

Critical Notice of Richard Moran, *Authority and Estrangement: An Essay on Self-Knowledge**

Sebastian Gardner

There is a way of understanding “the philosophical problem of the self” that makes it unquestionably central to and ineliminable from the entire endeavor of philosophy, ancient as much as modern: if by asking about the self we mean to ask what we really, fundamentally *are*, then the problem of the self comprehends a great range of problems, and every philosophical position worth the name will include a theory of the self. In analytic philosophy, however, the problem of the self has come to be understood differently: what it refers to in the first instance is the puzzle presented by knowledge claims regarding one’s own psychological states, a problem that is distinguished from, while enjoying some connections with, the problem of the conditions for personal identity and the mind-body problem, and that resolves itself into a set of finer problems concerning the conditions of psychological self-ascription, the reference of the first-person pronoun, the notion of immunity to error through misidentification, and so forth.

In a historical perspective, it is as if the self has ceased to be a lynchpin and become a philosophical epiphenomenon: the self is now regarded as presenting a problem whose solution depends substantially on what positions are taken in other areas of philosophy, and that does not play an important or extensive positive role in elucidating issues outside the philosophy of mind. The key steps through which this shrinking of the concept and problem has come about include of course the linguistic turn and Wittgenstein’s influential contribution, and the move away from Cartesian foundations in epistemology. But the decisive factor, it may be ventured, is acceptance, motivated from several quarters, of what may be called the Lockean or Strawsonian picture, according to which the methodologically and ontologically primary item is not the “I” of self-consciousness but the living, concrete, empirically determinate *person*, to which the self is conceived as annexed. While it is not denied, on the Lockean picture, that incorporation of an “I” is essential to personhood, it is not thought that the “I”

*Princeton: Princeton University Press, 2001.

as such contains an explanatory ground of personhood or of its philosophically important dimensions. Problems in, for example, moral psychology are thus not expected to receive elucidation through consideration of the first-person pronoun. In this way the problem of the “I” is dissociated from the grand question of what we really and fundamentally are, and a work like Charles Taylor’s *Sources of the Self* comes to seem, though harmlessly so, inaccurately titled: what Taylor deals with in that work is not strictly speaking sources of the self but of personhood.

Richard Moran’s endeavor, in this outstanding, stimulating, and original book, is to address the analytically circumscribed problem of the self in a way that undoes its isolation, restoring to its solution a connection both with the question of what we really and fundamentally are and with moral psychology, by reclaiming the thought that there is something “distinctive or irreducible about the perspective of the first-person,” that “the person’s own access and relation to his own mental life must be different in its possibilities and limitations from anyone else’s” (xxviii). In amplifying this idea, Moran breaks with some parts of the Lockean-Strawsonian picture.

Authority and Estrangement takes off from a familiar point, with consideration of the immediacy, non-inferentiality, and epistemic authority of claims to self-knowledge. In chapter 1, Moran reviews the asymmetries between self-knowledge and knowledge of others, and develops an account of self-knowledge that is independent from the perceptual model of self-knowledge, which Moran criticizes as defective on several fronts and in all its forms. Moran’s account retains the immediacy affirmed in Descartes’s picture while rejecting Descartes’s (widely supposed) commitment to its infallibility. Moran calls his epistemology “substantialist”—self-knowledge represents a “genuine cognitive achievement” (3) and “involves the awareness of some independently obtaining state of affairs” (17), “truth-conditions [that are] in some way independent of the making of the judgment” (18)—in contrast with the “deflationist” view that avowals of self-knowledge are not “expressive of first-person *judgments*” (13). On the question of whether first-person authority has an a priori basis, a conceptual and perhaps normative ground, Moran restricts himself to claiming merely that, if it does, then it is not at the cost of making self-knowledge something less than genuine cognition, that is, of deflationism or conventionalism.

The further chapters will be considered in some detail later, so here I will just state their content briefly. Chapter 2 pursues the argument between realism and idealism in the context of the objects of self-knowledge. This discussion leads to the central contention of Moran's book, developed with depth and brilliance in chapters 3 and 4, concerning the primacy of the practical in self-knowledge: "It is modelling self-consciousness on the theoretical awareness of objects that obscures the specifically first-person character of the phenomenon. ... A more complete characterization of the first-person perspective will require bringing the agent more explicitly into the picture" (32–33). Chapter 3 does the main work in making good on this claim: beginning with an analysis of Moore's paradox, Moran contrasts the theoretical stance with that of deliberation and commitment, and explains the notion of avowal, as distinct from self-attribution, in terms of the latter. Chapter 4 maps the ways in which the idea of first-person authority matters, showing how it is a condition of having mental states at all (108), how it is related to immediacy and to the priority of justificatory over explanatory reasons, and how the conditions of self-knowledge are connected with psychological health. Chapter 5 returns the discussion to the first- and third-person asymmetry that provides Moran's starting point, and shows how the preceding analysis allows us to grasp more clearly and deeply what it is to be a self, and how this corrects some of our ideas about the basic shape of morality. For it is only, Moran contends, when the first person is properly in focus that we know what it is to see oneself as "one person among others".

Reflecting the dissociation of analytic philosophy's conception of the problem of the self from that of earlier modern philosophy, Moran concentrates on mapping his relation to contemporary authors and dispenses with a historical framework. Passages from Wittgenstein are discussed, but in abstraction from what may be considered Wittgenstein's general philosophical intentions. The one figure outside the anglophone tradition with whose writings Moran does at several points compare his own account of self-knowledge is Sartre: Moran regards his own view of self-knowledge as, despite some important differences, fundamentally in tune with that which he discovers in *Being and Nothingness*.

Moran's study may be described as epistemology conjoined with metaphysics of a descriptive variety, but there are several familiar senses in which it counts as conceptual inquiry rather than metaphysics. First, the considerations that Moran appeals to in arguing for his

view are commonsense psychological and phenomenological in the nontechnical sense of concerning the appearances of mental life in the terms we ordinarily and most readily apply to them (Moran makes some use of psychoanalytic ideas, but he belongs to the school that regards psychoanalysis as an extension of commonsense psychology). Second, Moran does not venture any conceptual innovations or revisions, and invents no new philosophical concepts: he regards his conclusions as adequately expressible in the terms of our ordinary commonsense psychology. Third, there are certain questions about the self that Moran does not take up, or, it may be better to say, there are certain ways of formulating questions about the self that Moran does not employ: What is the self? What is self-consciousness? How is self-consciousness possible? Moran's attitude to questions like these, which ask after essences or for ultimate grounds and grounding explanations, is complex and will be considered below.

Moran's philosophical picture of the self is intended to match the self's own picture of itself. Moran regards the connection between his topic, self-knowledge, and his methodology, adherence to the framework of commonsense psychology, as *internal* (34–35). He accepts a thesis that one might call “internality of explanation with respect to the self,” according to which self-knowledge *ought* to be accounted for in the terms of commonsense psychology precisely *because* the self understands itself in those terms, and the self is something that ought to be explained in the same terms as those in which it understands itself.

What I will do now is discuss in turn four themes that figure in Moran's study: (1) the bearing of sub-personal psychology on understanding of the first person, (2) the issue of realism or idealism regarding the objects of self-knowledge, (3) Moran's conception of the practical-deliberative, and (4) the use made by Moran of Sartre. In each case I have some comments to make on the scope of Moran's treatment of the topic. These four themes do not by any means exhaust the content of the book, and my discussion of them will convey all too little of the subtlety of Moran's analyses. The reason for my selection will emerge at the end, when I will state the overall view of Moran's achievement that, I will suggest, should be extracted from my discussion.

1. *Sub-personal psychology*. In order to show how his project sits on a larger canvas of contemporary philosophical concerns, Moran situates the account that he gives of the self against two opposing positions: the

Cartesian conception of the self, and the sub-personalist's view of the self.

While Moran's distance from Cartesianism is straightforward, his relation to empirical psychology has complications. Moran regards (6ff.) his central theses that self-knowledge is asymmetric with knowledge of others and that it contains a node of essential interconnection with "ordinary rationality and personhood" (8) as contradicting the claim that the mind is in reality a configuration of sub-personal structures from which the self is absent. In illustration of this claim Moran cites Dennett, whose pronouncement that the "new" mind revealed by cognitive science shows the Self to be nothing but a species of fiction, an idealization or abstraction of some heuristic or practical value but lacking empirical reality and thus reality *tout court*, is well known.

At one level, however, it is not obvious that there is any inconsistency between Moran's theses and even a strong philosophical commitment to sub-personal psychology. Suppose Moran's constitutive thesis regarding self-consciousness and rationality or personhood is true. It follows, the sub-personalist will say, that there is a strongly unified sector of psychological life that falls under the rubric of self-knowledge and thereby provides prima facie evidence for the existence of a corresponding configuration at the sub-personal level. The existence of a distinctive personal-level sphere does not suggest immediately any philosophical limitations on sub-personal analysis; it shows only that there is no necessity to pursue the topic of self-knowledge at the sub-personal level. To think otherwise would be to conflate Dennett's specific, eliminativist-fictionalist sub-personalist thesis with the logical implications of sub-personalism *per se*. Moran is probably right to object that Dennett's claim goes beyond what we have reason to believe, but if so, the fault lies with Dennett for his extravagant inference, not with sub-personalism.

Moran's underlying worry is that sub-personal theory may be claimed to undercut, overtake, or supplant the sort of account that he wishes to offer, and he is of course right that there is no guarantee of the happy coexistence of a personal-level account of self-knowledge with sub-personal analysis: if self-knowledge is open to sub-personal analysis, then it is possible that its results will not confirm, or will contradict, the personal-level view, and in such a case the sub-personalist will resort to Dennett's eliminativism-cum-fictionalism. The inconsistency of Moran's view with sub-personalism thus lies in their respective modal claims: Moran wishes to deny, and the sub-personalist to affirm,

the *possibility* of the non-reality of the Self; their difference lies, on one conception of supervenience, between denial and affirmation that the first-personal sphere supervenes on sub-personal configurations.

If this is the nature of his disagreement with sub-personalism, Moran has the following options: (1) to give grounds for repudiating the very idea of sub-personal analysis of the mental in general; (2) to argue, not on general terms but just with reference to the first person, that, should sub-personal analysis contradict personal-level understanding, the latter must be deemed to survive unscathed; or (3) to leave open the possibility that the first-person sphere will be eliminated. Despite some suggestion that the normativity built into the first person blocks sub-personalization, Moran does not play the general anti-sub-personalist card (7 n. 5), so option (1) is not taken. Nor does he say anything to support (2) directly: though it may appear somewhat as if Moran believes that the degree of internal coherence of the first person gives proportional reason to affirm its reality, such an argument would not be to the point, and Moran does not make it. That leaves Moran with (3), the “fingers crossed” option. Does this matter? It means giving up entitlement to claim the autonomous reality of the first-person sphere, its nonsupervenience on the sub-personal, but beyond that Moran’s project is not directly affected: there is nothing to be done about the future development of cognitive science, so all that Moran can do is what he aims to do anyway, which is to exhibit to a maximal degree the distinctiveness, coherence, integrity and importance of first-personal understanding. A problem will arise only if incoherences in the first-person sphere come to light, for in such circumstances the sub-personalist, whose explanatory project, it should be remembered, is motivated and underpinned by a positive view of what we fundamentally and really are—namely, natural organisms—will be provided with the opportunity to cut the Gordian knot by declaring that the philosophical puzzles surrounding the “I” have no more cogency than the puzzles surrounding qualia, which, Dennett has argued eloquently, are best regarded as reflecting the unreality of the pseudo-objects that give rise to them.

There are indications, however, that Moran would not be satisfied with option (3) and wants to hold out for a stronger anti-sub-personalist conclusion (at the very end of the book, for instance, he says that it “as a whole has argued that the concept of the person as a reasoning agent is ... ineliminable” (193)). And this is understandable, since Moran agrees that self-knowledge does exhibit an apparently problematic character

and that philosophical labor is demanded to dispel its mystery, and so acknowledges the self's vulnerability. First, there is a strong suggestion that Moran regards the argument with sub-personalism as engaging a proto-axiological, or normative issue. While discussing the implications of giving up on first-person authority (121–24), Moran says that the “supersession of this assumption, perhaps made possible by some future science of the mind” (121), though not entailing any explanatory loss, would shift our picture radically, because “the primary thought gaining expression in the idea of ‘first-person authority’ [is] that it is *his business* what he thinks about something, that it is up to him” (123). Moran’s view may be, therefore—although the idea is not developed explicitly—that the *Faktum* revealed by awareness of self-responsibility expresses an interest of reason sufficiently authoritative to deflect sub-personalism.

Second, and connectedly, Moran seems to put a strong construal on the internality of explanation thesis: he considers, as I read him, not just that there ought to be some internal explanation of the first person, but that it ought *not* to have any external explanation. Again, Moran does not say this much explicitly, but commitment to the internality of explanation thesis in this strong form would provide a natural way of grounding his opposition to sub-personal explanations and elucidate his sense of their negative import (as it is in evidence on, for example, 6–7 and 85–87). However, Moran does not provide general support for the internality of explanation thesis strongly construed.

If Moran’s engagement with sub-personal psychology is limited in the respects indicated, then his account of the self cannot be claimed to have opened up or consolidated a general philosophical position of the same sort as that of the Cartesian and sub-personalist. This is to be expected, for in order to match sub-personalism, it would be necessary for Moran to engage in grounding explanations, specifically, to show that the self has—at least, that it can be given, if not that it must be given—a grounding explanation that precludes naturalistic sub-personalist possibilities.

2. *Realism vs. idealism.* In chapter 2, “Making up Your Mind,” Moran discusses the rival claims of realism and idealism regarding the objects of self-knowledge. The question is this: To what extent, if any, is either the existence or the character of my psychological states, as known by me, dependent upon my awareness (conceptualization, interpretation) of them? Moran, concentrating on the case of emotion, acknowledges that here our intuitions to some degree go two ways. But after

teasing out the main factors favoring a realist and an idealist answer to the question, he comes down on the side of realism: we need not and should not abandon “commonsense,” “everyday,” “ordinary realism about the mental” (37, 59).

As Moran configures the debate (39–40), there is an initial presumption in favor of realism: he points out that, after all, we believe that there is a fact of the matter regarding one another’s psychology, we regard self-knowledge as requiring effort (thus supposing there to be some fact the would-be self-knower is straining to attain), and there are at least some mental (emotional) states whose reality seems to be fixed quite independently of how its owner thinks of it. Thus the onus lies, per Moran, with the idealist view, of which Moran takes Charles Taylor as a representative exponent. On Moran’s account, while Taylor’s hermeneutical claims do pick out many genuine features and peculiarities of self-understanding, these can be regarded in realist terms once the practical-deliberative dimension of self-knowledge is put in place (59), and they can be taken to establish idealism only if they are not examined closely: Moran thinks that Taylor as it were agglomerates several different respects in which, in some areas of mental life, the identity of the state can seem to be determined by how we take or interpret it, while ignoring the realist background that makes this limited sphere of self-interpretative latitude, and the dynamic-transformative role of reflection on one’s mental states, possible in the first place.

Moran’s discussion here, while drawing attention to some limitations of Taylor’s articulation of his views, arguably does not work through the whole of the issue, and this may be in part because of Moran’s starting out with the special case of emotion. On the one hand, it can seem that emotions provide the right territory for pursuing the argument between idealism and realism to a definite conclusion: since emotion appears to grant the most scope for making the case for idealism, it would seem that if idealism fails in this context, then it can be inferred that it fails in other psychological contexts. However, the fact that emotions present special complications also distracts from the more fundamental difficulty surrounding the realist construal of first-person knowledge, allowing it to be thought that, so long as those complications can be shown not to entail idealism, realism is vindicated. The more fundamental difficulty is whether it is truly conceivable that I should regard my psychology as a realm of facts, ones with respect to which there are some special possibilities for me, but that are fundamentally of a kind with worldly facts at large. If self-

knowledge is construed in a plain and unqualifiedly realistic fashion, then the subject is obliged to conceive of the world as containing a set of facts concerning herself that are of the same order as any other set of facts, and that are distinguished for her not *in themselves* but solely in virtue of how she stands *in relation* to them. This requires the subject to operate with a uniform conception of “fact,” and to discount the seeming specialness of the facts that compose her—their seeming to be as it were recessed or set apart from the world—as a function of her “take” or “angle” on them.

Idealism, as a general philosophical position, can account for this situation in terms of its being a requirement on objective facticity, on the formation or application of the very concept of an objective fact, that it should stand in contrast with a subjective field: the objects of self-knowledge cannot be homogenized with the facts that compose objective reality at large, the idealist will say, because the latter are dependent (in a sense that goes beyond mere temporal epistemic order) on the former. From the same idealist quarters, the objection can also be put that a conception of fact that allows itself to be stretched in the way required by the realist cannot be taken for granted: if “real” states of affairs allow of modes of presentation and of epistemic access as remarkable as those exhibited in the first person, then no ordinary, humdrum, common-or-garden conception of reality is in play, and the realist is under pressure to first explain what this conception is and whence derives his entitlement to it.

Holding aside counterarguments deriving from a general idealism, it is true that there is no contradiction in the realist supposition that the seeming specialness of the objects of self-knowledge is a perspectival function, residing solely in the subject’s relation to them. But to say this raises the question, What *is* the subject’s own, positive conception of this relation? In order to adhere to the internality of explanation thesis, the conception must be one such as does or could appear on the inside of the subject, not just a conception fashioned by someone occupying the sideways-on view: it needs to be a way of thinking that is or can be exercised by the subject in her relating herself to her mental states. What then *is* this relation? Viewed one way, it seems that I have no positive conception of it: I cannot tell a story of how I come to stand in relation to my mental life; the only account I can give of the relation is exhausted in my ability to reflectively relate myself to my states. If this is so, then I cannot discount—at least, I cannot do so from the inside—the “seeming of specialness” of the objects of self-knowledge in

the way required by the realist: I cannot subtract my “take” on the facts that compose me from the facts themselves, if I do not have a conception of what it is that I am supposed to subtract.¹

That there is a deep difficulty with the idea that the seeming specialness of the objects of self-knowledge is contained in a relation to them, rather than residing in the objects themselves, is in fact a line of thought pursued by Moran himself, in criticism of D. H. Mellor’s account of conscious belief (28–33). But Moran does not regard the point as unsettling his ordinary realism: his claim, if I read him correctly, is that the practical-deliberative dimension of self-knowledge solves the difficulty without impugning realism. Whether it can play this role, I argue below, is doubtful.

That said, even if Moran’s notion of the practical-deliberative does not free his realism from difficulty, it certainly does not follow that idealism of any old sort gives the correct account of the matter at hand. On the contrary, and in conformity with the considerations cited by Moran as creating a presumption in favor of realism, the difficulty just indicated for realism reappears for idealism, under the construal that Moran gives it and supposes is to be found in Taylor: a conception of self-knowing along the lines of “I constitute myself thus,” “I make it so that I am thus”—a hermeneutically creative conception of my relation to the objects of self-knowledge, whereby interpretation is logically sufficient for one’s states (43), so that the objects of self-knowledge come to resemble products or instances of agency in the ordinary sense—clearly cannot play the requisite internal role. Idealism, so construed, is merely the obverse of realism, and accordingly mirrors its difficulty.

With regard to his criticism of Taylor, what Moran has arguably not factored in is Taylor’s intention to address in his idealist account of the objects of self-knowledge precisely the “transcendental” question that Moran puts into the margin by committing himself to “ordinary realism” in the domain of self-knowledge—the question of how there comes to be in the first place such a thing as persons, of how beings of our deeply peculiar sort can ever arise. Taylor may, therefore, agree with Moran that discussion of the relevant features of emotion does not, taken in isolation, warrant an inference to idealism, but add that his intention was instead to assign to his discussion of emotion a subordinate place in his attempt to construct an answer to a question that does not figure on Moran’s agenda.

The conclusion thus far would appear to be that Moran’s discussion has not settled the issue in favor of realism, and that our two sets of intuitions

remain unreconciled. The sub-personalist naturalist may propose dissolving the conflict of intuitions by treating the sphere of self-understanding, as Dennett proposes, on an analogy with a fictional world. A diametrically opposed suggestion is to be found in Sartre. The hope for Moran, who does not take over Sartre's metaphysics, must be that his conception of the practical-deliberative dimension will be able to accommodate the idealist intuition, that is, explain the seeming specialness of the objects of self-knowledge in a way that is compatible with ordinary realism about them.

3. *The practical-deliberative.* Moran's concept of the practical-deliberative dimension of self-awareness is supposed to account for the features of self-knowledge neglected or misunderstood by other theories of self-knowledge.

Moran's thesis that the basic form of self-knowledge is practical (stated at many points in the book (e.g., 150–51)) means that our having knowledge of ourselves consists in a practically or deliberatively oriented taking-up of our psychological states: consciousness of our attitudes is constituted by our interest in believing what is true, desiring what is desirable, etc., and our capacity for deliberating about those matters. Fundamentally, our relation to our psychological states is practical in the sense of *doing* something to our states, where this "doing" is conceived dispositionally or capacitatively: our "capacity for avowal" constitutes the fundamental relation of self-knowledge. In order for this capacity to render the relation practical in a significant sense, avowal must mean more than "make it be known," and Moran uses a variety of expressions to evoke the non-epistemological, quasi-conative dimension that he wants to fix: "endorse," "commit to," "speak for," "identify with," "acknowledge," "take responsibility for," "assume authority for," "see as 'up to me'" (xx). These are the marks of avowal as distinct from mere self-attribution. The reflexive stance or posture that Moran has in mind is therefore more fundamental than the sorts of reflexive practical attitudes exhibited more conspicuously in "second-order" psychological states such as resistance to one's evil impulses.

The index of being related to one's psychological states in the mode of endorsement or commitment is satisfaction of what Moran calls the "Transparency Condition," which is the condition that a piece of self-knowledge meets when its avowal—the affirmative judgment that *p* is as a matter of fact one's psychological state—is in tune with what is claimed by *p*, with what *p* says about the world or with what is needed

in order for *p* to be rational and justified. Failure to meet this condition is a mark of self-alienation of some sort and to some degree. Thus in the case of belief, Moore's paradox is an index of failure to meet the Transparency Condition. Transparency is, on Moran's account, not a "logical requirement" but an "achievement" due to the practical-deliberative stance (6–7).

As an analysis of self-knowledge competing with others on offer, Moran's account is extremely impressive, and its strengths are detailed convincingly in the book. It raises however a question connected with the issues discussed earlier.

Moran holds that it is the "practical point of view" on ourselves that secures our rootedness in ourselves and makes that relation asymmetric with our relation to others. One way that he goes about trying to show this is to invite us to conceive a subject that relates to its psychological life in a purely theoretical, cognitive mode (90ff). Moran argues persuasively that such a case, however complete and accurate the self-knowledge achieved by such a subject, would be one of "mind-reading" one's own mind, "self-telepathy," and not an instance of the "ordinary relation of self-knowledge"; the subject would be locked into the self-alienation evinced in Moore's paradox. Thus, Moran infers, it is the practical dimension that grounds our rootedness in ourselves.

Now this makes it look rather as if Moran is analyzing self-knowledge as a product of independent factors—as if he is inferring, from our discovery that a purely theoretical self-relation is insufficient for the ordinary relation of self-knowledge, that this relation must be practical; as if the idea were that it is because this psychological series is the one that I can *do something about* (or *with*, or *to*) and so must take responsibility for, that it is genuinely *mine*, that I am not distant from it in the manner of a mind-reader.

The notion that the bridge between the self as (knowing) "subject" and the self as (known) "object" consists of a practical tie is appealing, but it is important to recognize the explanatory limitation of the idea. In the first place, Moran's thought-experiment does not allow it to be inferred that the practical self-relation accounts for our rootedness in ourselves. The fact that if the practical self-relation is subtracted we no longer have self-knowledge does not mean that the practical self-relation is the source, the *explanation*, of what distinguishes self-knowledge from knowledge of others. For it may also be inferred from the fact that self-knowledge is destroyed by subtraction of the practical self-relation that the practical self-relation is a necessary *consequence* of whatever it is

that makes self-knowledge what it is: there may be something as yet unidentified that is the ground of *both* the theoretical and practical self-relations and that necessitates both. Also, and importantly, a circularity looms. The practical self-relation cannot coherently be *added* in thought, as supervening, to a pre-existing merely theoretical self-relation, since the subject of the practical self-relation (the “I” that endorses, commits itself, identifies) must identify the stretch of psychological life that it is its job to take responsibility for *as its own*, must already conceive it as a unity to which it relates *as itself*, and so must *already* be presented with the objects of self-knowledge in just the way that the practical self-relation was intended to elucidate.

A limit is thus set to the practical as an explanation of self-knowledge. Moran can claim that the practical self-relation is necessary for self-knowledge, and that it is a symptom or expression of whatever it is that explains self-knowledge, but not that it is what grounds our rootedness in ourselves, that avowal “is what makes the difference” between self-awareness and awareness of others (107).

Insofar as Moran’s project is not directed at uncovering explanatory grounds but rather at mapping conceptual connections, this is no point against him, but it has an important implication for an earlier issue: it means that the practical self-relation cannot be claimed to account for the seeming specialness of the objects of self-knowledge, and so cannot be employed to free Moran’s ordinary realism from the difficulty described earlier. There is, moreover, a respect in which the practical self-relation seems to highlight the difficulty encountered by Moran’s realism. If, as realism implies, the objects of self-knowledge have a real existence that extends beyond their being avowed and identified with, how do they come to present themselves as practically amenable, as occasions for deliberation, as surrounded with practical possibilities? How, once I have grasped something as a fact about me, can I go on to grasp it as “practicable” in the special way that Moran quite rightly says distinguishes the objects of self-knowledge? In the case of things composing the not-self, there is no conceptual difficulty in seeing how they can be the objects of both theoretical and practical attitudes—I see this is a hammer and I mean to pick it up—because a subject to whom we ascribe the capacity to unify the two perspectives is presupposed, but the existence of an entity with this capacity cannot be accounted for by modeling the subject’s relation to inner objects on its relation to outer objects. There must be, it seems, something within or about the facts composing the objects of self-knowledge that makes

possible and elicits their being taken up practically (avowed, endorsed, committed-to, etc.), but whatever this might be, its compatibility with ordinary realism remains undemonstrated and in doubt. (The question of how something presented as a fact becomes practicable can, of course, also be put in reverse: how does an object intended in practical self-determining become a fact to which ordinary realism is appropriate?)

Moran might simply regard the transcendental question that has surfaced here, concerning the intrasubjective unity of the practical and the theoretical moments of self-consciousness, as falling outside the bounds of his inquiry, albeit at the cost of leaving it unclear how a practical relation to the objects of self-knowledge can be squared with realism. Yet Moran does appear to take up the challenge of explaining the practical-theoretical unity: first, by disputing George Rey's sub-personalist explanation of how the theoretical and practical dimensions slot together (85ff.), and second, by adverting to the empirical/transcendental distinction in Kant and Sartre in order to gloss his dual perspectives (e.g., 89). Even if Moran is right that Rey's account is inadequate (though Moran's argument against Rey seems to me again to underestimate the possibilities of sub-personal explanation), this still leaves him without a positive account of his own. The comparison with Kant and Sartre seems to break down quickly. The transcendental point of view in Kant is not a point of view occupied by empirical subjects in their dealings with empirical objects, but a strictly philosophical view, designed to explain what makes it possible for us to inhabit our empirical point of view, while the dual-perspective element in Kant's system that does to some degree parallel Moran, the theory of freedom, for obvious reasons cannot be the model that Moran is intending to follow. The element in Sartre's system that again has some parallel with Moran's doubling of perspectives, Sartre's facticity/transcendence distinction, implies however, I argue below, a view of self-knowledge that conflicts with Moran's realism, and part of the reason for this is that Sartre's distinction—which should not be assimilated to either Kant's empirical/transcendental distinction or his theoretical/practical distinction—has an ontological sense and is not just a distinction of perspectives or “stances.”

4. *Sartre*. Moran's comparison of his account with Sartre's view that the sphere of self-consciousness is one of freedom and responsibility is obviously justified, but there is a key difference, which goes back to the realism *vs.* idealism issue.

Moran holds that the tension between realism and idealism regarding the objects of self-knowledge is a tension in *philosophical* reflection, from which it can be *eliminated* through an account that gives a proper place to the practical-deliberative dimension: the realism- and idealism-suggestive dimensions of self-knowledge are balanced out, the former weighing more heavily. Sartre, by contrast, holds that the tension is internal to natural consciousness, and that it must be *preserved*, in an appropriately elucidated form, in philosophical theory, precisely *in order* for that theory to uphold the self-responsibility of first-person understanding. The reason why Sartre's position is not a mere restatement of the idealist-plus-realist confusion that inhabits natural consciousness, but rather a philosophical explanation of it, is that his account of the objects of self-knowledge is metaphysical: on Sartre's account, the "idealist" and "realist" aspects of the objects of self-knowledge are written into their ontological status. Moran's difference from Sartre emerges sharply in a discussion on 79–80, where Moran treats as contingent—as peculiar to cases of self-alienation—a structure that Sartre regards as inherent in self-knowledge as such, and the same pattern is repeated in Moran's discussion of intersubjectivity in chapter 5: Sartre regards the general form of the self's relation with others as problematic in a way that Moran denies (Moran's view of intersubjectivity is, in Sartre's language, "ontological optimism").

It is notable that, although the metaphysics employed by Sartre do not support what Moran calls an "ordinary realism" about the mental, they are not contradicted by the considerations cited by Moran as creating a presumption in favor of ordinary realism, and nor is Sartre's position objectionable in the way that Moran supposes the idealism to which he opposes his own realism to be. First, while Moran is surely right to insist in his critique of the idealist theory that there is in self-knowledge a purposeful striving—self-knowing is conducted under an aegis, with an *end* in view, in relation to which the endeavor can fail or succeed—and that this end is not captured by talking of self-creation, it remains possible that the telos does not require the real existence of psychological facts: Moran's considerations preclude only what one might call an "ordinary idealism" about the mental, leaving it open that some other, "non-ordinary" form of idealism can articulate the telos of self-knowledge, and Sartre has a great deal to say on that subject.

Second, regarding Moran's (again, justified) insistence that an account of self-knowledge must allow the attempt to know oneself to "terminate in a true description" (63), Sartre agrees that there is "termination" in one

sense – a point can be reached where all the factual considerations have been counted in. The question, Sartre insists, while agreeing with Moran that “deflationism” is not the answer, is what *sort* of thing is determined at this point and whether the terminus should be conceived as knowledge of an “ordinarily real” fact. Whereas Moran answers in the affirmative, according to Sartre, to represent the terminus of self-knowing as an ordinarily real fact is to lose sight of its non-ordinary mode of being and so of the necessary condition of self-responsibility.

Let me now say where I think the discussion leads. As has been seen, there are a number of questions concerning the self raised by Moran’s account that are either not addressed by him or to which, if the demand for explanation is pressed far enough, his answer appears incomplete. These include the justification for affirming the impenetrability of the first-person sphere to sub-personalist naturalistic explanation or elimination, the explanation for the seeming specialness of the objects of self-knowledge, and the unity of the theoretical and the practical dimensions of self-consciousness. To the extent that Moran’s inquiry is circumscribed in such a way that these issues fall outside its bounds, this is no shortfalling, and I have taken pains to emphasize that the extended analysis that provides the core of *Authority and Estrangement* is unaffected by the points I have raised. However, Moran’s attitude to the transcendental questions is, we have also seen, tinged with ambiguity. On the one hand, as said at the outset, Moran can be viewed as pursuing conceptual inquiry in a sense that contrasts with metaphysical explanation, while on the other hand, engagement with metaphysical issues is implied by the various claims of his discussed above. Indeed, embedding of his analysis of self-knowledge in a bigger picture seems to be part of what Moran has intended to add in expanding his journal papers into a book.

The metaphysical issues must be faced eventually in any case, and what merits emphasis at this point, I suggest, is the strength of the case for supposing that, in order for them to be resolved in a way that coheres with Moran’s central theses concerning self-knowledge, speculative metaphysics of a sort that Moran avoids are needed.

This point begins to emerge from consideration of the difference between Moran and Sartre. The relevant question is not whether Sartre is right to suppose that an ontologically grounded libertarianism is necessary for human responsibility, but whether the way in which philo-

sophical consideration of the self provides a basic or *prima facie* warrant for the sort of approach adopted by Sartre. In favor of supposing that the answer is yes, the following can be said immediately. First, Sartre's introduction of different modes of being responds directly to the pressure exerted by the basic insight that I do not relate to the objects of self-knowledge as to worldly states of affairs: the reason why our relations to each sphere are so different, and why their epistemologies differ, is that they have the different modes of being theorized by Sartre. Second, the very same thesis allows Sartre to address the puzzle of Moran's conception of the practical-deliberative: the mode of being of the objects of self-knowledge precludes discrimination within them of two moments, one factual and one practical, so the problem of their "becoming practicable" does not arise. Third, and in one respect this is the most important point, Sartre's anti-naturalistic metaphysics achieves directly the aim of insulating the first-person sphere from sub-personal explanation.

To emphasize, the present question is not whether Sartre's metaphysics succeed or fail. Sartre is but one instance in a long tradition of post-Kantian speculation, and the point at issue is whether this tradition should be treated merely as an ancillary resource, from which ideas can be selected in piecemeal fashion to corroborate insights achieved through other, non-speculative philosophical means, or whether it should be regarded, as has been argued on many occasions by Dieter Henrich, as a unified philosophical program with enduring value that lies open to recovery and future development.² To take the latter view is to be presented with a wide range of possibilities, all of which reject, in a way more radical than Moran, the Lockean-Strawsonian picture, in favor of the idea that self-consciousness is metaphysically and methodologically foundational. They thus qualify, from the standpoint of ordinary realism, as varieties of idealism (though of a non-ordinary sort, which can claim to be also in some sense realistic). One way of stating the difference between this approach and Moran's is to say that, from the perspective of the former, the internality of explanation thesis has been applied by Moran only halfway, reflecting his continued, partial adherence to the Lockean-Strawsonian picture: Moran has given a view of self-knowledge that is internal with respect to *commonsense psychology*, but not with respect to *subjectivity* as such, and this is why the explanatory and grounding dimensions of his account are limited. In order to provide a view of the self that is fully internal

and comes closer to explanatory completeness, a philosophical step outside the bounds of commonsense psychology is needed.

From where many of those who find Moran's conclusions appealing stand, the attempt to reclaim philosophical interest for systems of this idealist sort is an attempt to unlearn decisive lessons from the history of philosophy, and to abandon the hard labor of elucidating and defending the reality of the personal sphere without resort to the quick fix of speculative metaphysics. What I have wanted to suggest is that, whatever else may be wrong with the post-Kantian project (and it is not, I think Moran would agree (xxxiv), that it is "Cartesian"), this judgment is not supported by consideration of the dimensions of self-consciousness that Moran brings to light. Viewed from inside the post-Kantian tradition, Moran's achievement, in attempting to develop his insight into the problem of the self within the analytic frame of reference, is to have pushed that framework, not to a point of collapse, but to a point from which the post-Kantian alternative returns to view.

One final comment, on a point connected with the issue of philosophical traditions. The implication, and in some ways also the assumption, of Moran's inquiry is that Transparency as such is a good, a necessary and proper goal, and that its absence, Alienation, is a deficiency or deprivation, to be overcome. Moran does not say anything as crude as this, but some such outlook orients much of his discussion: Transparency is connected directly and firmly to rational personhood, and when Moran gives examples of its failure, these are all cases that, everyone would agree, call for its restoration. Thus, in his (fascinating) treatment in chapter 5 of how the attempt to know oneself can miscarry, Moran explains the cases that he considers in terms of a false assimilation of self-knowledge to theoretical knowledge, or of the first person to the third. Moran does not see any inherent limitations to the project of self-avowal.

Now one might ask whether this is the whole story. One might think, first, that the demands of personal relationships, or of psychological flourishing, can provide not just psychological barriers to Transparency but reasons for setting limits to its scope. Second, there are philosophical and religious views from which again it would follow that Transparency is in certain domains an abortive or misguided endeavor. Should either prove to be so, then there are elements in and of the self with respect to which avowability in Moran's sense is not proper—objects of self-knowledge that I should not, of course, disavow in the sense of counting them into the not-self, but that equally I

should not strive to avow in the sense of “committing” myself to them. These objects I should regard not as “up to me” but as transcending the scope of what is properly subject to my deliberation.³ Obviously, and yet again, to raise this possibility is to make no criticism as such of Moran, who is of course quite right that with regard to the sectors of psychological life that he has in his sights, Transparency is what is wanted. What is indicated instead is that Moran’s outlook is not insulated from questions about the scope and meaning of rationality that the philosophy of mind on its own cannot settle, and that there are some conceptions of rationality that would require his account to be extended in a way that would allow “benign” or “rational” non-Transparency to be discriminated from its malign or irrational forms. As Moran recognizes when he aligns his view with the “tradition of Reflection” represented by Kant (138ff.), *Authority and Estrangement* is written in the spirit of Enlightenment. As such, it has its Counter-Enlightenment critics, some of whom can also claim Kantian ancestry. That a short book, on a topic that often seems to occupy only a small, technical corner in the philosophy of mind, should touch and illuminate so many nerve centers of philosophical concern, is cause for great admiration.

University College London

Notes

¹ For recent work on this topic, see David Carr’s exposition of Kant’s and Husserl’s views of the distinction of empirical from transcendental subjectivity, in chapters 2–3 of *The Paradox of Subjectivity: The Self in the Transcendental Tradition* (Oxford: Oxford University Press, 1999).

² See for instance Henrich’s discussion of Thomas Nagel in his review of *The View from Nowhere*. “Dimensionen und Defizite einer Theorie der Subjektivität,” *Philosophische Rundschau* 36 (1989): 1–24. My discussion of Moran draws on and in some respects parallels this piece.

³ In this connection, see Raymond Geuss, “Outside Ethics,” *European Journal of Philosophy* 11 (2003): 29–53.