Canadian Journal of Philosophy (2025), 55: 3, 247–260
doi:10.1017/can.2025.10046

SYMPOSIUM

Replies to Bar-On, Barnett, and Brink

Matthew Boyle

Department of Philosophy, University of Chicago, Chicago, USA
Email: mbboyle@uchicago.edu

Abstract
I reply to three critical discussions of my book, *Transparency and Reflection* (Oxford, 2024). The replies discuss the basic structure of my "reflectivist" account of self-knowledge, the bearing of my account on the distinction between rational and nonrational minds, the question of how to respond to Hume's challenge to our entitlement to attribute our thoughts to a single self, the relation between awareness of ourselves as conscious subjects and knowledge of our existence as embodied objects, and the relation of my views on self-awareness to the views of Immanuel Kant.

Keywords: transparency; reflection; self-knowledge; self-awareness; self-consciousness; Immanuel Kant; Jean-Paul Sartre

I want to start by expressing my gratitude to Dorit Bar-On, David Barnett, and Claudi Brink for their insightful comments. I am—predictably—going to try to resist many of their objections, but I feel honored to have my work discussed with such incisiveness and good sense. Each of their comments has helped me to see my project in a new light, and each has compelled me to pass beyond the formulations I reached in my book *Transparency and Reflection* (henceforth, T&R) to face new questions that I did not see. So I owe each commentator a real debt.

Since Bar-On challenges various basic ideas of my reflectivist approach to self-knowledge, it will be natural to address her comment first. Barnett and Brink both challenge my reflectivist account of our notion of the subject who has thoughts and attitudes, so it makes sense to turn to their concerns only once the viability of the general approach has been clarified. And of these latter two commentators, it seems appropriate to respond to Barnett first, since he challenges my very distinction between positing a subject who thinks and positing a being who is the thinker. Brink, by contrast, invites me to consider, in a more open-ended way, the relation of my account to various Kantian ideas about the nature of apperceptive awareness. I am grateful for the invitation, and since it will turn out that I need to draw on Kant in order to respond to Barnett, this will set the stage for the more open-ended reckoning for which Brink's comment calls.

1. Reply to Bar-On

Bar-On gives a lucid overview of my reflectivist approach to self-knowledge and raises several objections to it. I will focus on her main challenge: that there is an "inherent instability" in my position that results from my seeking to meet two quite different desiderata (Bar-On, n.d.). On the one hand, I aim to bring out the attractiveness of the Kantian idea that we human beings, as rational animals, possess a form of self-awareness that fundamentally distinguishes our cognition from the cognition of nonrational animals. On the other hand, I aim to give an account of self-knowledge that does not over-intellectualize human cognition, acknowledging that we can have first-order attitudes

© The Author(s), 2026. Published by Cambridge University Press on behalf of The Canadian Journal of Philosophy, Inc. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/can.2025.10046 Published online by Cambridge University Press

such as beliefs and desires without having any explicit beliefs about our own first-order attitudes, and indeed without having the concepts required to formulate such beliefs. Bar-On suggests that these two aims pull in contrary directions, so that when I seek to meet the first aim, I jeopardize the second, whereas when I seek to meet the second aim, I jeopardize the first.

Bar-On is certainly right that I have these two aims, and I admit that, on first encounter, they may appear to be in tension. But I think my reflectivist approach to self-knowledge shows that the tension between them is merely apparent. So to respond to Bar-On's challenge, I need to review some basic ideas of the reflectivist approach.

The first point I should stress is that the primary motivation for my project is the widely-discussed "problem of transparency" about self-knowledge, a problem that can be stated without presupposing any determinate views about the differences between rational and nonrational animals, the intellectual requirements on self-awareness, and so forth The problem of transparency arises when we observe that, for a wide range of mental states, a person can normally answer questions about her own mental state by treating them as "transparent" to corresponding questions about the nonmental world. Thus—to take Gareth Evans's familiar example—one can normally answer the question whether *one believes that there will be a third world war* simply by answering the question whether *there will be a third world war* and then prefixing "I believe that …" to one's answer. Similarly, one can normally answer the question *how one's perceptual experience presents one's environment* simply by answering the question *how things stand in one's environment* (while restricting one's answer to this latter question in certain specifiable ways) and then prefixing "It looks/sounds/feels to me as if …" to one's response. Such observations present a *prima facie* puzzle: What entitles us to suppose that we can answer questions about our own mental states simply by answering questions about the nonmental world? After all, these two topics are fundamentally different: the one concerns my own representational state; the other, a state of affairs that is independent of my representational state.

Of course, if I answer the latter, world-directed sort of question with sincerity, I will normally *express* whatever attitude I hold on that question, and Bar-On is a leading advocate of the "expressivist" approach of self-knowledge, according to which this point about expression holds the key to understanding our capacity for privileged self-knowledge. Expressivists argue that, inasmuch as we can learn to make psychological self-ascriptions in a way that expresses the very states we ascribe, such self-ascriptions will be reliably true, and indeed, they will plausibly express knowledge of our own mental states (cf. Bar-On, 2004; Bar-On & Wright, 2023). But I think this leaves a crucial part of the problem of transparency unresolved: it does not explain how the person who makes the relevant self-ascription can reasonably see her answer to the world-directed question as bearing on the corresponding representational question. On the face of it, the conclusion *There will be a third world war* does not support the claim *I believe there will be a third world war*: the latter might easily be false, although the former was true. To be sure, if in stating that there will be a third world war, I express my sincere view on the question, then it follows that I believe there will be a third war. But to know that this is my sincere view, I would need *already* to know my own belief about this topic, so this point cannot explain what enables me to know my beliefs in a transparent way.

So although I agree with Bar-On that transparent self-ascriptions of mental states characteristically express the very states they ascribe, I think this observation does not solve the problem of transparency, but merely highlights the point that needs to be explained. We need to explain how a person's disposition to answer questions about her own mental states by treating them as transparent to questions about the nonmental world can be understood by her, not just as a black box from which she can learn *post hoc* (as in "This is what I am disposed to say, so this is what my attitude must be"), but as a reasonable way to judge *ex ante*. That, anyway, is my view: Bar-On and I have argued back and forth about this matter for some years, and my formulations here are just variations of ones I have tried on other occasions, so I do not expect this latest effort to change her mind. My purpose in restating my position is simply to bring out what is distinctive in my

https://doi.org/10.1017/can.2025.10046 Published online by Cambridge University Press

reflectivist approach: it seeks to show, as the expressivist approach does not, that the relevant ways of thinking about the world contain an implicit *justification* for the corresponding ways of thinking about my own mental state. This would, I think, provide a particularly satisfying response to the problem of transparency, one that clarifies the reasonableness of the transparency transition from the subject's own point of view.

The easiest way to explain the reflectivist approach is with an example. Suppose someone asks me what I presently see.[1] The transparency observation says that I can answer this question, not by observing myself or using some special power of introspection to discern my own mental states, but simply by describing my environment (while setting aside extraneous information and confining my descriptions to the kinds of properties that are available to vision). But this observation raises the puzzle just noted: what entitles me to suppose that I can answer a question about my own visual experience by treating it as transparent to a question about the nonmental world? Suppose that what I see is a brown wooden tabletop with a white coffee mug resting on it. How does my observation of these non-mental items warrant me in asserting *that I see* these things? After all, none of the properties I can observe—the color of the table, the features of the coffee mug, the pattern of light falling upon them—imply that these things are seen. How does our visually-based knowledge of *what* we see justify us in supposing *that we see* it?[2]

The reflectivist replies that our visual experience not only makes us aware of the things we see; it presents these things in a distinctively visual *mode*, one that it is natural to express verbally by a certain use of the demonstrative "this," as in

(1) *This* is a brown table with a white coffee mug on it.

If we reflect on how we use the relevant "this," we will notice that we use it just when we stand in a certain kind of relation to the objects in question, namely the relation of *seeing* them. If this scene goes out of view, we will no longer speak of "this" table (etc.), and if (e.g.) the coffee cup moves, we will use a single "this" to refer to it only so long as it remains continuously visible. I submit that anyone who can describe what they see must know how to use "this" (or some equivalent marker of visual awareness) in the way just described. But this is not to say that anyone who can describe what they see must possess concepts of vision like *see*, *vision*, *visibility*, and so forth. A small child might be perfectly capable of describing *what* she sees (and so of using "this" in the relevant way) but might not yet possess concepts of her own power of visual perception. Indeed, any of us might in principle possess the capacity to describe what we see without possessing concepts of vision, although in practice most of us learn such concepts early in our lives. Be that as it may, there remains a distinction of principle between the kind of understanding required to describe *what* one sees and the kind of understanding required for explicit thinking *about* one's own seeing. The capacity to use "this" in the way just described is required for the former purpose; concepts of vision are required only for the latter.

That concepts of vision are not required for thinking thoughts such as (1) is evident from the fact that a person who thinks (1) does not think *about* herself or her own visual experience at all, but only about nonmental things: the table with the coffee mug on it. Nevertheless, she thinks about these things in a distinctive mode that expresses her visual relationship to them. We mark this mode by using "this" in a certain familiar way, but what should interest us is not the verbal device by which we mark this mode of presentation, but the mode we thereby mark. We have seen that we use "this"

[1]Philosophers may prefer to reformulate this question as "How do things visually appear to you?" so as to restrict attention on the subject's own visual impressions without committing her to claims about the actual state of her environment. All the points I make would also apply (*mutatis mutandis*) to this more cautious formulation, but in the text I will focus on the more natural formulation, "What do you see?"

[2]This way of posing (what I call) the problem of transparency was emphasized in a series of forceful papers by Fred Dretske. See for instance Dretske, 2003.

https://doi.org/10.1017/can.2025.10046 Published online by Cambridge University Press

in the relevant way when and only when we see (or presume ourselves to see) the demonstrated scene. So we could not distinguish cases in which we were in a position to think of *this table* or *this coffee mug* if we were not at least implicitly aware of our own seeing. But the required awareness of seeing is implicit rather than explicit: it does not take the form of a thought *that* I see such-and-such, but of a distinctive manner of representing *what* I see. Following Sartre, I call this implicit form of awareness of my own seeing "nonpositional consciousness," since it makes me conscious of my own seeing without "positing" my seeing as a topic of representation in its own right. This distinctively visual mode of presentation thus expresses an awareness of seeing that antecedes conceptual thought about vision.

My reflectivist response to the problem of transparency holds that this nonpositional consciousness of seeing expressed in the visual mode of presentation of objects is what justifies me in making the transition from an answer to the question how things stand in my environment to a thought about my own visual state. When I think the self-ascriptive thought

(2) *I see* a brown table with a white coffee mug on it

I am, so to speak, simply unpacking what was already implicit in the "this" in (1). In order to unpack the significance of the visual "this," I must of course possess the concept of seeing as well as a notion of myself, the subject who sees and has other mental states. I go on to argue, however, that these concepts of subjectivity (as we may call them) can themselves be acquired through mere reflection on what is implicit in the awareness expressed in world-directed thoughts like (1).[3] In this way, both the general concepts on which our capacity to think self-ascriptive thoughts depends and the awareness of the specific mental state that (2) ascribes are grounded in a nonpositional consciousness of seeing that is already involved in seeing itself. Moreover, I argue that analogous points can be made, not just about vision, but about a broad range of mental states (belief, desire, intention, etc.): we can know ourselves to be in such states simply by making explicit an awareness that is already implicit in the distinctive modes by which such states present their objects.[4]

"Reflection" is my term for this step from an implicit awareness *of* my own mental state to explicit knowledge *that* I am in a certain mental state, and I argue that reflection is a distinctive kind of cognitive step, different from perception, inference, and other familiar forms of knowledge-acquisition. When I come to know something on the basis of perception or inference, I acquire information that I did not already possess, but when I transform something of which I was already implicitly aware into reflective self-knowledge, I do *not* acquire any new information. Rather, I merely alter the form of understanding that I bring to bear on information I already possess, thereby transforming an awareness formerly expressed "nonpositionally" in the mode through which I

[3]Cf. Boyle, 2024a, Chs. 5 and 8.

[4]Cf. Boyle, 2024a, Chs. 1.4, 3.3, 6.5, 8.5, and 9.5. One kind of objection Bar-On makes to my reflectivist approach is that it is not clear how it might apply to certain kinds of conscious states: passing thoughts of the "Penny for your thoughts?" variety, sensations and moods that seem not to have an identifiable object, and so forth. I am more optimistic than Bar-On about the possibility of giving a reflectivist account of these forms of self-awareness, but there is not space for detailed discussion here. For brief discussion of a reflectivist approach to bodily sensations, see Boyle, 2024a, Ch. 6.5. For a broadly reflectivist way of thinking about emotions and moods, see Boyle, 2024b. I have not tried to develop an account of our awareness of passing thoughts, but I suspect that a first step would be to take seriously the idea that to be aware of such thoughts involves being aware of inwardly saying something to oneself. One kind of transparency-based account of awareness of one's own thoughts takes such awareness to involve something like hearing one's own inner speech (cf. Byrne, 2018, Ch. 8). A different kind of account, which I find more congenial, treats such knowledge a special case of agential knowledge of what one is doing (viz., inwardly saying something to oneself). Roessler, 2015 proposes an account along the latter lines, although he does not make a connection with the idea of transparency. But if there can be a transparency-based account of knowledge of our own actions, it should be possible to give a transparency-based account of this special subclass of action awareness.

https://doi.org/10.1017/can.2025.10046 Published online by Cambridge University Press

represent other topics into a topic of representation in its own right. I call this transformation a "mode-to-content shift."

Having outlined my reflectivist approach to self-awareness, I can now respond to Bar-On's objections. Let me start with the charge that my position is "inherently unstable," since

> the less intellectually demanding the Sartrean account of nonpositional awareness…[,] the less plausible it would be to regard Sartrean nonpositional awareness as a form of genuine *self*-awareness that could be readily 'transformed' into positional self-knowledge using epistemically and conceptually minimal resources. (Bar-On, 2025, 225)

Here, I just disagree. I think my reflectivist solution to the problem of transparency shows that there is a form of awareness of our own mental states that differs from explicit self-knowledge, but that nevertheless provides an intelligible ground for such knowledge. In the example we considered, this is the awareness of seeing expressed in the distinctive mode of presentation of what we see, the mode we mark with a perceptual demonstrative "this." (I will continue to focus on this perceptual case, but analogous points would hold for any type of mental state to which the reflectivist account applies.)

Our implicit awareness of our own seeing is not intellectually demanding: as we have seen, a subject can possess such awareness without possessing concepts of vision. Nevertheless, the capacity to represent objects in this distinctive mode implies an implicit awareness of one's own seeing, for seeing is precisely the relation to an object that the relevant mode of presentation marks. And the presence of this implicit awareness provides an intelligible justification for an explicit self-ascription of seeing. True, the mere fact *that there is* a brown table with a white coffee mug on it does not support the claim that I see the relevant table. But the awareness expressed in

(1) *This* is a brown table with a white coffee mug on it.

does justify the thought

(2) *I see* a brown table with a white coffee mug on it.

For the presentation of a table in the distinctive mode marked by "this" already expresses the subject's visual relationship to the table.

Bar-On suggests that I do not give a "properly epistemological account" of the reflective transition from (1) to (2), since I do not clarify "what are the conditions that ensure that one will engage in a cognitively effortless act of reflection upon being in M" or "what renders [the resulting self-ascriptive] belief *warranted*" (Bar-On, 2025, 232). About the first point I suspect there is little of a general kind to say. On my view, any subject who presently sees something will be in a position, other things equal, reflectively to self-ascribe the relevant condition. The qualification "other things equal" is important: doubtless special circumstances can arise (distraction, confusion, repression, bad faith, …) that interfere with a subject's normal power to reflect on her own mental states. I see no reason to demand—or to suppose there can be—a complete enumeration of such forms of interference, any more than we should demand a complete enumeration of, e.g., the possible forms of interference with our power to perceive. In each case, it suffices to give an account of how the relevant cognitive power puts us in a position to know facts of a certain type other things equal, and that is what my reflectivist account seeks to do.[5]

[5]Bar-On also asks whether, on my view, "cognitively equipped subjects who are in M" (i.e., subjects who are in M and who possess the concepts required to think the self-ascriptive thought *I am in M*) *always* reflectively transform their implicit self-awareness into explicit self-knowledge, and if they do not, whether they still have actual knowledge of the relevant states or are merely in a position to know them (Bar-On, 2025, 234). So for the record: I take it that even those of us who are cognitively equipped to make our own mental states reflectively explicit very often do not do so. Presumably most of our mental life passes

https://doi.org/10.1017/can.2025.10046 Published online by Cambridge University Press

As for explaining what renders our reflective self-ascriptions warranted, however, I do take myself to have addressed this question by pointing out that (a) there is an intelligible relationship between the soundness of thinking (1) and the soundness of thinking (2), and (b) this relationship is one that the subject herself would be in a position to appreciate, provided she had the relevant concepts of subjectivity. And as I have said, I go on to give an account of how she might frame such concepts of subjectivity through mere reflection, though I have not attempted to summarize the latter account here.

Bar-On's charge that my account is unstable rests partly on her view that I cannot vindicate the Kantian intuition that human minds are essentially self-aware, and in this way fundamentally different from the minds of nonrational animals, without jeopardizing my modesty about the intellectual requirements for nonpositional self-awareness. She argues that

> the more nonintellectual and outward-looking nonpositional awareness is understood to be, … the less plausible the Kantian link between (our) mentality and self-knowledge would seem… But the more our first-order mental states are portrayed as essentially dependent on some capacities for self-understanding, the more Boyle's account would be open to the charge of over-intellectualizing first-order mentality. (Bar-On, 2025, 225)

I do indeed think of my account as vindicating the broadly Kantian idea that a certain kind of self-awareness belongs essentially to our first-order mental states—though I interpret this idea in terms of a "nonpositional" self-awareness that may differ from what Kant had in mind by "apperception."[6] I also think of my account as vindicating Kant's idea that the capacity to *reflect* on our mental states is a necessary consequence of rationality—i.e., that nothing beyond general rationality is required to enable us to frame the concepts of subjectivity that allow us to "accompany our representations with the *I think*" (as Kant famously puts it). Finally, I think of my account as vindicating Kant's thought that our necessary self-awareness is the source of a distinctively philosophical type of knowledge whose special purview is the scope and limits of our own power of cognition. In T&R, I show that versions of all these thoughts can be defended within a reflectivist framework without sacrificing modesty about the intellectual requirements of pre-reflective self-awareness. This shows, I think, that my position is not unstable.[7] My position does, however, differ from the orthodox Kantian position, inasmuch as I hold that a rational subject need not *actually* possess the concepts of subjectivity that would enable her to accompany her representations with 'I think', though she must possess the *capacity* to frame such concepts through mere reflection.

These remarks do not address Bar-On's doubts about whether our first-order mental states are "essentially dependent on capacities for self-understanding," and so different in kind from the mental states of nonrational animals. This is a topic about which I have written elsewhere (cf. Boyle, 2012, 2016, 2017), and in T&R I seek to show that my solution to the problem of transparency is consistent with the view I have elsewhere advocated: that rationality transforms our cognition in general, rather than merely enabling us to engage in certain sophisticated forms of cognitive activity

---

by unreflected. Nor do I think we should demand a general account of the conditions under which people *will* actually reflect on their mental states: this will depend on whether the question of their own mental state becomes salient for them, and I do not see why there should be anything general to say about this.

So if a cognitively equipped subject does not reflect on some mental state M that she is in, is she aware of this state? Well, in a sense she is, and in another sense she is not. She is implicitly aware of M in virtue of being presented with its content in a distinctive mode, but she is not explicitly aware of M in the sense that requires representation of the self-ascriptive proposition *I am in M*. I reserve the term "self-knowledge" for the latter, explicit form of self-awareness, so I am committed to saying that such a subject does not have actual self-knowledge of her being in M. But this is just a stipulation about how to use a certain word; it does not affect the substantive point that such a subject is already aware of being in M before she reflects on it.

[6]Claudi Brink argues in her comment that my conception of self-awareness is more Kantian than I suppose. I am grateful for the support, but I think Kant's position remains ambiguous. For further discussion, see my Reply to Brink below.

[7]For a fuller account of these Kantian dimension of my view, see Boyle, 2023.

https://doi.org/10.1017/can.2025.10046 Published online by Cambridge University Press

that our nonrational brethren cannot. But T&R does not seek to make a positive case for this transformative view of rationality, and someone who disputed this view could still profitably engage with all of the main ideas of my book. For these are proposals about how to understand the basis and importance of forms of self-knowledge of which human beings are plainly capable, and these proposals do not depend on my claims about how to distinguish rational from nonrational mentality. Even if one thinks that the mental states nonrational animals are like ours in all important respects, one could still accept my proposals about how we rational knowers make the transition to reflection and why our capacity to do so matters. So I propose to set aside the question of transformation here, although I stand by what I said about it in T&R. Bar-On's comments have made me see, however, that I should have distinguished more clearly between my advocacy for the transformative view of rationality and my general reflectivism about human self-awareness: these two commitments are distinguishable, though I continue to believe they are complementary.

2. Reply to Barnett

In Chapter 5 of T&R, I address the question of how reflection might make us aware, not just of our various *mental states*, but of the single *subject* who is their bearer. To bring out the force of this question, I appeal to a famous remark of David Hume:

> For my part, when I enter most intimately into what I call 'myself,' I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never can catch myself at any time without a perception, and never can observe any thing but the perception.[8]

Hume's point is that what our consciousness presents is always some particular *object* of consciousness (e.g., heat or cold, light or shade), not the *subject* to whom such experiences are presented. This raises a problem about the source of our very idea of such a subject. When we express our self-awareness in judgments, we always ascribe the states of which we are aware to a single subject, "I"; but what is the basis of this idea of a subject who has many experiences, thoughts, feelings, etc.? What justifies me in saying "*I* think, feel, perceive, etc." rather than merely "*There is* thinking, feeling, perceiving, etc."?

In response to this Humean challenge, I argue that we can find a reflective basis for the notion of a subject, not in the contents of consciousness, but in its general structure or form. I illustrate this point by considering the way our spatial awareness presents things as located in a framework implicitly centered on a subject who perceives and acts. When objects in my environment are presented as *ahead* or *behind*, *near* or *far*, *on the left* or *on the right*, these spatial modes of presentation do not explicitly present the subject who perceives, but they are nevertheless implicitly "egocentric" modes of presentation whose significance can only be understood in relation to the perceiving and acting subject in relation to whom the relevant spatial framework is defined. Hence, I argue, each of us is warranted, on reflection, in formulating our egocentric spatial awareness, not just as (e.g.) "*There is* consciousness of a table on the left," but as "*I am* conscious of a table on the left." This response to Hume's challenge is available to each of us, not through an independent consciousness of our own existence, but simply through reflection on general structural presuppositions of our spatial awareness of the nonmental world.

I go on to suggest that this response to Hume's challenge still leaves open the question whether this conscious *subject* who is the bearer of mental states is in fact an *object* in the material world that is the primary focus of consciousness. This is the point at which David Barnett objects to my

[8]Hume, 1979, Section 1.4.6: 252. I also appeal to Lichtenberg's well-known challenge to Descartes's Cogito argument, but reading Gomes, 2024 has made me less certain whether Lichtenberg's challenge is really what I — in the company of many other readers — took it to be.

https://doi.org/10.1017/can.2025.10046 Published online by Cambridge University Press

position. As Barnett recognizes, I go on to argue (in T&R, Chapter 6) that each of us *can* know through mere reflection that, when she says "I," she refers, not to a disembodied subject, but to an embodied human being who inhabits the same material realm as the other objects of which she is conscious. But I treat this as a point in need of further argument, not one already settled by my response to Hume's challenge. Barnett thinks this is unacceptable: he forcefully argues that if I know that my thinking has a *subject*, then I must know that there is some *object*, some *being* who thinks my thoughts:

> Not to be dense here, but I cannot understand [what Boyle could mean]. If my self-knowledge is expressible as "I am thinking about a cat," then what I know is that I am thinking about a cat. But if so, what I know sure seems to entail that *something* is thinking about a cat, even if it is neutral on what kind of thing it is. (Barnett, 2025, 241)

As a first step in responding to Barnett's concern, let me fill in some background to my position. Here too, Kant serves as a source of inspiration. In his famous chapter on the Paralogisms of Pure Reason, Kant considers the attempt by "rational psychologists" such as Descartes to deduce a range of metaphysical conclusions—that the soul is an enduring substance, that it is simple and indivisible, etc.—from the pure apperceptive awareness of our own thinking that we express when we say "I think." He accuses these rational psychologists of confusing "the logical exposition of thinking in general" with "a metaphysical determination of the object" (Kant, 1998, B409), and he goes on to explain that

> the unity of consciousness … is here [viz., in rational psychology] taken for an intuition of the subject as an object, and the category of substance is applied to it. But this unity is only the unity of *thinking*, through which no object is given. (Kant, 1998, B421-2)

Kant's view is thus that rational psychologists are right to suppose that the apperceptive awareness expressed in "I think" is awareness of a certain unity in consciousness, but wrong to infer that this unity must correspond to an object. He does not deny that each of us in fact knows himself to be a single, enduring object (i.e., a substance), but he denies that mere apperception gives us sufficient basis for this knowledge. This is his criticism of Descartes's Cogito argument: the premise, *I think*, is indeed something each of us can rightly affirm on the basis of mere apperception, but the conclusion, *I am*, in effect brings the "I" that thinks under the category of substance, positing it as a single, enduring object, and no such object is given in pure apperceptive awareness.[9]

My reflectivist response to Hume seeks to take a position like Kant's. I argue that we can recognize, on the basis of mere reflection, that all our thoughts are grounded in a single subjective standpoint—a unity we mark by ascribing our thoughts to a single subject, "I." But I maintain that it is a further question whether this subjective unity *in our thinking* corresponds to some real *object* in the world about which we think. After all, our basis for including "I" in "I think" is not a "positional" consciousness *of* something that we use "I" to designate, but a "nonpositional" awareness *in* being conscious of a unity in consciousness itself. What we express with this term is thus a unity in apprehending, not an awareness of some unified entity existing among the things apprehended. So the question whether this subjective unity corresponds to some objective unity is still open: perhaps this unity in consciousness does not correspond to any real being in the world of which we are

[9]Although the point would require a more complicated development, I believe that Sartre holds a broadly similar view. For him, to say that something is a possible object of consciousness is to regard it as having "being in itself," a kind of being that simply "is what it is." Consciousness itself, by contrast, has "being for itself," which "is what it is not and is not what it is." Thus consciousness is the very opposite of a being-in-itself; it is a "nothingness" that exists only in positing objects as what it is not. So to the extent that the unity of consciousness can rightly be expressed with "I"—a point on which Sartre's views are too complicated to review here—this "I" does not signify an object, and to represent it as such involves a false reification of being-for-itself.

https://doi.org/10.1017/can.2025.10046 Published online by Cambridge University Press

conscious. Of course, this would be a mad account of the meaning of "I" to embrace at the end of the day, but the possibility is worth conjuring temporarily for the purpose of making clear what we have already proved and what still remains to be shown.

Barnett thinks this is confused: if consciousness has a "subject," then there must be *something* (some *being* in the metaphysically basic sense, and by the same token, some possible object of knowledge) which is this subject. So let me try to present my idea more carefully. Consider a judgment that ascribes consciousness of some object O. According to Anti-Egoists, this can be schematized as

(3) There is consciousness of O

where (3) merely asserts the existence of consciousness of O, not the existence of a subject who is thus conscious. I argue, however, that reflection supports the ascription of consciousness to a subject, as in:

(4) I am conscious of O.

Now suppose I am right about this. Does it follow that I am a "thinking thing" (*res cogitans*), as Descartes famously puts it? We can reformulate this question as follows: Does our justification for including a singular term on the left-hand side of "am conscious of" in (4), just as such, constitute a justification for introducing a term that can appear on the right-hand side of this phrase? Is a unified *subject* of consciousness, just as such, a possible *object* of consciousness?

In one sense, a subject is surely a kind of object. This is simply a logical point: if I think, then it follows by existential generalization that there is something that thinks, and by the same token, this thing, the thinker, is itself a possible topic of thoughts. Barnett uses this point as a hammer with which to batter my position. But it seems to me that this logical point does not yet clarify what sort of real being, if any, one posits in affirming "I think." Consider: If 2 is between 1 and 3, then there is a number between 1 and 3. If space is the all-embracing framework of places, then there is an all embracing framework of places. And if Hamlet is a fictional character, then there is at least one fictional character. Moreover, all these antecedents are clearly true, so evidently the number 2 and space and Hamlet are "objects" in the logical sense. But are they real beings? Numbers certainly exist in a sense, but the traditional view is that they are not *actual* beings, as they cannot change or stand in causal relations. The reality of space is even more tenuous: it seems not itself to be any kind of being, but a mere structure in which real beings exist. And fictional characters are not real beings at all: when we call them "fictional," we mean that they do not really exist, but are merely said to exist in some fiction. So the logical point that a true singular claim supports an existential generalization does not settle the question of what kind of being, if any, is thereby posited.

I think this is what Kant was getting at when he accused rational psychologists of confusing "the logical exposition of thinking in general" with "a metaphysical determination of the object." Kant held, as I do, that an adequate logical analysis of apperceptive awareness must represent our several thoughts as predicated of a single subject, as in (4) above. He took this logical point to be distinguishable from the metaphysical question of what sort of being, if any, we thereby posit, and he accused rational psychologists of conflating these two issues.

Kant accused rational psychologists of sliding from the logical point that thoughts must be predicated of a thinker to the metaphysical claim that thoughts must inhere in a *substance* that endures through time and can stand in real causal relations with other substances—as if a thinker were an immaterial analogue of the material substances we perceive via our senses. And this was indeed Descartes's view: his claim that he is a thinking thing is not a mere affirmation of the logical truth that *I think* implies *(∃x)(x thinks)*, but a positive step toward the conclusion that he is a

https://doi.org/10.1017/can.2025.10046 Published online by Cambridge University Press

substance "really distinct" from any body.[10] Kant, by contrast, denied that any such argument could succeed, since the mere truth of the proposition *I think* does not prove the existence of a metaphysically robust *thing* that thinks:

> [T]his unity [expressed by the apperceptive "I"] is only the unity of thinking, through which no object is given; and thus the category of substance, which always presupposes a given intuition, cannot be applied to it… Thus the subject of the categories cannot… obtain a concept of itself as an object of the categories. (Kant, 1998, B421-2)

Yet must not a singular proposition be made true by the existence of some real being that bears whatever property is ascribed? That is certainly the usual case, but I take the examples just given—numbers, space, fictional characters—to show that in special cases we may have reason to resist, or at least qualify, this inference from logic to metaphysics. My claim is that the apperceptive *I* is another such special case, for the reason recently given: our ground for introducing it is not a consciousness *of* some object, but a reflective recognition of unity *in* consciousness of any objects whatsoever. Our capacity to reflect does indeed allow us to make this unified I a topic of thoughts, and indeed of a limiting kind of knowledge. But our interpretation of these thoughts and this knowledge must remain true to their reflective source: they do not articulate consciousness of an object, but merely the unity implicit in consciousness itself.[11]

I have formulated this point by saying that, for all that sheer apperceptive consciousness tells us, the apperceptive "I" might not correspond to any real being, any worldly object. But I do not mean —as Barnett proposes on my behalf—that this "I" is indeed "something" but not a "certain kind of thing" (Barnett, 2025, 244), because it lacks some metaphysically significant property and so is disqualified from counting as a "being" in some special, demanding sense. Something that is not a real being is not a being that lacks some property any more than a fake Rolex is a Rolex that lacks some property. A fake Rolex is not a Rolex at all. A fictional person is not a person at all. Just so, to say that something is not a real being is to say that it is not a being at all, not merely that it is not a being of a certain kind. If I am right, showing that the content of apperceptive awareness is rightly formulated as "I think" does not yet show that there is any being whatsoever that thinks.

3. Reply to Brink

Claudi Brink's generous comment also engages with my response to Hume's challenge, and she rightly identifies the Kantian basis of my thinking about this topic. She seeks to argue, however, that my position is even more deeply Kantian than I recognize, and that embracing these Kantian roots would allow me to resolve a difficulty for my view.

As Brink notes, T&R grew out of an attempt to explicate Kant's famous claim that "It must be possible for the *I think* to accompany all of my representations" on pain of their being "nothing to me" (Kant, 1998, B131-2). I was attracted by this Kantian claim because I took it to express the idea, which I thought was crucial, that our capacity for privileged self-knowledge is not merely as a special capacity to discover facts about our own minds, but the articulation of an awareness we all necessarily possess simply in virtue of having rational minds at all. As I tried to give an account

[10]Note the etymological connection between *res* ("thing") and *realis* ("real"). In establishing that he is a thinking thing, Descartes takes himself to establish that he is a real being about which the question can arise whether it is identical to his body. This speaks to Barnett's question about what exactly I deny that Descartes affirms (Barnett, 2025, 244–45). Following Kant, I deny that the proposition *I think*, affirmed on the basis of mere apperception, implies the existence of a real thinking thing or substance: an independent existent that endures through time and can stand in causal relations. But this—to repeat—is only a temporary way-station in my argument: once bodily awareness is brought into the picture, we will be able to affirm that the thinker is an embodied being.

[11]Compare the final sentence of the passage from Kant just quoted.

https://doi.org/10.1017/can.2025.10046 Published online by Cambridge University Press

of this idea, however, I came to think that Kant's focus on explicit self-ascriptions of thinking gives a one-sided picture of self-awareness, one that focuses on the *reflective* expression of this awareness and neglects a more basic kind of self-awareness that is pre-reflective. So my book ultimately became a synthesis of a Kantian view of the significance of human self-awareness with a set of Sartrean ideas about how to conceive of such self-awareness.

Brink is sympathetic to my Sartrean notion of pre-reflective self-awareness, but she argues that this notion is more Kantian than I suppose. She points out that Kant himself speaks of the consciousness expressed in the representation *I* as "not even a representation distinguishing a particular object, but rather a form of representation in general" (Kant, 1998, A346/B404), and she contends that Kant's many remarks emphasizing this "formal" character of apperceptive awareness suggest a view on which apperception is not primarily a reflective awareness of my own thinking but a pre-reflective self-consciousness involved in any thinking whatsoever. She goes on to argue that this pre-reflective self-consciousness is a consciousness of a fundamental unity in our thinking, and she draws interesting comparisons between this consciousness of unity in thinking and our consciousness of the thematic unity in a work of art.

Brink then proposes that this Kantian idea of a consciousness of unity in thinking could help to resolve a problem faced by my account of the representation *I*. As I mentioned in my reply to David Barnett, I propose that we can respond to Hume's challenge to our entitlement to ascribe our various representations to a single self by arguing that the structure of our thinking about *objects* presupposes that these objects are presented to a *subject*. I illustrate this point by appeal to our egocentric thinking about things in space, and I go on to argue that it holds true in a different but related way for our thinking about what is the case more generally. Brink argues, however, that

> even if we can show that each representational domain exhibits the features that justify the introduction of the first-person concept, a more fundamental question concerns the identity of the subject across different representational domains… To make sense of the idea that it is the same subject who believes, perceives, desires, and so on, we would need to make sense of the idea that each of the points of view internal to these various modes are themselves integrated into an overarching point of view. This would seem to require yet another kind of nonpositional consciousness, one that is deeper still than the awareness of a unified epistemic view, which could incorporate the systematic relations not only within but between the various modes of presentation. (Brink, 2025, 215)

It is just here, she suggests, that we can profit from the Kantian idea of a consciousness of unity underlying all thinking. For precisely this consciousness provides a basis for ascribing our various representations to a unified subject.

Let me start with the question whether Kant's "transcendental apperception" amounts to a form of pre-reflective self-consciousness. This is not the place for a detailed discussion of Kant interpretation, but I certainly agree with Brink that Kant's emphasis on the "formal" character of apperceptive consciousness suggests that he views apperception as the organizing principle of all thinking, not just the basis of certain special thoughts we have about our own acts of representation. It must be said, however, that his remarks on this topic are rather cryptic. The main passage Brink quotes is from the Paralogisms chapter, where, in the course of explaining his grounds for rejecting the metaphysical conclusions of rational psychologists, Kant remarks that the consciousness expressed in the apperceptive "I" is "a mere consciousness that accompanies every concept," which is "in itself not even a representation distinguishing a particular object, but rather a form of representation in general, insofar as it is to be called a cognition" (Kant, 1998, A346/B404). My feeling is that this is the kind of remark one may recognize as expressing a profound truth once one has oneself grasped the point at issue, but which will not provide much help to someone seeking to understand the relevant truth in the first place. T&R is, in effect, my own attempt to bring into sharper focus the dark but suggestive idea that self-consciousness must play a "formal" role in all thinking.

https://doi.org/10.1017/can.2025.10046 Published online by Cambridge University Press

In general, I think Kant frequently needs, but does not himself give a clear account of, a distinction between a kind of representing in which some structure figures as the *form* through which other material is presented and another kind of representing in which this structure itself becomes a *topic* of representation in its own right. I will mention a few examples of this problematic in Kant, which I hope will be recognizable to aficionados, although I fear they will mean little to other readers. But speaking to the aficionados: Think of (1) the relationship between space considered as a "form of intuition" though which "outer appearances" are presented and space as constituting a "formal intuition" that itself presents a special kind of object for understanding, an object whose properties are explored in pure geometry; or again (2) the relationship between the "logical functions of the understanding in judgment" that come into focus when we "abstract from all content of a judgment in general, and attend only to the mere form of the understanding in judging" (Kant, 1998, A79/B95) and the corresponding "categories," which are said to be "pure concepts of the understanding" corresponding to these functions (Kant, 1998, A79-80/B105-6); or finally, (3) the relationship between the "synthetic unity of apperception" and the "analytic unity of apperception" (about which more in a moment). In each of these cases, I believe, Kant introduces a pair of concepts that stand to one another as *form-of-other-representing* to *form-as-topic-in-its-own-right*. Kant does not, however, theorize this special kind of shift: the shift of a representational structure from form to (distinctively formal) content. In Sartre's views on reflection, I find a more explicit account of this transition, and his account was the inspiration for my own.

I believe Kant's failure to theorize this reflective transition accounts for the ambiguous character of his views on apperceptive self-consciousness—the fact that they are read by some commentators as an account of a special kind of reflective awareness and by others (like Brink) as an account of an awareness operative even in unreflective thinking. I myself would suggest that they are both, but that because Kant did not clarify the difference and the relationship between these two topics, one can be left with the impression that one must choose sides. So what I find unsatisfactory in Kant is not that he positively embraces a reflective conception of self-consciousness to the exclusion of pre-reflective conception, but that he does not give us the resources to understand how these two topics differ and how they are related. So while I am grateful to Brink for her lucid presentation of the pre-reflective dimension of Kantian apperception, I remain convinced that this must be only one dimension of his topic, and that comprehension of the topic as a whole requires an inquiry that Kant did not himself undertake.

I have a similar attitude toward Kant's idea that apperception expresses a consciousness of unity in all thinking. This Kantian idea seems to me like a glorious antique garment that I know was once worn by people like me, but the practice of suiting up in which has been lost to time. I certainly would like to wear it, but how exactly would I put it on? I mean: How would I learn, not just to recognize this idea as something Kant thought, but to inhabit it as my own idea? How would I motivate for myself the idea that any act of thinking presupposes a consciousness of unity in all thinking? My reflectivist response to Hume's challenge seeks to motivate just this idea, and in this way, it can be seen as an attempt to reclaim a dark but appealing Kantian thought.

Kant famously distinguished two aspect of unity in apperceptive self-consciousness: the "analytic unity of apperception" that I recognize when I accompany each of my several representations with the same *I think*, and the "synthetic unity of apperception" that is achieved when I combine a manifold of given representations into a single, unified consciousness. He went on to argue that the former unity presupposes the latter:

> [I]t is only because I can combine a manifold of given representations *in one consciousness* that it is possible for me to represent *the identity of the consciousness in these representations* itself, i.e., the *analytic* unity of apperception is only possible under the presupposition of some *synthetic* one. (Kant, 1998, B133)

It would take more space than I have here to offer a systematic interpretation of these claims, but let me tentatively suggest that my response to Hume's challenge bears a striking resemblance to this

https://doi.org/10.1017/can.2025.10046 Published online by Cambridge University Press

Kantian line of thought. For I have argued that our ground for ascribing our various thoughts and experiences (analytically) to a single subject, "I," rests on the fact that these thoughts and experiences themselves stand together (synthetically) in a structure that presupposes a single subject.

Brink is right that I have exhibited this point only for certain special domains of representation (perceptual representations of things located in egocentric space, doxastic representations of propositional questions as open or closed), and she is right that it remains to be shown that the subjective standpoints presupposed in these different domains are all in fact the very same subjective standpoint, such that each of us can ascribe all of her representations to the very same "I." But I take this to be, not a defect in my approach, but simply an indication of a project which remains to be completed. As Brink notes, this incompleteness is something I emphasize in T&R: I seek to bring out a distinctive *strategy* for responding to Hume's challenge, but I pursue this strategy only far enough to show what it would be to have a purely reflective justification for ascribing a manifold of representations to a single subject, not to the point of offering a complete justification of the relevant kind (cf. Boyle, 2024a, 130–131). One of the merits of my reflectivist approach to self-awareness, I believe, is that it clarifies how we might approach this issue, and thus leaves us with a definite agenda of problems to be resolved and a definite method for resolving them. The task would be to reflect on the several ways in which any determinate thought presupposes a subject who thinks and on how these several presuppositions are related.

To see how this might work, consider a person who perceives a cat sitting in front of her and thinks an ordinary thought such as

(5) This cat is gray.

This thought presupposes the subject's own subjective standpoint in several ways. First, as I argued above in Section 1, the significance of her "this" depends on her present perceptual attention to a particular cat. But the relation of (5) to the subject's own standpoint goes deeper than this. Any perceived *this* is perceived as located somewhere in egocentric space, as near or far, up here or down there, and so forth. In T&R, I argue that such egocentric spatial representations presuppose a subject to whose location they are implicitly relative and in relation to whose powers their significance is implicitly defined. Moreover, I go on to argue that even judging that something is thus-and-so involves representing a propositional question as closed in a way that presupposes a subject who can know some but not all propositions that are true. So if I am right, thinking a thought such as (5) presupposes a located, perceiving, knowing, and potentially active subject, and it presupposes these several dimensions of subjectivity, not as independent postulates, but as linked in one act of holding true.

Pursuing this thread would, I believe, enable us to clarify why we are right to ascribe our beliefs, perceptions, desires, and so forth to a single subject, and in the process we might learn a great deal about the ways in which our capacities for perceiving, knowing, and acting are interconnected. So although I think Brink's question about why there are not "as many subjects as we have modes of presentation" (Brink, 2025, 215) is a good one, I do not think that answering it requires appealing to some independent Kantian thought about our necessary consciousness of the unity of thinking. Rather, our inquiry sheds light on what might be meant by this wonderful but elusive Kantian thought.

Matthew Boyle is Emerson and Grace Wineland Pugh Professor of Philosophy at the University of Chicago. He works on topics in the philosophy of mind and moral psychology, especially self-knowledge and the nature of rationality, and on topics in the history of philosophy, especially Immanuel Kant. He is presently writing a book on other minds and intersubjectivity.

References

Barnett, D. (2025). Boyle's elusive egoism. *Canadian Journal of Philosophy*, 55(3), 235–246.

Bar-On, D. (2004). *Speaking my mind: Expression and self-knowledge*. Oxford University Press.

Bar-On, D. (2025). Transparency to the world, the nature of mind, and self-knowledge: Commentary on Boyle's Transparency and Reflection. *Canadian Journal of Philosophy*, 55(3), 220–234.

https://doi.org/10.1017/can.2025.10046 Published online by Cambridge University Press

Bar-On, D., & Wright, C. (2023). *Expression and self-knowledge*. Wiley-Blackwell.
Boyle, M. (2012). Essentially rational animals. In G. Abel & J. Conant (Eds.), *Rethinking epistemology* (Vol. 2, pp. 395–428). De Gruyter.
Boyle, M. (2016). Additive theories of rationality: A critique. *European Journal of Philosophy*, 24(3), 527–555.
Boyle, M. (2017). A different kind of mind? In K. Andrews & J. Beck (Eds.), *The Routledge handbook of the philosophy of animal minds*. Routledge.
Boyle, M. (2023). Self-consciousness, transparency, and reflection. *Philosophisches Jahrbuch*, 130(2), 110–129.
Boyle, M. (2024a). *Transparency and reflection: A study of self-knowledge and the nature of mind*. Oxford University Press.
Boyle, M. (2024b). Anger, intentionality, and the view from within: Sartrean reflections. In B. Marusic & M. Schroeder (Eds.), *Analytic existentialism*. Oxford University Press.
Brink, C. (2025). Nonpositional consciousness and the form of thinking: Kant and Boyle on self-consciousness. *Canadian Journal of Philosophy*, 55(3), 211–219.
Byrne, A. (2018). *Transparency and self-knowledge*. Oxford University Press.
Dretske, F. (2003). How do you know you are not a Zombie? In B. Gertler (Ed.), *Privileged access: Philosophical accounts of self-knowledge*. Ashgate.
Gomes, A. (2024). *The practical self*. Oxford University Press.
Hume, D. (1979). *A treatise of human nature* (2nd ed., Eds. L. A. Selby-Bigge & P. H. Nidditch). Clarendon Press.
Kant, I. (1998). *Critique of pure reason* (Trans. P. Guyer & A. Wood). Cambridge University Press.
Roessler, J. (2015). Thinking, inner speech, and self-awareness. *Review of Philosophy and Psychology*, 7(3), 541–557.

Cite this article: Boyle, M. 2025. Replies to Bar-On, Barnett, and Brink. *Canadian Journal of Philosophy* 55: 247–260, doi:10.1017/can.2025.10046

https://doi.org/10.1017/can.2025.10046 Published online by Cambridge University Press