and conclude that receptive passivity is involved in one's access to other people's psychological condition and, therefore, that this kind of attitude could hardly make self-knowledge distinctive.

My reply to this objection is that what makes receptive passivity distinctively first-personal is not the abandonment of the Detachment Assumption as such, but the kind of agential articulation involved in this attitude, namely, the capacity to let oneself go once one has discerned and acknowledged the normative significance of one's own bodily experiences and emotional responses (Weil, 1963; Williams, 2002, ch. 2). The graceful dancer may thereby apprehend the order in a piece of music and the order in her own response too. The order in the music is not altered in the least by the dancer's effort to discern and acknowledge how best to respond to it given what she is. By contrast, she should expect her own response to the piece of music to be modified and shaped by this process of discernment and acknowledgement. Success in this process of self-transformation will count as a criterion of epistemic enlightenment, that is, of the fact that she has managed to apprehend how the piece of music resonates within her and, therefore, how she is forced to respond to it (Williams, 1981, 1993, 2002). Complementarily, failure in this process of self-transformation will count as a reason to engage in further processes of discernment along the lines suggested by the idea of receptive passivity.

As we see, this process of agential articulation is both strictly first-personal and epistemically distinctive. It is *strictly first-personal* due to the agential component, once it is conceived of in terms of receptive passivity, and is it *epistemically distinctive* because (a) it includes an epistemic component insofar as one's discernment is vulnerable to all sorts of mistakes and distortions but (b), unlike an external order such as that of a piece of music, the order in one's bodily experiences, emotional responses and character is shaped through the agential process of discernment and acknowledgement. In this respect, someone else's mental states and attitudes constitute an order as external to oneself as that of a piece of music. A third party may exercise receptive passivity to grasp such mental states and attitudes, but this party cannot shape them the way the graceful dancer does, once she has properly discerned the order in the piece of music that reveals and articulates who she is and what she is forced to become.¹⁶

¹⁶ I must emphasise that the epistemically vicious agent, like the neurotic agent, may in the end be confronted with a rather tragic situation. She may find herself in some circumstances to which her epistemic vices may count as a proportional response exactly in the same way in which an agent's neurosis might be the most appropriate attitude given her social environment and character. In other words, it may occasionally turn out that the preservation of a certain epistemic vice is the most appropriate response to the agent's plight. After all, there are further values than faithfulness to truth or to the available evidence, and they may eventually conflict with each other. An agent may confront a situation where a particular epistemic value or virtue could reasonably be neglected or denied for the sake of some other values. Some might

4 To sum up

In this paper I have distinguished between trivial and substantial cases of self-knowledge. In trivial cases, the impropriety of providing evidence enhances the agent's authority and makes such cases epistemically distinctive. On the contrary, substantial cases of self-knowledge must rely on evidence and, as a result, they should behave like any other sort of knowledge. This line of argument presupposes the Detachment Assumption, namely, that evidence must necessarily be gathered from a detached, theoretical perspective like that commonly attributed to a third-party.

Both Inferentialism and Rationalism are committed to the Detachment Assumption. In Cassam's case, this commitment shows both in his discussion about the distinctiveness of self-knowledge and in his analysis of the experience of alienation. Regarding the former, Cassam argues that the most we can do to accommodate the distinctiveness of self-knowledge is to emphasise that the agent and a third-party do not have access to the same kinds of evidence, since, for instance, internal promptings are only available to the agent herself and not to a third-party. Cassam seems then to assume that the agent and a third-party should bear the same relation R to any sort of evidence and, therefore, that R must involve the kind of detachment that sounds constitutive of a third-person perspective. Something similar happens with the idea of alienation. It is true that Inferentialism—and, in general, self-knowledge based on evidence—allows for an agent to be identified with an attitude that she has discovered inferentially, but it remains silent as to what should be added to the discovery for identification to occur. The worry is that, unless an alternative relation R^* towards one's own attitude is properly elucidated, we cannot make sense of the experience of identification. Moreover, I have argued that the same relation R^* is required to understand the sort of self-transformation that Cassam commends in his approach to epistemic vices, since, in recalcitrant cases, inferential awareness of one's epistemic vices will be a first step, but only a first. What stands in the way of discerning what other sort of self-awareness is required is precisely the Detachment Assumption, since this alternative kind of self-awareness involves a relation R^* to oneself that is both sensitive to evidence and engaged.

In the final sections I have tried to elucidate what this relation R^* might look like. In this respect, I have distinguished two kinds of passivity, that is, base and receptive passivity. To explore the latter, I have considered the contrast between the graceful and the unimaginative dancers. The dance in the unimaginative case is governed by a set of rules and, therefore, it comes with a certain degree of rigidity. The graceful dancer is, instead, in tune with the order in the music and with her bodily experiences as well. There's no way in which the dancer could detach herself from her bodily experiences and still be graceful. This engagement with her bodily experiences is part of the process by which she discerns the order in the piece of music and articulates

reply that the source of an agent's epistemic vices, especially those associated with her character, could eventually be traced back to some sort of psychic impairment or trauma, so that receptive passivity might help her to recover from it and lead, in the end, to a higher sort of agential articulation with no epistemic cost. Still this possibility, even though it may eventually work for a certain epistemic vice or blind spot, could hardly be granted as a general procedure unless one were ready to assume that all human values can ideally be squared, which I rather doubt (Berlin, 2000).

her dance. Her attitude is both strictly first-personal and epistemically distinctive. It is strictly first-personal due to the agential component, once it is conceived of in terms of receptive passivity, and is it epistemically distinctive because her own bodily experiences are shaped through her process of discerning how the piece of music resonates within her.

Acknowledgements I am indebted to my colleagues in the Valencia Philosophy Lab and to two anonymous reviewers for their insightful comments on earlier versions of this paper. Research for this paper has been funded by the University of Valencia and by the Spanish Ministry of Science and Innovation (PID2019-106420GA-I00, PID2020-119588GB-I00).

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

Declarations The author has no relevant financial or non-financial interest to disclose.

Conflict of interest I declare that there are no financial or non-financial interests that are directly or indirectly related to this manuscript submitted for publication.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Anscombe, G. E. M. (1963). *Intention* (2nd edition.). Cornell University Press.
- Berlin, I. (2000). The pursuit of the ideal. In I. Berlin *The Proper Study of Mankind*. Farrar, Straus and Giroux, 1–16.
- Blackburn, S. (1998). *Ruling passions. A theory of practical reasoning*. Clarendon Press.
- Boyle, M. (2009). Two kinds of self-knowledge. *Philosophy and Phenomenological Research*, 77, 133–164. <https://doi.org/10.1111/j.1933-1592.2008.00235.x>.
- Boyle, M. (2011). Transparent self-knowledge. *Proceedings of the Aristotelian Society Supplementary Volume, LXXXV*, 223–241. <https://doi.org/10.1111/j.1467-8349.2011.00204.x>.
- Boyle, M. (2015). Critical study: Cassam on self-knowledge for humans. *European Journal of Philosophy*, 23(2), 337–348. <https://doi.org/10.1111/ejop.12117>.
- Brewer, T. (2009). *The retrieval of ethics*. Oxford University Press.
- Byrne, A. (2011). Transparency, belief, intention. *Aristotelian Society Supplementary Volume*, 85, 201–221. <https://doi.org/10.1111/j.1467-8349.2011.00203.x>.
- Cassam, Q. (2014). *Self-knowledge for humans*. Oxford University Press.
- Cassam, Q. (2017). What asymmetry? Knowledge of self, knowledge of others, and the inferentialist challenge. *Synthese*, 194, 723–741. <https://doi.org/10.1007/s11229-015-0772-7>.
- Cassam, Q. (2019). *Vices of the mind*. Oxford University Press.
- Corbí, J. E. (2012). *Morality, self-knowledge, and human suffering*. Routledge.
- Corbí, J. E. (2017). Self-knowledge, deliberation and the fruit of Satan. *Acta Analytica*, 32(2), 245–261.
- Dunn, R. (2006). *Values and the reflective point of view: On expressivism, self-knowledge and agency*. Ashgate.
- Edgley, R. (1969). *Reason in theory and practice*. Hutchinson.

- Evans, G. (1982). *The varieties of reference*. Oxford University Press.
- Fernández, J. (2013). *Transparent minds: A study of self-knowledge*. Oxford University Press.
- Finkelstein, D. (2003). *Expression and the inner*. Harvard University Press.
- Finkelstein, D. (2012). From transparency to expressivism. In J. Conant, & G. Abel (Eds.), *Rethinking Epistemology* (2v. vol.). De Gruyter. 101–18.
- Frost, K. (2014). On the very idea of direction of fit. *Philosophical Review*, 123, 429–484. <https://doi.org/10.1215/00318108-2749720>.
- Gallois, A. (2008). *The world without, the mind within*. Cambridge University Press.
- Gertler, B. (2011). *Self-knowledge*. Routledge.
- Gertler, B. (2018). Self-knowledge and rational agency: A defense of empiricism. *Philosophy and Phenomenological Research*, 96, 91–109. <https://doi.org/10.1111/phpr.12288>.
- Gertler, B. (2021). Self-knowledge. In: Edward N. Zalta (ed.). *The Stanford Encyclopedia of Philosophy*, <https://plato.stanford.edu/archives/win2021/entries/self-knowledge/>.
- Gregory, A. (2012). Changing direction on direction of fit. *Ethical Theory and Moral Practice*, 15, 603–614. <https://doi.org/10.1007/s10677-012-9355-6>.
- Hampshire, S. (1957). *Freedom of the individual*. Princeton University Press.
- Hieronymi, P. (2006). Controlling attitudes. *Pacific Philosophical Quarterly*, 87, 45–74. <https://doi.org/10.1111/j.1468-0114.2006.00247.x>
- Humberstone, I. L. (1992). Direction of fit. *Mind*, 101, 59–83. <https://doi.org/10.1093/mind/101.401.59>.
- Korsgaard, C. (1996). *The sources of normativity*. Cambridge University Press.
- Korsgaard, C. (2009). *Self-constitution: Agency, identity, and integrity*. Harvard University Press.
- Lawlor, K. (2009). Knowing what one wants. *Philosophy and Phenomenological Research*, 79, 47–75. <https://doi.org/10.1111/j.1933-1592.2009.00266.x>.
- Little, M. (1997). Virtue as Knowledge: Objections from the philosophy of mind. *Noûs*, 31, 59–79. <https://doi.org/10.1111/0029-4624.00035>.
- Moran, R. (2001). *Authority and estrangement: An essay on self-knowledge*. Princeton University Press.
- Paul, L. A. (2014). *Transformative experience*. Oxford University Press.
- Rushmer, R., & Davies, H. T. O. (2004). Unlearning in healthcare. *BMJ Quality and Safety in Healthcare*, 13(Suppl II), ii10–ii15. <https://doi.org/10.1136/qshc.2003.009506>.
- Ruskin, J. (2004). *Selected writings*. Oxford University Press.
- Schueler, G. (1995). *Desire: Its role in practical reason and the explanation of action*. MIT Press.
- Smith, M. (1994). *The Moral problem*. Blackwell.
- Stroud, B. (2011). *Engagement and metaphysical dissatisfaction*. Oxford University Press.
- Walton, K. (1990). *Mimesis as make-believe*. Harvard University Press.
- Weil, S. (1963). *Gravity and grace*. Routledge.
- Wiggins, D. (1987). Deliberation and practical reason. *Needs, values, Truth* (pp. 215–238). Oxford University Press.
- Williams, B. (1981). Practical necessity. *Moral Luck* (pp. 124–131). Cambridge University Press.
- Williams, B. (1993). *Shame and necessity*. University of California Press
- Williams, B. (2002). *Truth and truthfulness*. Princeton and Oxford: Princeton University Press.
- Zangwill, N. (2008). Besires and the motivation debate. *Theoria*, 74, 50–59. <https://doi.org/10.1111/j.1755-2567.2008.00001.x>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.