Murdoch on Misunderstanding

Kieran Setiya

MIT

September 25, 2025

(Draft; do not cite without permission)

In *The Sovereignty of Good*, Iris Murdoch explores the nature of moral thought detached from overt agency—as in the well-known example of the mother, M, who revises her conception of her daughter-in-law, D, without altering her outward behaviour (Murdoch 1970: 16-23). For Murdoch's opponent: "Morality must be action because mental concepts can only be analysed genetically" (Murdoch 1970: 15). On the genetic theory, "the possession of [a] concept is a public skill" (Murdoch 1970: 11). The theory is "genetic" because it holds that there is no more to grasp of concepts than is manifest in their acquisition: if you've learned what a word means by conventional standards, you fully grasp the corresponding concept.[1] This leaves no room for the change in M's relationship with concepts as she redescribes D—"not vulgar but refreshingly simple, not undignified but spontaneous, not noisy but gay, not tiresomely juvenile but delightfully youthful, and so on"—to count as the improvement in conceptual mastery it is (Murdoch 1970: 17).

Despite her shift away from observable action, Murdoch goes on to propose a striking view about acting for reasons. The upshot of activity like M's, when it goes well—"a refined and honest perception of what is really the case … which is the result not simply of opening one's eyes but of a certain perfectly familiar kind of moral discipline"—is that the world is "*compulsively* present to the will" (Murdoch 1970: 37, 38). "If I attend properly," Murdoch argues, "I will have no choices and this is the ultimate condition to be aimed at" (Murdoch 1970: 38). In a slogan that appears in the second essay of *Sovereignty*: "true vision occasions right conduct" (Murdoch 1970: 64).

[1] See Murdoch 1970: 12-13 on how we learn the concept of decision.

Murdoch seems to countenance here a radical form of "motivational internalism": not that reasons must be capable of moving us to act (as per "internalism about reasons"), or that one is invariably moved by the belief that one has reason to φ (as per "judgement internalism"), but that one is bound to respond to a reason correctly if one fully grasps the fact in which that reason consists. Thus:

> RADICAL INTERNALISM: If the fact that *p* is a reason for A to ϕ, and A fully grasps that *p*, A is moved to ϕ in proportion to its weight as a reason.[2]

Assuming that facts about others' needs provide us with reasons to act, and that it is possible to know the meaning of "need" by conventional standards and yet to be indifferent to others, Radical Internalism conflicts with the genetic theory of concepts.

If Radical Internalism is true, the fact-value distinction breaks down, not just because there are normative facts—facts about what is a reason for what—but because, as Murdoch puts it, "mental concepts enter the sphere of morality" (Murdoch 1970: 23-4). Facts typically treated as non-ethical; the facts on which, we are told, the ethical supervenes; facts that provide us with reasons to act: for Murdoch, these facts share a feature often thought definitive of normative or evaluative facts, that full grasp of them is necessarily motivating. (Following Murdoch, I draw no substantive distinction between "ethics" and "morality"—the topic is how we should live, where that includes our relation to others.)

What is more, Murdoch gestures towards a theory of concepts that would vindicate her position—a form of "conceptual role semantics" on which "[the] entry into a mental concept of the notion of an ideal limit destroys the genetic analysis of its meaning" (Murdoch 1970: 28).[3] Each concept is associated with standards for its ideal use, both practical and theoretical, norms that specify where the concept should be applied and what follows from its application, in relation to cognition and

[2] This formulation is adapted from a slightly different one in Setiya 2013: 7. For a similar position, see McDowell 1978.

[3] I defend this interpretation of Murdoch at greater length in Setiya 2013: §3.

the will. To grasp a given concept is to approximate, in one's dispositions of thought, conformity with these norms. Concept-possession thus comes by degree, and points to an ideal limit: perfect compliance with the norms by which our concepts are defined. On this Platonic theory, we must use concepts well, or as we should, at least by approximation, in order to grasp them at all. The norms are at once a condition of thought and an ideal to which we aspire. In this way, the good is inscribed within each concept. The truth of Radical Internalism follows from the Platonic theory of concepts, assuming that the norms of concept-possession require one to respond to reasons that their application gives. As Murdoch puts the point, abstractly and elusively: "*The necessity of the good is then an aspect of the kind of necessity involved in any technique for exhibiting fact*" (Murdoch 1970: 64). The motivational necessity of reasons is an aspect of the conceptual mastery required to fully grasp the facts.

On this reading, Radical Internalism is at least not obviously false. The form of conceptual role semantics on which it rests is a close cousin of views that philosophers of mind and language still take seriously.[4] As I have argued elsewhere, the truth of Radical Internalism bears on debates about the nature of practical reason, where it dissolves an otherwise difficult puzzle about our reasons to care about the rights and needs of others (Setiya 2013). In this essay, I turn instead to the epistemological consequences of Murdoch's view.

Despite Murdoch's attention to the inner life, her intimations of its privacy and idiosyncrasy have not been much discussed.[5] How private is our life with concepts and what is the significance of this fact? There are two sides to this.

One issue is about convergence or its absence. Murdoch writes about the "specialized personal *use* of a concept" (Murdoch 1970: 25). And in dismissing the genetic theory, she suggests not only that our grasp of concepts may be partial even when we master them by conventional standards, but that the journey towards greater conceptual grasp involves divergence from one another:

[4] Perhaps the closest cousin is Ichikawa and Jarvis 2013.

[5] Wiseman 2020 is a recent exception.

> We do not simply, through being rational and knowing ordinary language, 'know' the meaning of all necessary moral words. We may have to learn the meaning; and since we are human historical individuals the movement of understanding is onward into increasing privacy, in the direction of the ideal limit, and not back towards a genesis in the rulings of an impersonal public language. (Murdoch 1970: 28)[6]

How radical is Murdoch's doctrine of privacy in passages like this?

The second issue is related to the first. It's about moral knowledge and how Murdoch can account for it. Since for Murdoch, "mental concepts enter the sphere of morality," it is also about our knowledge of other minds (Murdoch 1970: 23-4).

I'll explore the tension between Murdoch's remarks on idiosyncrasy and her conception of love. I'll explain how her focus on "secondary" moral concepts—including concepts of mental life—speaks to problems in moral epistemology. And I'll raise puzzles about the extension of this account to "primary" moral concepts, such as the concept *good*. I'll end by asking how far Murdoch's insights can be extricated from the Platonic theory.

I

Murdoch associates Radical Internalism, perhaps surprisingly, with freedom. For Murdoch, freedom lies not in the absence of constraint, the liberty of indifference, but in responding to reality: "Freedom is not strictly the exercise of the will," she writes, "but rather the experience of accurate vision which, when this becomes appropriate, occasions action" (Murdoch 1970: 65). Instead of disputing the value of freedom, Murdoch reconceives it: having no alternative is perfectly compatible with freedom if the

[6] See also Murdoch 1956: 78 on "the notion of a private or personal vision."

compulsion under which one acts is the compulsion of reasons, full grasp of the facts in light of which one ought to act as one does.[7]

The idea that freedom is, or is compatible with, determination by reason may seem perplexing. Does it portray the virtuous person as a rational automaton, a machine whose behaviour is dictated by the world?[8] Murdoch responds, at least in part, by reminding us of the work required to reach this point, the exertion of will in thinking how to describe one's situation justly, refining one's grasp of the concepts through which it is given: "Man is not a combination of an impersonal rational thinker and a personal will. He is a unified being who sees, and who desires in accordance with what he sees, and who has some continual slight control over the direction and focus of his vision" (Murdoch 1970: 39). The compulsion with which we act when we grasp the reasons for acting as we do is the product of arduous and uncompelled attention to the world. It does not bypass or subvert our agency.

But are there hints of something more? On the Platonic theory, each concept is associated with standards of proper use, both practical and theoretical. It is natural to assume that there is just one way to meet these norms, one way of applying a concept and responding to its use, that is ideal or perfect. But we need not take that view. We could say instead that there are many ways to meet the norms, as there may be many portraits of a single person, substantially different from one another, no one of which is more accurate, or beautiful, than the rest. In other words, we could adopt a radically pluralist view:

> RADICAL PLURALISM: There are incompatible ways to apply a moral or mental concept, and respond to its use—incompatible modes of application and response—such that none is more ideal or perfect than any other.

[7] For a related claim, see Wolf 1990.

[8] This question is framed in Stern 2022, as is the response sketched in this paragraph.

For the Radical Pluralist, we are free to understand the world differently, even when our grasp of concepts is flawless. We are not rational automata whose behaviour is dictated by the world, even when we fully grasp the world and its description. For we may use moral and mental concepts in conflicting ways, none of which is better than the others. We have liberty even here.[9]

Does Murdoch accept this view? There are passages that tempt one to say yes. It is natural to see Radical Pluralism in Murdoch's claim that "the movement of understanding is onward into increasing privacy" (Murdoch 1970: 28). This fits well with what she says about M:

> M's activity is peculiarly her own. Its details are the details of this personality; and partly for this reason it may well be an activity which can only be performed privately. M could not do this thing in conversation with any other person. … [If] M says D is 'common' [this] can only be fully understood if we know not only D but M. … Language is far more idiosyncratic than has been admitted. Reasons are not necessarily and *qua* reasons public. (Murdoch 1970: 22, 32)

Moral language, she concludes, is "often unavoidably idiosyncratic and inaccessible" (Murdoch 1970: 33).

But in the end, I hesitate to read Murdoch as a Radical Pluralist. There are three points to make here: textual, philosophical, and interpretive.

The textual point is that the language of privacy and idiosyncrasy is absent from the last two essays in *Sovereignty*: it is downplayed, not developed further.

The philosophical point is that there is tension between Radical Pluralism and Murdoch's conception of love. "Love is knowledge of the individual," she writes; in loving them, we grasp "the separateness and differentness of other people" (Murdoch 1970: 27, 64). According to an earlier

[9] In "The Darkness of Practical Reason," Murdoch associates the work of the will in shaping our concepts with fabrication: "Each of us lives and chooses within a private, partly fabricated world … When moments of decision arrive we see and are attracted by the world we have already (partly) made." (Murdoch 1966: 199, 200)

formulation: “Love is the extremely difficult realization that something other than oneself is real” (Murdoch 1959: 215). But to love someone is not simply to acknowledge their reality in the abstract; it is to attend to their particularity. Love is intimate knowledge of an individual, knowledge of the sort that orients us towards them. This will involve knowing their inner life, since facts about their thoughts and feelings give us reasons. To the extent that one’s knowledge of these reasons is imperfect, one’s love for another is imperfect, too.

It follows that, unless perfect love is in principle impossible, we must be able to describe the thoughts of others—“M believes that D is spontaneous, gay, and delightfully youthful”—understanding the concepts that figure in their beliefs in the same way they do. Radical Pluralism would preclude that: if I understand the concept of spontaneity differently from M, applying and responding to it in different ways, I can’t ascribe to her the thoughts she actually has, understood in the way she understands them. I might perhaps describe her understanding from the outside, explaining how she operates with the concept; but I cannot share her understanding. This may be the best we can do for the less-than-perfect, whose flawed understanding prevents us from loving them perfectly. The problem is that, for Radical Pluralists, perfect love may be impossible even for those with perfect grasp of moral and mental concepts. That can’t be right.

The interpretive point is that an alternative reading of Murdoch is available. It is true that she emphasizes idiosyncrasy in our grasp of concepts, and that she insists on the difficulty of love, even to the point of “an inevitable imperfection … an ideal limit to love or knowledge which always recedes” (Murdoch 1970: 27). But her view may be that we take idiosyncratic routes towards the same “infinitely distant standard” (Murdoch 1970: 30). There is one ideal way to apply a given moral or mental concept and respond to its use, but since none of us can ever reach it, even the best of us will only approximate this standard in differently imperfect ways. This fact is an obstacle to love, since it is an obstacle to perfect grasp of concepts. But it doesn’t follow that if, impossibly, we were to reach that perfect grasp, to know reality in full, our love would remain imperfect.

As evidence for this interpretation, we can cite a remark about “concrete universals”:

> My view might be put by saying: moral terms must be treated as concrete universals. And if someone at this point were to say, well, why stop at moral concepts, why not claim that all universals are concrete, I would reply, why not indeed? Why not consider red as an ideal endpoint, as a concept infinitely to be learned, as an individual object of love? A painter might say, 'You don't know what "red" means.' (Murdoch 1970: 28-9)

Murdoch is drawing here on a way of thinking due to F. H. Bradley and the British Idealists, on which full grasp of a universal would involve acquaintance with every object that falls under it.[10] To fully grasp the concept *red*, one would have to be acquainted with every red thing in its concrete particularity. The reason why full grasp of concepts is impossible, on this view, is that the extension of a given concept is never fixed: there are endlessly more possible individuals that fall under it, and one cannot be acquainted with them all. This is what Murdoch has in mind when she speaks of "the indefinability of good": "Good is indefinable not for the reasons offered by Moore's successors, but because of the infinite difficulty of the task of apprehending a magnetic but inexhaustible reality. … The indefinability of Good is connected with the unsystematic and inexhaustible variety of the world" (Murdoch 1970: 41, 96).[11] She would say something similar about moral and mental concepts like the ones invoked by M. If they stand for concrete universals, our grasp of them at any moment will be idiosyncratic, since we are acquainted with different instances, even though we would converge at the ideal, impossible limit.

How does this idea fit with the other elements of Murdoch's view? It is distinct from, but compatible with, the Platonic theory of concepts. The thought would be that perfect grasp of the standards of application and use for a concept would involve acquaintance with everything that falls under it. Radical Internalism would remain intact—a consequence of the Platonic theory—though it is

[10] On the relationship between Murdoch, Bradley, and Hegel on "concrete universals," see Mac Cumhaill 2020: 237-8.

[11] See also Murdoch (1956: 87) on "the inexhaustible detail of the world, the endlessness of the task of understanding, the importance of not assuming that one has got individuals and situations 'taped,' the connection of knowledge with love and of spiritual insight with apprehension of the unique."

not clear that its conditions would ever be met. In realistic cases, our grasp of concepts would always be partial. Still, as our grasp of a concept improves, we would come closer to applying it, and responding to its application, properly—which means, *inter alia*, being moved by the facts in which it figures in proportion to their weight as reasons. "The more the separateness and differentness of other people is realized, and the fact seen that another man has needs and wishes as demanding as one's own," Murdoch writes, "the harder it becomes to treat a person as a thing" (Murdoch 1970: 64). Finally, there would be a single state of perfect conceptual grasp, free of idiosyncrasy, albeit one we never achieve. The upshot is that we need not read Murdoch as a Radical Pluralist: the idiosyncrasy in our grasp of concepts may be both permanent and provisional.

II

That we can in principle share moral and mental concepts does not mean that mutual misunderstanding is rare; for Murdoch, it is pervasive. In "Vision and Choice in Morality," she writes that we differ morally "not only because we select different objects out of the same world but because we see different worlds … it is surely true that we cannot always understand other people's moral concepts" (Murdoch 1956: 82). But this must be handled with care. For there are several things to mean by "mutual understanding": partial possession of the very same concept; full or perfect grasp of how another person grasps the concept; and the disposition to use the concept in the very same way. The last two may be vanishingly rare. But the first is commonplace: when our grasp of a concept like *love* or *repentance* differs, it's the same concept we grasp by approximating, differently, the standards for its proper use.

Part of the interest of Murdoch's Radical Internalism is that it shifts the burden of ethical thought dramatically from so-called "thin" or "primary" concepts such as *reason*, *good*, and *right*—the ones that come first in lists of moral concepts drawn up by philosophers—on to "secondary" moral concepts, including mental concepts, that figure in the reasons that determine what we should do. At one point, Murdoch frames this consequence starkly:

> On my view it might be said that … the primary moral words could be dispensed with entirely and all moral work could be done by the secondary specialized words. If we picture the agent as compelled by obedience to the reality he can see, he will not be saying 'This is right', i.e., 'I choose to do this', he will be saying 'This is A B C D' (normative-descriptive words), and action will follow naturally. As the empty choice will not occur the empty word will not be needed. (Murdoch 1970: 40-1)

Murdoch's point here is easy to misinterpret. The idea of "normative-descriptive words" may call to mind "thick moral concepts" such as "*coward, lie, brutality, gratitude*" in which description and evaluation are bound together (Williams 1985: 140). But as we've seen throughout, Murdoch's claims are not limited to thick concepts. They apply more generally to the concepts with which we describe our circumstance and the people with whom we interact, concepts that figure in the reasons for our responses—concepts such as *need*, *repentance*, *love*. If Murdoch calls them "moral" that is because they are subject to Radical Internalism: the thoughts in which they figure are internally related to the will. The knowledge that constitutes virtue, for Murdoch, is not knowledge of the good but of the real existence of other people and the world we share.

Suppose, then, that we focus on Murdoch's "secondary specialized words" and the concepts they express, many of which we share but understand differently and to differing degrees. These are the concepts with which we describe the circumstance in which we find ourselves: our social situation and the lives of those around us. If we do so justly and lovingly—meaning not just that we use the right concepts but that we understand them well enough—we will thereby act as we should.[12] We will not need to ask, explicitly, what reasons we have, or what it would be good to do, what actions are right or wrong: we do not need to know.

[12] See Murdoch 1970: 22: "What M is *ex hypothesi* attempting to do is not just to see D accurately but to see her justly or lovingly."

As I have argued elsewhere, the deepest problem in the epistemology of ethics is about knowledge as non-accidental truth. This idea is sometimes expressed in terms of "modal safety": when S knows *p,* S could not easily have been wrong, in that there is no close possible world in which S falsely believes *p*. But this condition is vacuous for necessary truths, including those in ethics, since there is no possible world in which the relevant proposition is false. When *p* is a necessary truth, the method of believing *p* could not easily lead one astray, even if the fact that one believes *p* has nothing to do with the fact that *p*. Hence the turn to an explanatory rather than modal interpretation of "accident":

> K: When S knows *p*, she knows it by a reliable method, and it is no accident that her method is reliable; there is an explanatory link between the fact that *m* is reliable and the fact that S uses *m*.[13]

The question is how ethical beliefs could count as knowledge, given K. If ethical facts are causally inert and constitutively independent of us, there is no room for an explanatory link between the methods by which we form our ethical beliefs and the fact that those methods are reliable. Ethical knowledge is impossible.[14] You might think the solution is to go reductionist: for any ethical concept, E, there are non-ethical concepts, N, with which we can say what it is to have the property picked out by E. If we can give causal explanations that appeal to N, ethical facts will not be causally inert. But despite appearances, this maneuver will not help. It remains true that the fact that being E is being N is causally inert, so we have no account of how such truths could ever be known.[15]

There are, however, theories that would solve the problem. Some are "anti-realist": they meet condition K by giving an account of ethical facts in terms of our beliefs. For instance:

---

[13] I argue for K, and explore its impact more extensively, in Setiya 2012: Ch. 3.

[14] Unless, perhaps, we appeal to God; I discuss this move in Setiya 2012: 114-5.

[15] I spell this argument out more fully in Setiya 2012: 112-3.

CONSTRUCTIVISM: For something to be good is for us to be disposed to think it is.

If this is true, it is no accident that we form beliefs about the good by a reliable method, since our reliability follows from the nature of ethical facts. Constructivism may appear too simplistic to take seriously, but there are more subtle views that share its spirit, such as Sharon Street's "Humean Constructivism," on which the facts about what there is reason to do are a function of our judgements about reasons, corrected for coherence of various kinds (Street 2008).

One need not think the facts are a function of our beliefs in order to explain how ethical knowledge is possible. The constitutive link could run the other way: our beliefs may be a function of the facts, as in externalist views of content.[16] According to the simplest such view:

EXTERNALISM: Part of what it is to have the concept *good* is to be such that one's method for identifying things as good is sufficiently reliable.

Again, if this is true, it is no accident that we form beliefs about the good by a reliable method, since our reliability follows from the nature of those beliefs.

The problem with Constructivism and Externalism, at least in these simple forms, is that they predict less disagreement than there seems to be: anyone who has the concept *good* is sufficiently reliable about the good. But there seem to be individuals whose beliefs about the good are wildly off the mark. If Murdoch is right, however, we need not confront this problem, focusing instead on secondary ethical concepts. On the Platonic theory, grasp of a concept depends on approximating the standards for its proper use: one must be sufficiently reliable in applying the concept and responding to its application. The bar for sufficient reliability must be low enough to accommodate those who fail to act as they should; but it could vindicate Externalism about such concepts. To grasp a secondary ethical concept, one must be sufficiently reliable; but there is room for improved conceptual grasp.

[16] Although it may seem unfamiliar, versions of this thought have been pervasive in realist meta-ethics; see Boyd 1988, Jackson 1998: Chs. 5-6, Brink 2001.

A critic may object that the application of certain secondary concepts—for instance, virtue concepts such as *justice*—elicits disagreement almost as deep as that occasioned by *reason*, *good,* and *right*. Externalism falters here, as well. But we can solve the problem by going ever more secondary, dispensing with *justice* and other virtue concepts in favour of concepts that describe the facts on which the virtues of an action supervene. Eventually, we will reach concepts not traditionally seen as ethical: the social and psychological facts that make our actions right or wrong, just or unjust, courageous or cruel, facts about the particular situations in which we act and the effects of our actions on others. Here disagreement is far less pronounced and Externalism more defensible. Our reliability about such facts, while often limited, is not an accident; nor is the truth of our beliefs about them, when it flows from the methods involved in the proper use of the relevant concepts, or approximation of these methods.

If Murdoch is right, we can thus have the knowledge we need in order to act as we should: knowledge of the social and psychological facts that ultimately give us reasons, falling under Radical Internalism. We need not face the question, how we know what is a reason, what is good, what is right or wrong, just or unjust—since we do not need to know.

III

And yet we seem to have beliefs about these subjects. We make purported claims about what is a reason to act, what is good or bad, the principles of justice, right and wrong. Murdoch's theory itself relies on primary moral concepts: it is about how we *should* think and act.

It would be awkward, at best, to throw this ladder away, concluding that the right and the good are ineffable. But it is not unprecedented: the move is reminiscent of Wittgenstein's (1929/1965) "Lecture on Ethics." And it would give sense to some of Murdoch's more enigmatic views: that the "transcendence of the good" places it beyond experience, that "Goodness is not in this world" (Murdoch 1970: 58, 59). It's as though we can never quite say what we mean to say when we talk about "the good," that we fail to state plain facts, not because we are merely expressing our will—as the non-

cognitivist might say—but because the good lies ever beyond our conceptual grasp. There is a strain of mysticism in *The Sovereignty of Good*.

In the end, though, Murdoch does not seem to take this line. "It would however be far from my intention to demote or dispense with the term 'good,'" she writes, "but rather to restore to it the dignity and authority which it possessed before Moore appeared on the scene" (Murdoch 1970: 41). The question is how we can have this concept, or be capable of knowing what is good, in light of radical disagreement: the apparent existence of individuals whose beliefs about the good are substantially unreliable. Such individuals threaten the truth of Externalism—on which one cannot possess the concept *good* unless one is sufficiently reliable in identifying what is good—making our reliability about the good an accident. Knowledge of the good becomes impossible.

There are sophisticated versions of both Constructivism and Externalism that allow more room for disagreement. To focus on Externalism, we might tie concept-possession not to individual reliability but to one's linguistic or conceptual community:

> SOCIAL EXTERNALISM: Part of what it is to have the concept *good* is to belong to a linguistic or conceptual community whose method for identifying things as good is sufficiently reliable.

If the community uses method *m* and I use *m* because the community does, it is again no accident that my method is reliable. There is an explanation of why I use it that entails as much.

But apart from going against the spirit of Murdoch's views, Social Externalism cannot make room for whole societies to go astray. If there are individuals whose beliefs about the good are substantially unreliable, why not social groups or civilizations? More likely, Murdoch would deny appearances: however it may seem, nobody who grasps the concept *good* is less than sufficiently reliable with it—they must approximate the standards for its proper use, both practical and theoretical. Those who appear to use the concept in radically divergent ways mean something else by "good," if they mean anything at all.

This response is not altogether *ad hoc*. It is consonant with one of Murdoch's more resounding aphorisms: "Man is a creature who makes pictures of himself and then comes to resemble the picture" (Murdoch 1957: 75). The picture she associates with existentialism and emotivism is one on which the use of "good" becomes "a function of the will" (Murdoch 1970: 4).

> Moore was quite right (it was said) to separate the question 'What does "good" mean?' from the question 'What things are good?' though he was wrong to answer the second question as well as the first. He was right to say that good was indefinable, but wrong to say that it was the name of a quality. Good is indefinable because judgments of value depend upon the will and choice of the individual. (Murdoch 1970: 3)

It is not absurd to fear that where this picture is pervasive, it remakes us in its image: we come to use the primary moral word "good" as an "empty action word which is the correlate of the isolated will," not for the ideal limit of perfection to which conceptual grasp aspires (Murdoch 1970: 8). If this is the predicament of those who seem to disagree with us in radical ways, we can plausibly deny that they mean what we do by "good." For them, it is a term of approbation or prescription unconstrained by any facts; for us, it maps the normative structure of mind and world. Understood in this way, our "disagreement" with emotivists or existentialists is consistent with Externalism: those who grasp the concept we express by "good"—as they do not—must be sufficiently reliable. Our reliability is no accident; nor do we disagree with them, since they have no idea what we mean.[17]

But the story cannot end here. It is one thing to dismiss emotivists and existentialists as having the lost the concept *good*. It's another to make sense of our apparent disagreement with moral realists whose professed beliefs about the good are shocking. Imagine, if you will, a Calliclean society, inspired by Callicles' "great speech" in Plato's *Gorgias*. "What you call 'justice' is no virtue," they will say. "In the condition of 'natural justice' the powerful dominate the weak, who acquiesce in the justice of their

[17] For a related thought, see MacIntyre's "disquieting suggestion" (MacIntyre 1981: Ch. 1).

subordination. This is what virtue looks like in cold, hard truth. Given Radical Internalism, to use concepts as one should is, among other things, to conform to natural justice. You are simply wrong about what is good." On the face of it, we disagree with the Callicleans. What they say is false: "natural justice" is not an expression of virtue but of vice, part of a whole system of virtues, and a conception of the good, from which we recoil. The problem is that, if the Callicleans share our concepts, then Externalism—even Social Externalism—cannot be true. For they are not reliable about the good. How then, can our reliability be more than accidental? How can we know what is good?

To preserve the possibility of knowledge, one would have to say about the Callicleans what I said Murdoch should say about the society of emotivists or existentialists: they simply do not have the concept *good*; whatever they mean by "good," it isn't that. After the passage quoted above, in which she details "how severely and in what respects Moore was corrected by his successors," Murdoch writes that "on almost every point I agree with Moore and not with his critics" (Murdoch 1970: 3-4). A point on which she disagrees with both Moore *and* his critics, I submit, is the very first: that we can "separate the question 'What does "good" mean?' from the question 'What things are good?'" (Murdoch 1970: 3). That we cannot do so is the lesson of our Calliclean rivals. Their view about what things are "good" is so far off the mark that they cannot mean by "good" what we do.

The question that remains is what they do mean, if their words have meaning at all. For the emotivists and existentialists, "good" is meaningful, but not descriptive: it is used to express approbation or to issue prescriptions. But the Callicleans do not seem to use "good" in this way. They do not treat it as an "empty action word which is the correlate of the isolated will" (Murdoch 1970: 8). They use it, so they say, to trace the standards for the ideal use of concepts—though presumably, by parity of reasoning, they do not mean by "ideal" what we do. Does the Callicleans' use of "good" describe the world, not in terms of what is good, but through some quasi-evaluative concept to which we have no access, one defined by standards for proper use (in our sense) that determine where it should be applied and what follows from its application for cognition and the will? If so, there are quasi-evaluative facts of which we are ignorant, perfect grasp of which would lead us to act as the Callicleans do. But surely there are no such facts.

The best way forward here is to deny that the Callicleans mean anything by "good." Their word expresses a "defective concept," full possession of which would involve reacting or thinking in ways that we should not—which is to say that "good," used by the Callicleans', does not express a genuine concept, after all. Since the norms definitive of any concept are norms for thinking and responding as we should, real concepts cannot be in this way flawed.[18] The Callicleans may appear to be thinking and asserting things when they profess to disagree with us; but they are not.

IV

At this point, I step back from exegesis to ask: how plausible is this? Not especially, I fear. We have been led by Murdoch's view of concepts—assuming it is meant to ground the epistemology not just of secondary but of primary moral knowledge—to the claim that we talk past the Callicleans, since their words do not make sense. On the contrary, I am pretty sure we disagree with them.

The Callicleans are fictional. But I don't see what prevents more realistic instances of disagreement about virtue and the good life, across the span of human history, too radical to reconcile with Externalism. How to make sense of ethical knowledge in light of this is a puzzle—but that is the puzzle we face. It would take work to spell the difficulties out in detail, and more work to resolve them.[19] Instead of attempting that here, I want to close by considering, in brief, what we can learn from Murdoch if we reject the Platonic theory of concepts, Radical Internalism, and the account of ethical knowledge I have sketched on her behalf.

What we learn above all is that "secondary specialized words" play a central role in ethical thought. Even if mental concepts do not "enter the sphere of morality" because "true vision occasions right conduct," it remains the case that, in our relationships with others and ourselves, finding the

[18] For a similar argument, see Wedgwood 2007: 168-9; and for an opposing view, Williamson 2003.

[19] I attempt to do this in Setiya 2012: Chs. 3-4. As I argue in Setiya 2012: 134-8 and at greater length in Setiya 2024, the resolution leads to a form of Aristotelian relativism inconsistent with Murdoch's more Platonic view.

right description of our circumstance lies at the heart of moral reflection. Only by doing so can we discern the reasons that determine how to act and what to feel. It doesn't matter whether perfect grasp of reasons is inevitably motivating, as the Radical Internalist believes. What matters is that the reasons we actually have look more like descriptions in novels, or M's description of D, than those that figure in schematic cases, or principles, framed by moral theorists.

Murdoch frequently puts her insight in terms of vision: "I can only choose within the world I can *see*, in the moral sense of 'see' which implies that clear vision is the result of moral imagination and moral effort. … One is often compelled almost automatically by what one *can* see" (Murdoch 1970: 35-6). The second sentence hints at Radical Internalism. But it need not be read that way, and even if it is, the previous sentence does not rest on or entail it. One can agree with Murdoch that "goodness is connected with knowledge … a refined and honest perception of what is really the case" without believing that, if one fully grasps the facts that provide one with reasons one is bound to act on them (Murdoch 1970: 37). The hard-won knowledge of one's circumstance "which is the result not simply of opening one's eyes but of a certain perfectly familiar kind of moral discipline" may be necessary, not sufficient, for acting well (Murdoch 1970: 37).

It is probably not an accident that Murdoch focuses, in *Sovereignty*, on personal interaction, not public policy. The centrality of sustained, attentive description is most salient in the ethics of close relationships, where the insights we gain may be private or idiosyncratic—difficult to communicate or share with strangers. But Murdoch's insight extends to social-political thought. One need not be a Radical Internalist to think that language matters here, too: that the words we deploy when we describe our social world affect what sorts of reasons we discern, and so how we respond. Do we think in terms of structural injustice, exploitation, ideology, public reason? Are these concepts alien to us? Do we dismiss them—or some of them—as unhelpful? Are our concepts adequate to the political reality we confront? What these questions press us to acknowledge is the need for philosophical imagination, as well as argument, in ethics and in politics. The lines between philosophy and imaginative literature begin to blur. Thus Murdoch writes in "Vision and Choice": "From here we may see that the task of moral philosophers has been to extend, as poets may extend, the limits of language, and enable it to

illuminate regions … which were formerly dark" (Murdoch 1956: 90). One need not accept the Platonic theory of concepts, or Murdoch's Radical Internalism, in order to accept this task.[20]

References

Boyd, R. (1988). How to be a Moral Realist. In G. Sayre-McCord, ed., *Essays on Moral Realism*. Ithaca, NY: Cornell University Press, 1988, pp. 181–228.

Brink, D. (2001). Realism, Naturalism, and Moral Semantics. *Social Philosophy and Policy*, 18: 154-76.

Ichikawa, J. and Jarvis, B. (2013). *The Rules of Thought*, Oxford: Oxford University Press.

Jackson, F. (1998). *From Metaphysics to Ethics*, Oxford: Oxford University Press.

Mac Cumhaill, C. (2020). Getting the Measure of Murdoch's Good. *European Journal of Philosophy*, 28, 235-47.

MacIntyre, A. (1985). *After Virtue: A Study in Moral Theory*, Notre Dame, IN: University of Notre Dame Press.

McDowell, J. (1978). Are Moral Requirements Hypothetical Imperatives? Reprinted in *Mind, Value, and Reality*. Cambridge, MA: Harvard University Press, 1998, pp. 77-94.

Murdoch, I. (1956). Vision and Choice in Morality. Reprinted in G. Steiner, ed., *Existentialists and Mystics*. London: Chatto and Windus, 1997, pp. 76-98.

Murdoch, I. (1957). Metaphysics and Ethics. Reprinted in G. Steiner, ed., *Existentialists and Mystics*. London: Chatto and Windus, 1997, pp. 59-75.

Murdoch, I. (1959). The Sublime and the Good. Reprinted in G. Steiner, ed., *Existentialists and Mystics*. London: Chatto and Windus, 1997, pp. 205-20.

Murdoch, I. (1966). The Darkness of Practical Reason. Reprinted in G. Steiner, ed., *Existentialists and Mystics*. London: Chatto and Windus, 1997, pp. 193-202.

Murdoch, I. (1970). *The Sovereignty of Good*, London: Routledge.

Setiya, K. (2012). *Knowing Right From Wrong*, Oxford: Oxford University Press.

Setiya, K. (2013). Murdoch on the Sovereignty of Good. *Philosophers' Imprint*, 13, 1-21.

Setiya, K. (2024). Human Nature, History, and the Limits of Critique. *European Journal of Philosophy*, 32, 3-16.

[20] I am grateful to an audience at the Eastern Division Meeting of the APA in Montreal; special thanks to Youngbin Yoon for written comments on an earlier draft.

Stern R. (2022). 'How is Human Freedom Compatible with the Authority of the Good?': Iris Murdoch on Moral Agency, Freedom, and Imagination. *Proceedings of the Aristotelian Society*, 122, 1-26.

Street, S. (2008). Constructivism about Reasons. *Oxford Studies in Metaethics*, 3, 207-45.

Wedgwood, R. (2007). *The Nature of Normativity*, Oxford: Oxford University Press.

Williams, B. (1985). *Ethics and the Limits of Philosophy*, Cambridge, MA: Harvard University Press.

Williamson, T. (2003). Understanding and Inference. *Proceedings of the Aristotelian Society, Supplementary Volume*, 77, 249-93.

Wiseman, R. (2020). What if the Private Linguist Were a Poet? Iris Murdoch on Privacy and Ethics. *European Journal of Philosophy*, 28, 224-34.

Wittgenstein, L. (1929/1965). A Lecture on Ethics. *Philosophical Review*, 74, 3-12.

Wolf, S. (1990). *Freedom Within Reason*, Oxford: Oxford University Press.