Virtue, Authenticity and Irony: Themes from Sartre and Williams

Alan Thomas[1]

Abstract

In the course of criticizing indirect forms of consequentialism Bernard Williams argued that because virtues of character enter into the very content of the self, they cannot be instrumentalised. They must, instead, be viewed as cognitive responses to intrinsic value. This paper investigates this argument and relates it to similar claims in the work of Sartre. The inalienability of the first personal point of view represents a common theme and informs a further argument that an agent can only think of him or herself as merely one amongst others via a distinctive ethical use of the trope of irony.

In this paper I would like to explore the ramification of a single paragraph in Bernard Williams's *Ethics and the Limits of Philosophy*. (Williams 1985) It is concerned with indirect consequentialism's treatment of the virtues:

> The theory finds a value for these dispositions, but it is still an instrumental value. The dispositions are seen as devices for generating certain actions, and those actions are the means by which certain states of affairs, yielding the most welfare, come about. This is what those dispositions look like when seen from outside, from the point of view of the utilitarian consciousness. But it is not what they seem from the inside. ....The dispositions help to form the character of an agent who has them, and they will do the job the theory has given them only if the agent does not see his character purely instrumentally, but sees the world from the point of view of that character. (Williams 1985, pp. 107–8)

This contrast between a schematic "inside" and "outside" view seems to run throughout Williams's work: the passage presents this distinction—under one of its aspects—and

---

I will follow the convention of referring to concepts with small capitalization.

✉ Alan Thomas
  ap.thomas@york.ac.uk

1 Department of Philosophy, University of York, Sally Baldwin Building, York YO10 5DD, UK

gives an indication of how Williams hoped to generalize its application.

The explicit target of the passage is the indirect consequentialism of the kind familiar from the work of R. M. Hare—one of Williams's Oxford tutors. (Hare 1981) Views of this class putatively work by telling agents to cultivate virtues that are, in fact, consequentially justified at a reflective level as productive of the most value. Williams diagnoses a problem with this view: it seems to value the virtues only instrumentally while the agent *cannot help* but view them non-instrumentally. Williams tells us why immediately after the quoted passage: the explanation for *that* fact involves the further thought that "other things" have to be seen non-instrumentally—and that is allegedly *why* you cannot think of your own character as merely "devices …. for generating certain outcomes" (ibid.). Williams claims that it is constitutive of virtue possession that the virtuous person values both the virtue and its proper objects "for their own sake". Yet the way in which these issues are thematized in this paragraph from Williams gives rise to several challenges for this argument.

What sense of "cannot" does Williams invoke when we are told that agents "cannot" think of their virtues as devices with an instrumental-cum-consequentialist justification? Is this "cannot" compatible with these virtues *in fact* being devices that work this way—a truth somehow inaccessible to the agent? If this is, indeed, the constitutive truth about the nature of value and rightness, then what is it about human psychology that prevents our access to these truths? Given

the plasticity of human psychology—as notably thematised in Mill's account of 'Of the Ultimate Sanction of the Principle of Utility'—can we not imagine that the necessary adjustment to make these truths ones we can live with have to come from us—rather than from the rejection of the truth of some form of consequentialism? (Mill 1871/2015).
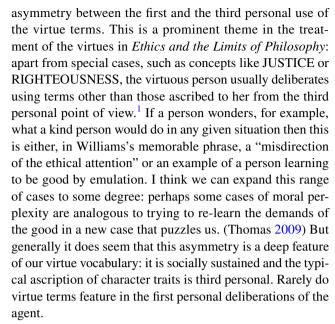
In this paper I will present a defense of Williams's position: I will argue that the issue is not whether one can take up an instrumental attitude to one's own character. Reflection on the similarity between Williams's moral psychology and that of Sartre suggests that the deeper problem is that of the possibility of an objectified view of one's own character—where objectification is prior to instrumentality. One's own character cannot be an object for one's own reflection in the way that Sartre first diagnosed in his critique of one conception of authenticity as "truth to one's real self". Irony will play a crucial role in explaining Williams's resistance of one manifestation of the impulse to impartiality such that one thinks of one's own character "as if" it were the character of another person.

This is not, literally, a metaphysical thesis about the absorption of the individual self into a psychic totality, nor a heuristic that advises the agent to act as if she were Hare's "World Agent" for whom all the practical projects of others are aggregated into a single model (Schopenhauer 1840/2000; Hare 1981; Williams 1985, pp. 83–84, 87–88). The plausible versions of this view emphasise the "as if"; the goal is the exploration of the idea that the impartial perspective works to discipline ethical actions that can occur only at the level of engagement and not that of reflective detachment (Nagel 1986, 1991). But not even that moderate version of the ideal, Williams implies, survives scrutiny and irony helps to explain why not.

1 Instrumentality and Objectification

Williams tells us that a virtuous person cannot think of her character merely as an instrumentally valuable device for producing certain outcomes; how unfortunate, replies the indirect consequentialist, because her virtues *are* merely devices for producing valuable outcomes! In order to address this concern, I think we need to go a little deeper into Williams's moral psychology, which is more indebted to Sartre's moral psychology than is usually acknowledged. (Indeed, more deeply indebted to Sartre than one might infer from Williams's occasional invocation of the idea of a "project".) Prior to thinking of one's character as an instrument is thinking of it as an object for one's own reflection: this thought is subjected to sustained criticism as part of Sartre's rejection of the ideal of authenticity.

The key to establishing any parallel between Williams's and Sartre's views is the former's claim that there is a deep asymmetry between the first and the third personal use of the virtue terms. This is a prominent theme in the treatment of the virtues in *Ethics and the Limits of Philosophy*: apart from special cases, such as concepts like JUSTICE or RIGHTEOUSNESS, the virtuous person usually deliberates using terms other than those ascribed to her from the third personal point of view.[1] If a person wonders, for example, what a kind person would do in any given situation then this is either, in Williams's memorable phrase, a "misdirection of the ethical attention" or an example of a person learning to be good by emulation. I think we can expand this range of cases to some degree: perhaps some cases of moral perplexity are analogous to trying to re-learn the demands of the good in a new case that puzzles us. (Thomas 2009) But generally it does seem that this asymmetry is a deep feature of our virtue vocabulary: it is socially sustained and the typical ascription of character traits is third personal. Rarely do virtue terms feature in the first personal deliberations of the agent.

Sartre has a more general explanation of why this should be so: it is connected to the fact that, for Sartre, there is a sense in which all psychology is moral psychology: any account of mentality must reconcile the first and third personal use of mental predicates to a synchronically unified mental subject; it is through the moral psychology of shame that Sartre illustrates his solution to his fundamental problem (Sacks 2005). Sartre sought to conjoin two claims: that conscious mentality involves both a phenomenological openness to the world and that mentality is an objective phenomenon. By the latter, I mean the acknowledgement of the minimal objectivity imposed by the grammar of our practice of mental ascription: that my ascription to myself of a mental predicate and your ascription to me of a matching predicate are both made true by the same worldly fact. (Thomas 2010).

A tension arises between these two commitments in the following way: Sartre claims that is it of the very nature of intentional awareness that it both actively posits—and yet is extinguished in—its intentional objects. (This is his radical externalism about intentionality.[2]) This explains the systematically elusive character of consciousness as Sartre conceives of it: this is his distinction between (first order) consciousness and (higher order) "consciousness-reflected-on".

---

[1] This paper reproduces, in its exposition of Sartre's ideas, some sentences of Thomas (2010). I have not rewritten these sentences because they are merely expository of another philosopher's views, as opposed to my own views.

[2] The alternative is to posit, as the truthmakers of perspectival judgements, perspectival "facts"; that this would be a serious error is demonstrated by John Biro in his (1991, 1993, 2006) see also (Thomas 2008, pp. 28–29).

Sartre's metaphor for consciousness—that he also calls a "nihil"—is light.[3]

> I am as inseparable from the world as light, and yet exiled as light is, gliding over the surfaces of stones and water, never gripped and never held. (Sartre 1947/1992)

Light makes vision possible without being an object of vision; consciousness enables the intentional representation of objects of awareness while being systematically elusive to higher order, or "second order", reflection.[4] The problem, then, is that if this is the character of consciousness, then what does reflection find when it directs its attention to self-knowledge? The elusiveness of first order conscious mentality explains why it eludes our grasp: all we find, in its place, are those worldly objects upon which it was intentionally directed. But how can that intuition be cashed out when it comes to knowledge of oneself—when reflection turns on itself? Mark Sacks has a subtle appreciation of the problem—as expressed in this passage:

> As long as my awareness is rooted solely in first personal thought I cannot, by reflection, dissociate myself from my subjective stance and identify myself as an object. Every object that I come to perceive as an object, is presented as essentially distinct from me, the perceiver; and just insofar as it is presented to me as a distinct object, it is presented as in principle alienable from me. The only item that is not presented as separable from me in that way, the only thing that is given to me with an immediacy that precludes the coherence of my thinking it away, is my own point of view. But that is always given as a point of pure subjectivity: it is behind the lens, so to speak, whereas anything captured as an object is always in front of it. (Sacks 2005, p. 288)

My surmise is that when Williams refers, in the passage with which I opened this paper, of the world "seen from the point of view of that character" he is referring, precisely, to this "inalienability" of one's first personal point of view. But if anything that is placed "before the lens" is alienable, what happens when the object before the lens is one's own character?

There will either be something alienable—the problematic sense of appearing to oneself as an "object"—in which case we have simply misplaced our subject matter.

As Richard Moran has argued the distinctiveness—not any putative Cartesian "privilege"—of the first person will have been lost. (Moran 2001) But a recoil to the point of view of "pure subjectivity" is no help either: our metaphor of a point of view is an inherently realist metaphor. (Thomas 2010) It is a point of view on something; on this pole of the problem when we try to focus on our own character we are, instead, delivered to the world of objects in a way that reflects Sartre's commitment to radical externalism. So on Sartre's general conception of mentality it is not only consciousness that is systematically elusive; one's own character is, too. We seem to be delivered into a paradox: the unacceptable conclusion that, as Moran notes, that which is closest to us and which ought to be known the most easily—ourselves—cannot be known at all.

Sartre's solution to this problem is ingenious: we find the substance of self-knowledge in the moral psychology of shame. As I have reconstructed elsewhere, in his famous account of "the Look" in *Being and Nothingness* Sartre exploits the inter-dependence of the first and third personal perspectives on mentality to argue that those patterns of your past behaviour that constitute your character are, for you, an interpretative artefact of how others view you (Sartre, 1947; Thomas 2010). The only solution to the envisaged problem compatible both with the phenomenology of the first person and with treating the first- /third-person asymmetry as basic to our idea of mentality is to form a conception of oneself as one is seen by others. It is in the experience of being ashamed that I feel ashamed of myself through and through, as a whole person, in the gaze of another[5] (Sacks 2005).

The Other in this example need not actually be present; shame can be triggered by the fact that I could be observed. That phenomenological fact, in turn, is part of seeing the world as structured around other centres of consciousness each of which is a "for-itself" that transcends its own reality just as I transcend mine. Sacks emphasizes that there is a positive aspect of one's presentation, as an irreducibly whole person, to the gaze of the other. (Sacks 2005) It is as oneself that one is so presented. (Theunissen 1984, pp. 224–227, 230, 236–237).

This point is obscured because we associate Sartre's moral psychology with the so-called "pathologies of self-regard" that constitute bad faith. But this critical diagnosis

---

[3] For a similar view, defended in a contemporary idiom, see van Fraassen (2004).

[4] This puts me at odds with the interpretation of Sartre's views in Gennaro (2002); this is not the place to rehearse our differences, but I think the distinction between thetic and non-thetic consciousness is a serious problem for Gennaro's interpretation.

[5] As Williams put it in later work "The root of shame lies in exposure in a more general sense, in being at a disadvantage: in what I shall call, in a very general sense, a loss of power." (Williams 1993, p. 220). Williams continues "(The) other still performs a function, of recalling to the subject a person in the eyes of whom the subject has failed, has lost power, is at a disadvantage. In contrast to guilt, there is no need with shame that the viewer should be angry or otherwise hostile" (ibid, p. 221).

of these pathologies depends on the prior, positive, account of self-knowledge as expressed in the account of "the Look". The Sartrean diagnosis is that bad faith arises only when one views oneself merely as an object—the word "merely" is crucial here. For Sartre, our never being "self-identical" in choice is that we know ourselves both as transcendence and as facticity. The pathologies of self-regard result from a collapse of one of these aspects of authentic choice into the other.

How literally ought we to take this foundational Sartrean contrast between the "in itself", a world of objects which are necessarily self-identical, and the tragedy of conscious intentionality, the world of the "for itself" that strives for, but can never realise, this condition of self-identity?[6] If this metaphysical contrast is the price Williams has to pay for an attractive moral psychology one wonders if he would be prepared to pay it.

Fortunately, however, I think Charles Larmore has shown that there is a way to deflate this metaphysics so that it is no more than a commitment to the kind of Davidsonian interpretationism about the mental to which Williams became increasingly sympathetic as his work developed[7] (Larmore 2010). In Larmore's account Sartre's "tragedy of consciousness"—his diremption thesis—is explained wholly in terms of an account of normative commitment. For Larmore, to be a self is to be located in the space of reasons, where this is to be able to take up various normative commitments. Commitment to contents in thought is commitment to the inferential consequences of accepting that content as true (for belief) or desirable (for desire). This very idea of taking up a commitment, however, captures the idea of the non-coincidence of self with self that Sartre mistakenly took to be a metaphysical notion.

For Larmore, "non-identity" is not a metaphysical notion, but the appeal to commitment explains why one might mistakenly think that it is. That is because in commitment there is *difference within unity*: to place the self in a space of inferential commitments involves a distinction between the *committing* and the *committed* self. The committing self has to envisage, in advance, the space of inferential possibilities that it will occupy as it takes up one, or the other, metaphorical "location" in this space of rational commitments consequent on accepting a content.[8] Commitment is

to hold oneself open to a set of inferential pathways such that the self is "committing itself to be what it has not yet become" (Larmore 2010, p. xv). Sartre's opaque claim about the essential non-coincidence within the for-itself has a more straightforward explanation. I think Larmore's transposition of a putatively metaphysical claim into a wholly normative one is a real insight.

While there are deep affinities between Williams's view and those of other recent interpreters of Sartre in the analytic tradition, notably Moran and Larmore, the key difference between Williams and these other commentators is whether or not practical reasoning is essentially first personal (Williams 1985). Both Moran and Larmore are normativists about the mental; both want to emphasise a role for the first person; this is combined in Larmore, but not in Moran, with a conception of one's self-relation as essentially practical (Moran might prefer to say "deliberative"). But only in Williams do we get the further thesis that even within this normative conception of mind there is a further distinction between theoretical and practical reasoning. The former, Williams tells us, can be first personal, but is never essentially so. I have elsewhere argued that the grounds for this thesis—that a practical conclusion true for "anyone" always requires supplementation by the further conclusion that it is true for oneself—requires the Aristotelian thesis that the conclusion of practical reasoning is the content expressed by what one actually does. (Thomas, forthcoming) But whatever the grounds for the thesis—and Williams tells us that he later repudiated it—it goes beyond Larmore's and Moran's normativism about the mental. In my view, the first person is not only distinctive; in the case of practice, it is distinctively expressed by action—what the agent actually does.

2 Reason Sensitivity and 'Strict Liability'

Larmore's plausible account of normative commitment helps to deflate one Sartrean thesis: I also think it helps to explain another, namely, the claim that there is no "inertia" in consciousness. (A thesis that I will, in the next section, also put to use in defense of Williams.) This latter claim can be interpreted as the normativist thesis that, for any mental state of yours, if you tolerate it, then you are underwriting

[6] For a very helpful discussion of this Sartrean thesis see Webber (2009), pp. 108–110.

[7] After his early book on Descartes, where Williams endorsed Quinean indeterminacy about the mental, Williams was thereafter committed to Davidsonian interpretationism. This is clearest in the late *Truth and Truthfulness*, but arguably his earlier thesis that all practical reasons are internal depends on a Davidsonian view of action explanation. See Thomas (2006, chapter four) and Finlay (2009).

[8] To be committed is to undertake various possible ways of developing one's commitments. This brings in, for example, the imagination as an essential component of normative commitment. This is another point of overlap with Williams: central to the formulation of his internal reasons thesis is the claim that the internal sense of reasons ascriptions is necessarily indeterminate because of the role played in them by the imagination. Williams here took himself to be expositing the views of his one time Bedford College colleague Aurel Kolnai (1973).

it. That is because the radical interpreter of your thought and talk can interpret you as committed to it. Openness to the world—including openness to value—is a rational orientation towards that world that is subject to a norm of "strict liability". So endorsement takes two guises: active and passive: both contrast with *repudiation*.

On the view I am reconstructing here, any state to which you acquiesce, or voluntarily underwrite, is reason-sensitive; furthermore, all reason-sensitive states can be integrated into Sartrean "projects". Such projects are structured hierarchies of commitments, where "commitment" here bears the expanded sense that correlates with Sartre's strict liability norm: if you tolerate a state, then you are committed to it. As Jonathan Webber has pointed out, Sartre's idea of a rationally assessable project to which one can take an agent to be committed extends to such "projects" as staying alive. (Webber 2009, p. 49) That is not a purposive, goal directed activity, but a commitment rationally to underwrite acquiescing in a given state of affairs. It is, in that sense, a helpful example of the reason-sensitive nature of a mental commitment even in a case where it seems "passive". This model of reason-sensitivity also explains why Sartre is not a superficial "voluntarist".[9] This is not some neo-Stoic moral psychology where the contents of states are accompanied by a mental "toggle" of endorsement or repudiation; so the fact that, for example, I am currently being assailed by a range of beliefs formed on the basis of my immediate perceptual experience does not embarrass the Sartrean view. I am, in one sense, "passive" towards these beliefs, but given that they are reason-sensitive, to acquiesce to them can be interpreted as endorsing them.

Combine these claims with the asymmetry between first and third personal access to the mind and an important corollary for the Williams argument from which I began is that in integrating another person's conception of you as your conception of yourself, then the latter must have the capacity to outstrip the former:

Consciousness does not know its own character – unless in determining itself reflectively from the standpoint of another's point of view. It exists its character

in pure indistinction non-thematically and non-thetically in the proof which it effects of its own contingency and in the nihilation by which it recognizes and surpasses its facticity.[10] (Sartre 2003, p. 372)

This typically Gnostic paragraph can be explained in two ways: the first point invokes the norm of "strict liability" in that any conception of yourself to which you acquiesce you thereby endorse. But that point must be essentially qualified by a second: the pattern in your conduct that the other has discerned in your past actions and reflexively "returned" to you was, after all, a pattern *in your past rational responses to situations*. It is only in so far as you are an appropriate subject for such rational interpretation that this conception of yourself for another has any salience for you. But you have not ceased to be such a rational subject and are thereby open to the continuing demands, not only of your own past character, but of the world. In knowing your own character via the way you are interpreted by the Other, then you must, thereby, transcend that character.[11]

Again, absent Sartre's special terminology, I take the point to be straightforwardly expressible in terms that mention a normative stance and a commitment to reasons: the rational interpreter found a pattern in your responses because your attention was directed outwards, towards the world. This pattern in a set of rational responses lacks the fixity of the state of a mechanical device: the dispositions that constitute your character are an inter-locking set of multi-track dispositions that constitute a mode of response.[12] They were a response to the evaluative features presented to you by past situations. When Sartre says, deliberately courting paradox, that you know yourself in *not* merely being this conception of yourself the explanation is not, in fact, paradoxical. The explanation is that as an on-going subject of rational assessment your character is an "open" pattern *for you* because you are uniquely responsible for the impact of that previous pattern on your future choices. Sartre's strict liability norm, as I have called it, is essential to the formulation of this

---

[9] Dick Moran offers an interesting gloss on this Sartrean claim that "Choice and consciousness are one and the same thing. But if it has been well established that consciousness is a nihilation, the conclusion is that to be conscious of ourselves and to choose ourselves are one and the same" (Sartre, 2003, p. 595). Moran comments: "(Sartre's) idea is not without its own unclarities, but it is quite different from saying I am free to pick and choose among the convictions, values and desires I would like to have. His language does suggest, however, that it is a consequence of such reflection that a situation of choice is somehow *forced* on the person, such that whatever he does now with respect to this "psychic given" must count as either his commitment to it or his acquiescence in it" (Moran 2001, p. 140).

[10] As I have explained, by "non-thematic" and "non-thetic" Sartre implies that knowledge of consciousness is non-propositional and perhaps better compared to knowledge by acquaintance. But this is less a case of "knowing how" than an insight into an act of intentional positing: so it is direct, acquaintance-like, insight into the exercise of an agent's capacity for intentional representation. By the "nihilation" of its "faciticty" Sartre seems to mean that this systematically elusive act of intentional positing is both exhausted by, and absorbed into, its intentional object.
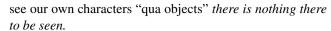
[11] It is also the entry point for Sartre's account of embodiment that, for reasons of space, I cannot discuss here. For a general perspective see Widder (1997). For an analysis with a direct bearing on transcendence and facticity see Longuenesse (2008, pp. 10–12).

[12] To my mind virtue ethics does not pay enough attention to the multi-track nature of those capacities of mind that we refer to as the virtues; for a sophisticated recent account see Vetter (2013).

thesis as once again, mere acceptance will be construed as endorsement. As I will explain below, if you ever identify with being "merely" how the other conceives of you, then the result can only be an instance of irony. The strict liability norm has, as one of its corollaries, the claim that there is no such thing as "mere" acquiescence.

Authentic choice always involves both facticity and transcendence, bearing in mind Sartre's extended use of the idea of "choice"—one that does not involve a superficial voluntarism. Your aim, qua deliberator, is to be rationally sensitive to the demands of the situation.[13] This is largely pre-determined by the patterns of relevance and salience in the situation to which you are attentive that have been established by your virtues. This represents your facticity: the real pattern within your rational responses to past situations that forms the reality of your character for another. However, from your first personal perspective your engagement is with *the situation*. Your action must be a correct response to this situation regardless of how much your apprehension of this situation has been structured in advance by your prior facticity. Terminology aside, Sartre's theory is an account of responsible choice from within an interpretationist framework: to accept the conception of yourself merely as another sees you would be to abnegate responsibility for choice—one form of bad faith in which facticity displaces transcendence. To affirm a limitless freedom not "answerable to evidence", as Moran puts it, would be the complementary irresponsibility that constitutes its converse in which transcendence displaces facticity. (Moran 2001, pp. 87–88; Hampshire 1975; Gardiner 1977; Thomas 2010) But underpinning these familiar—if obscure—pathologies of self-regard is Sartre's positive account of self-knowledge and its ethical implications.

3 From Sartre to Williams?

How might these Sartrean arguments help Williams in his defence of the claims with which I opened this paper? First of all, theyoffer a deeper grounding for the asymmetry between first and third personal virtue ascriptions: that asymmetry is grounded in postulating such an asymmetry for a wide range of mental predicates. Secondly, it explains the problem of "instrumentality" adverted to in the passage I have taken as my starting point: the Sartrean gloss on this text runs as follows: when we attempt, first personally, to

see our own characters "qua objects" *there is nothing there to be seen.*

That seems too radical: what of "the world (as seen) from the point of view of that character"? But that remark, too, cancels out simply into "the world": you do not "occupy" your point of view on the world. It would be less misleading simply to say you *are* your point of view on the world, but that simply takes one to (in the light of Sartrean externalism) the world. (Thomas 2010) That point, too, has a deflated but no less interesting explanation: the moral agent is phenomenologically orientated to a world that is already shaped by her concerns and interest. However, that process of shaping is invisible to her: it is a determinant of the "world as she finds it", but not represented within it. Its centers of value, patterns of salience and relevance, its manifestation of motivational potential are, for the agent, simply there; what cannot be "there" for her is the equally revealing domain of the *unthinkable*—a category of thoughts to which Williams repeatedly drew attention. (Williams, 1973, pp. 92–3) I, the interpreter, take it as important to my conception of your character that there are some considerations that it does not, or could not, even occur to you as ethically relevant. This necessarily evades your first personal grasp of the deliberative situation.

For Sartre, the explanation of this expansive conception of moral phenomenology is the extensive role he gives to the imagination not simply in practical deliberation (as in Williams's account of internal reasons), but in shaping the field of possibilities within which action is orientated: for example, you can "see" your garden lawn as exhibiting "to be mowedness" —worldly states of affairs as *already* freighted with motivational content. That is why, like Scanlon, Sartre thinks that moral philosophers over-dramatise the work of deliberative reasoning: the work of such deliberation has already been done by one's overall conception of the "Gestalt" of the evaluative features presented by the situation. (This also explains Sartre's particular focus on radically dilemmatic situations where these resources fail.)

Overall, then, we have a deeper critique of the ambitions of indirect theory. The reason we cannot think of our own character traits as forming, overall, a set of instrumentally-valuable dispositions is that instrumentalityis dependent on a prior idea: objectification. It is not then, that we "cannot but help" think of our own characters as not merely a device when they *are*, in fact, merely devices for the promotion of outcomes. The "cannot" here is constitutive and not merely psychological.

---

[13] As Moran expresses this ideal: "When I *avow* a belief, I am not treating it as just an empirical psychological fact about me; and to speak of a *transcendental* stance towards it is meant to register the fact that it commits me to the facts beyond my psychological state; and as a commitment it is not something I am assailed by, but rather is mine to maintain or revoke" (Moran, 2001, p. 89).

4 Hybrid Theory and the Trope of Irony

One of the reasons for the eclipse of Williams's thought in recent normative ethics is that it can seem to have lost out to one of its closest competitors: hybrid theory. (Scheffler 1983; Nagel 1986) The hybrid theorist claims that Williams's critique of the impersonality of moral theory is well taken; theory must encompass the values of the personal. But while it is obvious that an impersonal theory cannot do that, it is less obvious that an impartial theory cannot do so. Hybrid theory is a form of impartialism that falls within the class of "moderate moralities": it incorporates the importance of the personal point of view because each of us is an individual with our own lives to lead. A full consideration of this view is outside the scope of this paper; here I want to address its foundational commitment to taking a conception of "oneself merely as another" as the privileged way in which a discipline of objectivity is brought to bear on ethics—even if, for the hybrid theorist, it is not the only way in which objectivity bears on the ethical. The hybrid theorist can accept Williams's trenchant critique of Sidgwick: he was correct to insist that the "point of view of the universe" cannot be a standpoint of agency. (Williams 1989) But a standpoint of agency—necessarily first personal and engaged—can nevertheless still be informed by a discipline imposed, at the reflective point of view, by a de-centered view of the world where information is rendered inaccessible. Paradigmatic of such restricted information is the centering of those concerns on *oneself*. (Nagel 1986) That thought can be deployed as part of a creative re-interpretation of Sidgwick's point: the practical is not the theoretical, agreed, but an analogue of a theoretical view of the world can play a disciplinary role in our ethical thinking. This conception puts to use a point on which Williams and Nagel can agree: that theoretical reasoning, if first personal, is not essentially so. An objective conception of the world removes indexical information that centers that world on any particular agent.

For this style of impartialist, thinking of yourself as one person amongst others equally real is what it *is* to think objectively in the domain of the practical. Whereas in Nagel's early work all genuine reasons were to be explained as the combination of a first personal perspective on a non-perspectival content, in his later work—and in that of Samuel Scheffler—the impartial point of view is not all encompassing, but it is all constraining (Nagel 1970, 1986; Scheffler 1983). There are the reasons it directly vindicates, and those it merely tolerates, but even in the latter case the de-centred standpoint retains its authority. It is no longer true that all values and reasons are "objective"; but they all have to placed in *some* relation to this authoritative standpoint.[14]

There could hardly be a more direct contrast between two views than the Nagelian claim that the objective point of view, in which one thinks of oneself merely as another, is a uniquely privileged form of objective ethical thinking while the view I have attributed to Sartre and Williams is that you can *never* think of yourself in this way (without irony). Even worse, common sense seems to support the intuition that motivates the hybrid theorist: after all, it can seem to be a matter of mere common sense that it can be ethically important to look at the world, ethically, while not privileging one's own point of view. Natural partiality can easily become tainted partiality just as, as Williams puts it, fantasy may naturally subvert belief. (Williams 2002) If it is a commonplace thought that we can need to discipline our thinking in this way, then how can Sartre and Williams declare it to be impossible? However, I think there is a further point to be made in favour of the Sartre-Williams conception: an explanation of how irony functions in this context.

This proposal draws upon an important observation made by Roger Scruton in his characterisation of "the culture of forgiveness":

> From the culture of forgiveness springs the other habit that helps us to be at home in the society of strangers. This is irony, by which I mean the habit of acknowledging the otherness of everything, including oneself. However convinced you are of the rightness of your actions and the truth of your views, look on them as the actions and views of someone else, and rephrase them accordingly: such is the principle by which, in our better moments, we wish to live .... Irony is quite distinct from sarcasm: it is a mode of acceptance rather than a mode of rejection. And it points both ways: through irony I learn to accept both the other on whom I turn my gaze, and also myself, the one who is gazing. *Irony is not free from judgement: it simply recognizes that the one who judges is also judged by himself.* (Scruton 2010, pp. 82–3, emphasis added)

For Sartre, and for Williams, as I interpret them the trope of thinking of oneself merely as another is, indeed, merely that: a specific ethical trope. Indeed, it is the ethical appropriation of the trope of irony as Scruton describes its workings in this quotation.

What does this trope involve? It involves both doubling and unsuspecting confidence. There must be two points of view: that of the subject of irony (the ironized) and the perspective of the ironist who identifies an unsuspecting confidence on the part of the ironized about their predicament

---

[14] Although Nagel is sensitive enough to the claims of moral phenomenology in *The View from Nowhere* to acknowledge that the

Footnote 14 (continued)

extensive category of agent-relative values are in its "blind spot" (Nagel 1986; Thomas 2008).

(Thomas 2006). Irony is central to Sartre's conception of the mind: From an apperceptive, normative, perspective the self is essentially non-identical with any of its empirical predicates (Rosenberg 1981). In the more localized case of his moral psychology, the locus of ethical judgement knows herself as one person amongst others equally real, but as not merely that. To act with unsuspecting confidence would be to act as if one were merely oneself-as-another; ironized confidence both acknowledges and transcends that stance. If any ethical concept has a claim to be of central importance to Williams, it is confidence (Williams 1985, p.170):

> One question we have to answer is how people, or enough people, can come to possess a practical confidence that, particularly granted both the need for reflection and its pervasive presence in our world, will come from strength and not from the weakness of self-deception and dogmatism. (Confidence is not the same as optimism; it could rest on what Nietzsche called the pessimism of strength.) (Williams, ibid.)

Ethical confidence, in this conception, is not the unsuspecting confidence of the person ironized—any more than the realism Williams finally endorses avoids being a "chastened" form of realism in an echo of the epigraph he selected from Wallace Stevens (Krishnan & Queloz 2022).

Ethical irony is, then, a particular instance of Sartre's claim that there is no inertia in consciousness. I have interpreted this claim by appealing to Sartre's brand of normativism: any mental state in which you acquiesce an interpreter may attribute to you *as endorsed*. So Scruton's ironist is accepting a conception of him or her self "as foranother". But she does so ironicallyand therefore takes a risk endemic to irony:

> The ironic figure of speech cancels itself .... inasmuch as the one who is speaking assumes that his hearers understand him, and thus, through a negation of the immediate phenomenon, the sense becomes identical with the phenomenon. (Kierkegaard 1989, p. 248)

You may, for specific and tactical purpose, think of yourself as *that*—as others see you; but it remains true that you can never accept yourself as *merely* that. So ironic acceptance, following Sartre's more general point, is as Scruton describes it: "it is not free from judgement". Judgement is not suspended as it would be if you were to accept that you were *merely* that; ironic acceptance is a mode of acceptance *for a specific ethical purpose*. But, as a specific trope, its role is local, tactical and hence limited. It is not the uniquely privileged way in which we bring the demands of objectivity to bear on the ethical; it is, rather, one way to avoid the temptations of a tainted partialism. The hybrid theorist was wrong to take Williams's critique of consequentialism to depend solely on the point that any such view is objectionably impersonal; the underlying point is that there can be irreducibly partial reasons in ethics (Cottingham 1983, 1986, 1991, 1997; Thomas 2005) Partiality, too, has its own discipline: the trope of irony plays a role in this.

Indeed, the situation seems to me to involve a yet deeper irony: while hybrid theory takes the impartial point of view to encompass the personal, either by direct vindication or toleration, for Sartre and Williams the situation is precisely reversed. The personal point of view—the standpoint of engagement, authentic choice and action—encompasses the particular ethical trope of a de-centered view of the world that it tolerates for one specific ethical purpose. It can be ethically useful, on occasion, to think of oneself as, in Scruton's terms, a stranger in a society of strangers: to be other to everyone, including yourself. However, this is not, as the hybrid theorist claims, the all encompassing form that objectivity must take: any temporary suspension of the "magic in the pronoun 'My'" is a particular way in which a conception of oneself "returned" to you by the other is put to ethical use.

5 Openness to the World and Value

If the account I have presented so far is attractive qua philosophical argument, then can it also be made convincing qua interpretation, specifically, as an interpretation of Williams? Sartre is hostile to any "material" theory of value as his tradition calls it; I seem to be interpreting Williams as a cognitivist virtue ethicist for whom the virtues are a mode of evaluative *response*. Only with this conception in place can we understand virtue as a way of knowing anything and hence as a form of openness to the specific evaluative features manifested in a person's experience. This seems to ignore Williams's "non-objectivism" about the ethical as presented in *Ethics and the Limits of Philosophy.*

I certainly have my doubts about non-objectivism and its suspiciously foundationalist approach to the reflective displacement of putative ethical "knowledge", and I have expressed them elsewhere, but I do not need to pursue them here. (Thomas 2006) That is because all I need, for current purposes, is the minimal and uncontroversial observation that Williams is not a projectivist about the ethical. From the phenomenological point of view he thinks that he can agree fully with the critique of the projectivist developed by the so-called "secondary property" realists such as Wiggins and McDowell. He certainly has his disagreements with them, or he would not be a non-objectivist, but it is not over their cognitivism. (Where, by cognitivism here, I mean the view that a core of ethical judgements are truth-apt, often true and often known to be so.) For the right kind of agent— the virtuous person—the virtues are modes of evaluative response; that can be granted. Now, Williams also famously says that ethical thought cannot be "all that is seems"; but

that is compatible with it being at least what it seems and for present purposes that is enough. The virtues certainly seem to be a mode of evaluative response and that thought can survive the very general and abstract thought that motivates the non- objectivist: that some particular ethical way of going on might be merely one way amongst other forms of ethical life that are equally admissible as expressions of our distinctively human forms of ethical life.

6 Conclusion

Many of the arguments of *Ethics and the Limits of Philosophy* have a negative and skeptical tone; the culmination of the positive, reconstructive phase of Williams's thinking occurs later in *Shame and Necessity* and *Truth and Truthfulness* (even if the former is an act of historical recovery of an underlying historical truth from its philosophically motivated misrepresentation). (Williams 1993; 2002) In *Truth and Truthfulness* in particular, Williams defends an account of the epistemic virtues that re-deploys many of the points of the paragraph with which I opened my discussion. Truth telling, and the practices in which it is embedded, requires valuation for its own sake as does its supporting dispositions of Sincerity and Accuracy. Mis-conceptions of the truth about self-knowledge have, in our own philosophical tradition, led not only to philosophical error, but also to political errorin the guise of dangerous political fantasy. I have tried to suggest here that alongside the genealogical vindication of our ethical dispositions (and their supporting ethical emotions invoked in thoughts such as "it is shameful to lie") we need also to appeal to arguments in support of Williams's position that could, broadly, be called phenomenological. Interpreting Williams as committed to a moral psychology that he largely shares with Sartre seems to me to make the best overall sense of his philosophical commitments.

Declarations

References

Biro J (1991) Consciousness and subjectivity. In: Villanueva E (ed) Consciousness: philosophical issues. Ridgeview Publishing, Atascedero

Biro J (1993) Consciousness and objectivity. In: Davies M, Humphreys GW (eds) Consciousness. Basil Blackwell, Oxford

Biro J (2006) A point of view on points of view. Philos Psychol 19(1):3–12

Cottingham J (1983) Ethics and impartiality. Philos Stud 43:83–91

Cottingham J (1986) Partiality, favouritism and morality. Philos Q 36(1986):357–373

Cottingham J (1991) The ethics of self-concern. Ethics 101:798–817

Cottingham J (1997) The ethical credentials of partiality. Proc Aristotelian Soc XCVIII:1–21

Finlay S (2009) The obscurity of internal reasons. Philos Imprint 9(7):1–22

Gardiner P (1977) Sartre on character and self-knowledge. New Literary Hist 9(1):65–82

Gennaro R (2002) Jean-Paul Sartre and the HOT theory of consciousness. Can J Philos 32(3):293–330

Hampshire S (1975) Freedom of the individual. Princeton University Press, Princeton

Hare RM (1981) Moral thinking: its methods, levels and point. Oxford University Press, Oxford

Kierkegaard S (1841/1989) The concept of irony with continual reference to Socrates, Edited and translated by Howard V Hong, Edna H Hong, Princeton University Press

Kolnai A (1973) Deliberation is of ends. In: Ethics, value and reality. Athlone Press, London, pp 44–62

Krishnan N, Queloz M (2022) The shaken realist: Bernard Williams, the war, and philosophy as cultural critique. Eur J Philos 31(1):226–247

Larmore C (2010) The practices of the self. University of Chicago Press, Chicago

Longuenesse B (2008) Self-consciousness and self-reference: Sartre and Wittgenstein. Eur J Philos 16(1):1–21

Mill JS (1871/2015) In: Philp M (ed) On liberty, utilitarianism and other essays. Oxford World's Classics. Oxford University Press, Oxford

Moran R (2001) Authority and estrangement: an essay on self-knowledge. Princeton University Press, Princeton

Nagel T (1970) The possibility of altruism. Oxford University Press, Oxford

Nagel T (1986) The view from nowhere. Oxford University Press, New York

Nagel T (1991) Equality and partiality. Oxford University Press, New York

Rosenberg J (1981) Apperception and Sartre's "Pre-Reflective Cogito." Am Philos Q 18(3):255–260

Sacks M (2005) Sartre, Strawson and others. Inquiry 48(3):275–299

Sartre J-P (1947/1992) The reprieve (trans: Sutton E). Vintage Press, New York

Sartre J-P (2003) Being and nothingness: an essay in phenomenological ontology (trans: Barnes HE), revised edn. Routledge, London

Scheffler S (1983) The rejection of consequentialism: a philosophical examination of the considerations underlying rival moral conceptions, revised. Oxford University Press, Oxford

Schopenhauer A (1840/2000) On the basis of morality (trans: Payne EJF). Hackett, Indianapolis

Scruton R (2010) The uses of pessimism. Oxford University Press, Oxford

Theunissen M (1984) The other: studies in the social ontology of Husserl, Heidegger, Sartre and Buber (trans: Macann C). MIT, Cambridge

Thomas A (2005) Reasonable partiality and the agent's point of view. Ethical Theory Moral Pract 8(1):25–43

Thomas A (2006) Value and context: the nature of moral and political knowledge. Clarendon Press, Oxford

Thomas A (2008) Thomas Nagel. Routledge, London

Thomas A (2009) Consequentialism, integrity and demandingness. In: Chappell T (ed) The problem of moral demandingness. Palgrave Macmillan, New York, pp 123–147

Thomas A (2010) Alienation, objectivity and the primacy of virtue. In: Webber J (ed) Reading Sartre. Routledge, London, pp 161–179

van Fraassen Bas (2004) Transcendence of the ego (the non-existent knight). Ratio XVII:453–477

Vetter B (2013) Multi-track dispositions. Philos Q 6(251):330–352

Webber J (2009) The existentialism of Jean-Paul Sartre. Routledge, New York

Widder K (1997) The bodily nature of consciousness: Sartre and contemporary philosophy of mind. Cornell University Press, Ithaca

Williams B (1985) Ethics and the limits of philosophy. Fontana, London

Williams B (1989) Sidgwick and the point of view of the universe. In: *Making sense of humanity and other essays*. Cambridge University Press, New York

Williams B (1993) Shame and necessity. California University Press, Oxford

Williams B (2002) Truth and truthfulness: an essay in genealogy. Princeton University Press, Princeton

Williams B, Smart JJC (1973) Utilitarianism: for and against. Cambridge University Press, Cambridge