Philosophers' Imprint

volume 13, no. 9
may 2013

Murdoch on the Sovereignty of Good

Kieran Setiya

University of Pittsburgh

Though she is as well-known as any figure in British philosophy since 1950, Iris Murdoch's impact on the philosophical fields in which she wrote is hard to make out. She is cited as an influence by Cora Diamond, John McDowell, Hilary Putnam, Charles Taylor, Bernard Williams, and Susan Wolf.[1] But she is rarely mentioned in recent work. There must be many moral philosophers who have never read, or hardly thought about, her best and most systematic essays, in *The Sovereignty of Good*. The reasons for this neglect are various. Murdoch left her academic position at Oxford in 1963. She stopped publishing in philosophy journals. And she produced fewer essays of any kind, for the most part writing novels instead.

Along with these sociological facts, there are difficulties internal to Murdoch's work. Her writing can be opaque, her views obscure. It is not easy to identify arguments, if she has them, or clear objections to opposing views. And where we do find objections and arguments, their targets often seem dated or irrelevant. "The Idea of Perfection" is framed as a critique of Stuart Hampshire's *Thought and Action*, which these days is surely read and cited less than Murdoch herself.[2] Ryle's alleged behaviourism is in the background: no longer a live concern. In moral theory, Murdoch's interlocutors are R.M. Hare, whose influence has also waned, along with caricatures of existentialism and of Kant. The result is that Murdoch is not cited or discussed by most contemporary work in moral psychology, a topic whose current incarnation owes much to her.[3] She is, if anything, less visible in moral epistemology, in the metaphysics of morals, and in the study of practical reason.

If Murdoch is to speak more audibly to contemporary philosophers, so that she cannot be ignored, her ideas must be reframed as interventions in existing disputes, her arguments must be recovered,

1. See Diamond 1995; McDowell 1979; Putnam 2002; Taylor 1989; Williams 1985; Wolf 1990.
2. Originally published in 1964, "The Idea of Perfection" is the first essay in *The Sovereignty of Good* (Murdoch 1970). The others are "On 'God' and 'Good'" and "The Sovereignty of Good Over Other Concepts". I will cite these essays as they appear in the book.
3. On Murdoch and moral psychology, see Broackes 2012: 36n.77.

and her conclusions made clear. With notable exceptions, few have taken up this task; there is a lot to be done. In this essay, I try to make progress here, to bring out the force and cogency of Murdoch's thinking by relating it to an issue that is certainly live: that of ethical rationalism and the question, "Why be moral?" My aim is not to evaluate Murdoch's views but to make them plausible, intelligible, and, so far as possible, argumentatively sharp. What emerges is, I believe, a conception of great originality and theoretical power. If it is wrong, it is wrong in interesting ways; to say what they are lies outside my present scope.

This essay has four main parts. In the first, I introduce the idea of ethical rationalism and relate it to the problem of moral reasons. In the second, I set out Murdoch's striking claim that "true vision occasions right conduct" and explain its importance for the rationalist debate (Murdoch 1970: 64). Though its truth would be significant, Murdoch's claim is problematic. It seems possible, off-hand, for someone to share the vision that "occasions right conduct" without acting well. In section three, I show how Murdoch anticipates this charge by tracing it to a defective theory of concepts. That the theory goes wrong is the moral of her most famous example, that of M and D.[4] I go on to explore the entanglement of fact and value in Murdoch's view, how it differs from the appeal to "thick moral concepts", and how it draws on a Platonic theory of concept-possession. Section four then turns to the "image of the Good as a transcendent magnetic centre" (Murdoch 1970: 73), working through the odd mix of empirical psychology, moral exhortation, and speculative metaphysics in *The Sovereignty of Good*. Murdoch's theory of concepts, introduced in the first essay of *Sovereignty* and developed in the second, provides the basis for a moral argument whose conclusion is metaphysical: that the Good is not, after all, illusory. If this is right, Murdoch's work has implications not only for practical reason and moral metaphysics but for the scope and ambition of moral philosophy.

4. Murdoch 1970: 16–23.

1. How to Be an Ethical Rationalist

In my perhaps unconventional usage, ethical rationalism is the project of deriving standards of practical reason from the nature of agency or the will, as the capacity to act for reasons.[5] This project is ethical in the widest sense. The rationalist need not make claims for morality as a source of reasons or for the rationality of concern for others. But his subject is practical: how we should live and act. His treatment of this subject is a form of rationalism in that it generalizes a familiar reading of Kant in *Groundwork* III.[6] On this reading, it belongs to agents, as such, to act "under the idea of freedom" and so to be responsive to the moral law. Not that every agent does respond, but the propensity to do so is contained in the capacity to act. Its realization is thus an aspect of practical rationality. The generalization of this approach leaves room for species of rationalism on which we act intentionally not under the idea of freedom but under the guise of the good,[7] or on which we aim at self-knowledge or the satisfaction of desire, so that it is the object of practical reason to achieve the good, to gain self-understanding, or to get whatever you want.[8] In an alternative idiom, ethical rationalism is "constitutivism" about practical reason, where the constitutive view need not be Kantian in result. The standards that flow from the nature of agency may be ones of self-interest or means-end efficiency, not moral duty.

Whatever its specific content, the rationalist project moves from metaphysical premises to normative conclusions, from the metaphysics of agency to the norms of practical reason. A central task for ethical rationalists is to explain this transition. How do we get from the *is* of

5. This section draws on Setiya 2012: 15–21 and on Setiya forthcoming; its framework derives from Setiya 2007.

6. Kant 1785.

7. In that our action strikes us as good in some respect; for 'the guise of the good' see Velleman 1992.

8. See *Reasons without Rationalism* (Setiya 2007: 14–5), citing Korsgaard 1996; Railton 1997; Velleman 1989, 2000; Williams 1979; Dreier 1997. To this list, one might add Korsgaard 2009; Raz 1999, 2011; Vogler 2002; and, on one interpretation, Smith 2009, 2010.

what it is to be an agent, to have a will or the capacity to act for reasons, to the *ought* of practical rationality?

In one of the most explicit developments of ethical rationalism, J. David Velleman presents intentional action as having a "constitutive aim", which is self-knowledge or self-understanding.[9] More recently, Christine Korsgaard describes the aim of agency as "self-constitution".[10] In working out the basis of the rationalist approach, we can generalize these claims. On each account, agency is, in effect, a functional or teleological kind, defined by an end or goal. This schema can apply to other views. It might be in the nature of agents to aim at perfect autonomy, at means-end coherence, or at the highest good. Once we think of agency in these terms, we can trace the pivotal move of the ethical rationalist to a function argument inspired by Aristotle. Recall that, for Aristotle, human beings have a defining function or activity, which is the use of reason, and whatever has a function finds its good in performing that function well. There are standard objections. Is it right to speak of a human function? Does the argument conflate what is good *for* an F with being good *as* an F? Even if they are sound, however, these questions do not touch the principle we need.

> Excellence: When Fs have a defining function or activity, a good F is one that performs that activity or function well.

If what it is to be an agent is to aim at autonomy, or self-knowledge, or whatever, what it is to be good as an agent is to aim at these things effectively. Since practical reason is the virtue of agency, as such, one is to that extent practically rational. The rationalist account goes through.

We can spell out the demands of this argument more slowly. First, it relies on a certain structure in the function of agency: not just that agents are defined by an activity — doing things for reasons — but that agency has a target, like happiness or means-end coherence, of which it can fall short. It belongs to the nature of agents to be directed by, or tend towards, an aim that they may not realize even when they succeed in acting for reasons.[11] The achievement of this aim does not sort behaviour as intentional or not: it offers an ideal to which intentional action may or may not conform. It is not a condition of acting for reasons that one hit the target in question, only that one aim at doing so. This structure allows for defective action in the framework of ethical rationalism: the capacity to act for reasons can be exercised imperfectly, in ways that do not fully achieve its end.[12]

The idea of an aim or function that figures in the rationalist argument need not be mysterious. Since rationalists hope for standards of practical reason that apply to agents, as such, regardless of their biological form, they do not appeal to specifically biological function.[13] Nor do they appeal to a particular intention or desire. Instead they turn to dispositions that constitute agency or to rules and principles that guide us if we act for reasons at all. There is room to be flexible here. If we think of the aim as fixed by a disposition of every possible agent, we end up with what is called "internalism about reasons": agents have the capacity to be moved by any reason to which they are subject. (More on this below.) Alternatively, to be an agent is to *approximate* the possession of dispositions whose target is thereby constituted as the aim of agents, as such.[14] When agents fall short of full possession, internalism fails. Either way, one can manifest the dispositions that constitute agency, to the extent that one has them, either imperfectly or in full. Intentional action is the product of such dispositions, which set a target for agents to meet in what they do.

The second assumption of the rationalist argument is what I have called "Excellence". It is worth stressing how modest this principle is. Not only does it not require the more contentious elements of

9. See, especially, the introduction to Velleman 2000.

10. In Korsgaard 2009.

11. For this requirement, framed as an objection to Velleman, see Clark 2001: 581–85.

12. On the apparent difficulty here, see Railton 1997: §3; Korsgaard 2009: Ch. 8.

13. Smith 2010: 124–5.

14. The role of approximation is emphasized in Dreier 1997: 89–91, 97–8, and Smith 2009: 66–9.

Aristotle's function argument, it does not purport to be a general account of *good*. That "good" has a functional use is quite consistent with its being used in other ways, too, as when we speak of "good outcomes", what is "good for" an individual, or even what is "good" *simpliciter*.[15] The applications of Excellence are innocent enough. If the function of clocks is to tell the time, a good clock does so both legibly and reliably. If the defining activity of a thief is to steal others' property, a good thief is one who gets away with the loot.

Putting the first two steps together: when the nature of a kind is defined in dispositional terms, and where it has a target of which it can fall short, to be good of that kind is to manifest that disposition in full. It is not enough for the application of Excellence that the disposition can fail to be exercised altogether: there must be such a thing as its imperfect or incomplete manifestation. Where an object meets these conditions, it can operate well or badly as the kind of thing it is. Thus, if being an agent is being disposed to a certain end, at least by approximation, and one can exercise this disposition, to the extent that one has it, more or less well, to be good as an agent is fully to achieve that end.[16]

In its final step, the argument identifies practical rationality — in the sense that involves, but is not exhausted by, responding to reasons — with being good *qua* agent. This premise draws on a compelling thought: that judgements of practical reason are assessments of agency, not some other aspect of our lives. To say this is not to presuppose the truth of rationalism; one can accept it even if one doubts that the nature of agency is a source of rational norms. If standards for being good *qua* agent do not flow from the nature of agency, as such, their grounds must lie elsewhere, perhaps not in the function of anything. Still, they are standards of practical reason.

Schematically, then, the argument for ethical rationalism takes this form: First, agency is defined by a function or activity, an aim or end to which it is directed, just as such — though it is not a condition of acting for reasons that one realize this aim. It follows by Excellence that to be good as an agent is to achieve the aim by which agency is defined. Practical rationality consists in being good *qua* agent. So to be practically rational is to achieve the aim of agency. Since it is a defect of practical reason to be unmoved by a reason to act, this conclusion will constrain what there is reason for us to do. If you can achieve the aim or end of agency while being indifferent to the fact that p, this fact is not a practical reason. Finally, if the aim is fixed by the dispositions of any possible agent, we can derive a more familiar constraint:

> Internalism about Reasons: If the fact that p is a reason for A to ϕ, A is capable of being moved to ϕ by the belief that p.

Since every agent is disposed to achieve the aim or end of agency, as such, and so to be practically rational, every agent has the capacity to respond to the reasons for her to act.

The interest of ethical rationalism, internalist or otherwise, lies primarily in two things. One appears to be a virtue. Ethical rationalists can assimilate the metaphysics and epistemology of practical reason to the functional use of "good". The idea of a reason to act is, on this conception, no more mysterious than that of a good thief or a good clock.[17] The other threatens to be a vice. By placing conditions on what there is reason for us to do that bind it to the nature of agency, ethical

15. Compare the discussions in Geach 1956 and Setiya 2007: Part Two, §2.

16. Does Excellence apply to the dispositions of objects that are not artifacts or living things? In principle, yes, but only when the conditions in the text are met. If a magnet is by nature disposed to orient itself in certain ways, and this disposition is operative when it orients itself in *roughly* the right way, it is functioning well as a magnet so far as it manifests its disposition to the highest degree. If this sounds odd, the problem is not with the application of Excellence but with the conception of magnets on which it rests. Magnets are defined by their intrinsic properties, not by dispositions that they manifest more or less well. If there are physical kinds that do have a suitable nature, it will make sense to evaluate their functioning — though doing so will not have implications for practical reason, as the function of agency does.

17. For metaphysical and epistemic arguments for ethical rationalism, or something like it, see Velleman 2000: 173–9 and Smith 2010.

rationalism creates a problem for the generality of moral reasons. It is this problem that provides the context for my interpretation of Iris Murdoch. I spend the rest of the section spelling it out.

Suppose, to begin with, that A is confronted with a circumstance in which someone is in serious need, and the right the thing to do is to offer them help. It would be morally wrong to ignore their difficulty, and the sacrifice involved in answering it is not extreme. We might be tempted to conclude that A *should* offer help, all things considered — that this is what he has most reason to do. At any rate, if moral reasons are categorical, if they apply to everyone, A has some reason to help the person he finds in need. Suppose, however, that A is completely unmoved by this. He knows what the circumstance is: that the need is serious and the cost is modest. (Let us set aside, for now, beliefs about right and wrong.) But this has no impact on him. If we hope to defend the conviction that A has reason to help, what can we say? In the rationalist framework, our options are limited. We could hope to show that A is practically irrational, that he cannot fully achieve the aim or end of agency, as such. But this amounts to a heroic task: that of deriving, from the metaphysics of agency, a sensitivity to moral concerns. We need an account of the aim of agency, as such, on which it cannot be met by an agent like A. The function of agency cannot be the satisfaction of desire, regardless of its object; it cannot be autonomy or flourishing, conceived in ways that allow for indifference to others. It must belong to the nature of the will, as the capacity to act for reasons, that its aims make one susceptible to others' needs. Some will accept this claim, hoping to derive from the constitutive aim of action some commitment to the moral law.[18] But most have doubts. There is a problem of moral reasons on the rationalist approach.

It might seem, so far, that the problem can be contained. We have focused on reasons that consist in the needs of others, asking whether one must respond to such needs in achieving the aim or end of agency, as such. In doing so, we may have ignored the more salient reason to help, that failing to do so is morally wrong. Alternatively, we might appeal to the fact that it would be good to help. It is this fact that justifies action, not the bare fact of human need. It is not clear, however, how much we gain by this move. There are three difficulties. First, how can the fact that someone is in need make it wrong not to help them, or good to do so, if this fact is not, by itself, a reason to help? Second, by turning to facts about right and wrong, or what is good, we forgo the metaphysical and epistemic virtues of ethical rationalism. These facts raise the usual puzzles: they cite aspects of practical normativity that have not been absorbed by the functional use of "good". Finally, even such loaded reasons place demands on the theory of action that will not be easy to meet. The picture of agents on which they act under the guise of good is historically influential. It derives from Plato and Aristotle, and it persists in more recent work.[19] But it is controversial, and I have argued against it, at length, elsewhere.[20] In short: while it may be true that representations of the good are essentially practical, we can say what it is to be an agent without them. The disposition to be moved by appearances of the good is not one we must possess or approximate in order to act for reasons at all. If this is right, it is not a defect of agency, and so not a defect of practical reason, in rationalist terms, to be indifferent to the good. The corresponding claims are even more dubious for moral right and wrong. Why should agency be impossible without specifically moral thought?

The upshot of these reflections is that ethical rationalism animates one version of the question, "Why be moral?" It provides the basis for a sceptical challenge to the generality of moral reasons. On the face of it, we can resist this challenge only by disputing the argument for rationalism or by deriving moral reasons from the nature of the will. Since the second task looks difficult, we may be tempted by the first. But this is equally hard, for what the argument assumes is weak. While many doubt that agents have autonomy or self-knowledge

18. Most notably, Korsgaard 1996, 2009.

19. For variations, see Railton 1997; Raz 1999, 2011; Tenenbaum 2007.

20. In Part One of *Reasons without Rationalism* (Setiya 2007) and more recently in Setiya 2010.

as an essential aim, let alone conformity with the moral law, most assume that they are disposed to take means to their ends: to be means-end coherent in their intentions or desires.[21] Even this claim is enough for the ethical rationalist. There are dispositions we must approximate in being agents, and they can be cast as dispositions to conform to an ideal. To be good *qua* agent is to meet this ideal and so to be practically rational.

The issues surrounding ethical rationalism are complicated, and I cannot hope to resolve them here.[22] My aim has been more modest: to show what is compelling about the rationalist view and to make the options with which it presents us sufficiently troubling that we are willing to look for another way out. We seem to confront a dilemma: Given the argument for ethical rationalism, deny the generality of moral reasons or derive them from the nature of the will. It is in search of a further option that we turn to Murdoch's book.

2. True Vision and Right Conduct

In reading *The Sovereignty of Good*, it will prove useful to start at the end of the first essay, not the beginning. What Murdoch works towards in "The Idea of Perfection" is an alternative to the picture of freedom on which the agent "chooses his reasons in terms of, and after surveying, the ordinary facts which lie open to everyone" (Murdoch 1970: 34). It is this picture that serves as common ground for her several antagonists — behaviourist, existentialist, and Kantian. And it is one that she rejects.

> I suggest [that] we introduce into the picture the idea of *attention*, or looking […] 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 place of choice is certainly a different one if we think in terms of a world which is *compulsively* present to the will, and the discernment and exploration of which is a slow business. […] If I attend properly, I will have no choices and this is the ultimate condition to be aimed at. (Murdoch 1970: 38)

In the stark formulation at the heart of the following essay, "realism […] is a kind of intellectual ability to perceive what is true, which is automatically at the same time a suppression of self […] true vision occasions right conduct" (Murdoch 1970: 64).

That Murdoch makes these striking claims is sometimes recognized. What is less well understood is why they matter. We can bring this out by finding the disputed conception of choice in the background of section 1. The picture of freedom as autonomy, not determined by the plain facts of one's circumstance, is shared by rationalists of different kinds, by instrumentalists and Kantians alike. The problem of moral reasons could be solved by giving it up. In raising this problem, we imagined someone who is fully aware of the circumstance that requires him to offer help, of serious need and modest cost, but who remains unmoved. The task that looked difficult, within the confines of ethical rationalism, was to explain why he *should* be moved, why these facts provide him with reasons to act. We must trace the necessity of being moved to the nature of the will. Murdoch's intervention is to find an alternative view. For Murdoch, the necessity of being moved by moral reasons lies not in the nature of the will but in the motivational import of cognition. Despite appearances, the agent we imagined is impossible. One cannot fail to be moved by an adequate conception of the facts that require a response: "true vision occasions right conduct". There is thus no need to enter the maze of options in which we got lost before.

21. This thought is explicit in Dreier 1997: 95–8, Vogler 2002: 135–46, and Smith 2009: 66–7.

22. In Setiya 2007, 2010, forthcoming, I defend a minimalist conception of agency on which the rationalist argument cannot begin.

According to a simple version of this idea:

> HYPER-INTERNALISM: If the fact that *p* is a reason for A to ϕ, and A knows that *p*, A is moved to ϕ in proportion to its strength as a reason.

Murdoch's final view is more complex. But the simple formulation brings out some crucial points. First, in the usual contrast between judgement and existence internalism, between claims about the motivational significance of thinking that there is reason to ϕ and motivational conditions on the existence of reasons, Murdoch's internalism falls on the side of existence.[23] It follows from something's *being* a reason to ϕ, according to the hyper-internalist, that if one knows the fact that is a reason, one is suitably moved. We need not add the further condition that one believes this fact to be a reason. Second, Hyper-Internalism is vastly stronger than Internalism about Reasons. It is not just the capacity to be moved but actual motivation that follows from knowledge of reasons. Finally, the truth of Hyper-Internalism does not tell us what there is reason to do. It does not imply that facts about the needs of other people provide us with reasons to act. Its role is not to support an argument for that claim but to prevent it from being threatened by ethical rationalism. If facts of this kind justify action, it follows that we are moved by the relevant beliefs. We need not derive their status as reasons, considerations by which we are moved insofar as we are rational, from the aim of agency, as such. Ironically, by tightening the connection between cognition and choice, we make this connection in one way easier to defend: it need not go through the metaphysics of the will.[24]

A similar thought can be found in an early paper by John McDowell, "Are Moral Requirements Hypothetical Imperatives?":

23. For this distinction, see Darwall 1983: 53–4.

24. This claim is qualified towards the end of section 3, where I speculate on Murdoch's conception of agency.

> To explain an action we regard as virtuous, we typically formulate a more or less complex characterization of the action's circumstances as we take the agent to have conceived them. Why should it not be the case […] that the agent's conception of the situation, properly understood, suffices to show us the favourable light in which his action appeared to him? […] There may seem to be a difficulty: might not another person have exactly the same conception of the circumstances, but see no reason to act as the virtuous person does? If so, adverting to that conception of the situation cannot, after all, suffice to show us the favourable light in which the virtuous person saw his action. […] We can evade this argument by denying its premise: by taking a special view of the virtuous person's conception of the circumstances, according to which it cannot be shared by someone who sees no reason to act as the virtuous person does. (McDowell 1978: 80)

The virtuous person's conception of the circumstance includes the reasons for which he acts — for instance, that someone is "shy and sensitive" and so needs to be put at ease (McDowell 1978: 85–6). It does not claim that they are reasons, or there would be nothing "special" about the view on which it cannot be shared by someone who "sees no reason to act as the virtuous person does". On a modest reading, the claim is merely that knowledge of the circumstance *can* motivate action without the need for a desire that is not itself explained, and made intelligible, by this knowledge.[25] But what is actually said is more ambitious: that "the relevant conceptions are not so much as possessed except by those whose wills are influenced appropriately" (McDowell 1978: 87). What we appear to have is then an instance of Hyper-Internalism:

25. On the motivation of action by belief alone, see Nagel 1970: 27–30; Wallace 1990: 373–4; Setiya 2007: 100–6.

> MORAL INTERNALISM: If the fact that p is a decisive *moral* reason for A to ϕ, and A knows that p, A is decisively moved to ϕ.[26]

Murdoch restricts the scope of Hyper-Internalism in a similar way:

> [We] are not always responding to the magnetic pull of the idea of perfection. Often, for instance when we pay our bills or perform other small everyday acts, we are just 'anybody' doing what is proper or making simple choices for ordinary public reasons. (Murdoch 1970: 41)

The contours of this restriction are not entirely clear. For simplicity, I will work with Hyper-Internalism for decisive moral reasons, stressing that the point is about the motivational force not of the belief that an act is right or wrong but of the facts that give it that status.[27]

As I have portrayed it, the attraction of Moral Internalism is to prevent the problem of moral reasons from being posed. We can save the generality and decisive force of such reasons without having to worry about the argument for ethical rationalism and without attempting the heroic derivation of moral commitment from the aim or end of agency, as such. Its only flaw, you might suspect, is that it is false — indeed, obviously so.[28] The special view on which the virtuous person's conception of the circumstance cannot be shared by the indifferent is too special to be believed. It is a fact of life that people are unmoved even by decisive moral reasons. As McDowell acknowledges, "[failure] to see what a circumstance means, in the loaded sense [which entails motivation] is of course compatible with competence, by all ordinary tests, with the language used to describe the circumstance" (McDowell 1978: 86). He does not seem perturbed by this, saying only that it "brings out how loaded the notion of meaning involved in the protest is" (McDowell 1978: 86). But many will be put off. If someone is competent by ordinary tests, they grasp the relevant concepts and know the relevant facts. If they are not moved by them, Moral Internalism fails. What good is it that Murdoch makes sense of moral reasons if she does so with an indefensible claim?

3. A Platonic Theory of Concepts

A key to the structure of *Sovereignty* is that Murdoch anticipates this complaint.[29] In the first half of "The Idea of Perfection", the part whose targets can seem distant from us, Murdoch criticizes a "genetic" analysis of mental concepts that derives from a broadly behaviourist reading of Wittgenstein. On the genetic theory, "[mental] life is, and logically must be, a shadow of life in public" since, in general, "the possession of [a] concept is a public skill" (Murdoch 1970: 7, 11). Murdoch illustrates this theory with the concept of decision.

> How do I *learn* the concept of decision? By watching someone who says 'I have decided' and who then acts. How else could I learn it? And with that I learn the essence of the matter. I do not 'move on' from a behaviouristic concept to a mental one. […] A decision does not turn out to be, when more carefully considered, an introspectible movement. The concept has no further inner structure; it *is* its outer structure. (Murdoch 1970: 12–3)

26. In later work, McDowell's commitment to Moral Internalism is less clear. Thus, in "Virtue and Reason", the knowledge that issues in right conduct is knowledge of an action "under some such description as 'the thing to do'" on the basis of "a consideration apprehended as a reason" (McDowell 1979: 51, 54; see also 57). Here the virtuous person's conception of the circumstance includes explicitly normative claims.

27. Because it is restricted to decisive reasons, this is not a full account of moral motivation, even when we ignore beliefs about right and wrong. In the case of so-called "imperfect duties", decisive reason applies not to discrete or localized acts, such as helping right now, but to ones that occupy indefinite periods, like doing enough to help those in need. The motivation of the latter falls under Moral Internalism; the motivation of the former does not.

28. This charge is vividly expressed by Smith 1994: 121–5.

29. Here I am indebted to Broackes 2012: 39–48, though I differ from him, at least in emphasis, on the role of moral perception in Murdoch's view. See the discussion of "thick concepts" below.

Although this is something of a caricature, the idea that mental and other concepts are anchored in their public use is not anachronistic. Versions of it are still proposed. What matters here is that theories of this kind support the charge against Moral Internalism. On the genetic theory, nothing that is not apparent in the public acquisition of a concept can be essential to its content. Someone who goes through the ordinary training, and who is competent by ordinary tests, has everything required to grasp the concept expressed by a word. Thus, the person we imagined in section 1 can know about his circumstance exactly what the virtuous person knows. He can know the facts that provide a moral reason without being moved. This is what Murdoch must deny, as she does in language echoed by McDowell:

> There are two senses of 'knowing what a word means', one connected with ordinary language, the other very much less so. […] We do not simply, through being rational and knowing ordinary language, 'know' the meaning of all necessary moral words. (Murdoch 1970: 28)

The question is: What is her argument? How does she dislodge the genetic theory and others like it, theories on which the objection to Moral Internalism goes through? The answer lies in a second key to *Sovereignty*: a correct interpretation of the example of M and D. The example itself is more frequently cited than anything else in the book. A mother, M, feels hostility to her daughter-in-law, D, finding her "pert and familiar, insufficiently ceremonious, brusque, sometimes positively rude, always tiresomely juvenile" (Murdoch 1970: 16–7). This does not affect M's outward behaviour, which is perfect throughout. Yet she experiences moral progress: M "reflects deliberately about D until gradually her vision of D alters. […] D is discovered to be not vulgar but refreshingly simple, not undignified but spontaneous, not noisy but gay, not tiresomely juvenile but delightfully youthful, and so on" (Murdoch 1970: 17). We are to imagine the case as one in which M's vision of D becomes more loving and more just, so that something of moral significance has gone on.

What is the point of this example? Because two issues are run together, one of them is easy to miss.[30] It is obvious that Murdoch is contesting the neglect of private moral activity by the behaviourist obsessed with outward deeds. But this is only part of it. More important is the nature of M's activity, which is refining the way in which she sees the world. Her grasp of the concepts with which she operates — *pert, ceremonious, undignified, gay* — is transformed and improved, and her descriptions change accordingly. It is this phenomenon, in which one's understanding of a concept goes beyond what one knew in acquiring it, or being competent by ordinary tests, that the genetic theorist cannot comprehend. The argument against the genetic theory is that the phenomenon is real: full possession of a concept can transcend the mastery of its public use. The example of M and D is supposed to make this vivid. Innocent of theory, Murdoch believes, we will be tempted to describe the case as one of progress towards perfection in the grasp of mental concepts.

It might be argued, instead, that the story is one of progress towards perfection in understanding another person. But there is no conflict here: we can say that, too. And we have to read Murdoch as concerned with grasp of concepts in order to make sense of her book. The example of M and D is framed by the genetic theory, and its insights are applied to repentance and love.[31] In each case, her topic is conceptual mastery, and it is this topic, perhaps among others, to which the example speaks. Murdoch's conclusion makes this clear:

> The entry into a mental concept of the notion of an ideal limit destroys the genetic analysis of its meaning. […] Is 'love' a mental concept, and if so can it be analysed genetically? No doubt Mary's little lamb loved Mary, that is it followed her to school; and in some sense of 'learn'

30. See, again, Broackes 2012: 47–8.

31. On repentance, see Murdoch 1970: 25.

> we might well learn the concept, the word, in that context. But with such a concept that is not the end of the matter. […] A deepening process, at any rate an altering and complicating process, takes place. (Murdoch 1970: 28)

Though it is not my purpose to defend her view, Murdoch surely has a point. Her descriptions of concept-possession ring true. We say that our understanding of repentance and love has grown, not merely that we have learned new facts about them.[32] If we are right to do so, the genetic theory is wrong. The objection to Moral Internalism thus cannot rely on the genetic theory, or anything like it, for support. There is room for the view that, while someone may appear to be unmoved by knowledge of decisive moral reasons, their grasp of these reasons is imperfect, since they do not possess the relevant concepts in full.

Still, we may doubt that the only source of resistance to Moral Internalism is a bad philosophy of mind. The idea that one can grasp the needs of others without being moved by them can be elicited without the genetic theory. Consider, for instance, a virtuous person who undergoes moral deterioration, becoming indifferent or weak-willed. Is it plausible to claim that his grasp of the relevant concepts, previously flawless, has failed? More generally, we are entitled to ask *why* full possession of certain concepts, along with their correct application to the circumstance, entails motivation or choice. How does this follow from the conditions of concept-possession, on a more adequate account? It is not enough for Murdoch to reject a theory that conflicts with her claims; she needs to sketch an alternative that sustains them.

Here we reach a third key to *Sovereignty*: the false assimilation of Murdoch's theory to an emphasis on "thick moral concepts" such as "*coward, lie, brutality, gratitude*" in which description and evaluation are ineluctably bound together (Williams 1985: 140). Bernard Williams, who makes a great deal of such concepts in *Ethics and the Limits of Philosophy*, credits his appreciation of their importance to a seminar taught by Philippa Foot and Murdoch in the 1950s (Williams 1985: 218n.7). For Williams, thick moral concepts are distinctive in being at once "world-guided" and "action-guiding". They are world-guided in that there are necessary limits to divergence in their use. Those who grasp the relevant concepts are bound to agree in their application, except at the margins (Williams 1985: 140–1). In this respect, they are meant to differ from "thin" concepts like *ought* and *good*, though the difference is presumably one of degree (Williams 1985: 151–2).[33] Thick concepts are action-guiding in that they are "characteristically related to reasons" and because one cannot grasp them unless one shares, at least through imagination, their evaluative point (Williams 1985: 140–2). In possessing a thick moral concept, one participates, perhaps vicariously, in a sensibility that may have motivational force. In effect, it is through this sensibility that users of the concept are able "go on in the same way": this is how motivation or affect is built into the concept.

It is undeniably tempting to interpret Murdoch in light of Williams.[34] Murdoch invites this reading in her reference to "normative epithets" and "secondary moral words" in connection with M and D, and in disparaging "the impersonal world of facts" (Murdoch 1970: 18, 22, 24). The temptation is especially strong with passages like this:

> 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

32. This point is emphasized by Alice Crary in *Beyond Moral Judgment* (Crary 2007: 41–3). Although she cites Murdoch and draws attention to the practical upshot of concept-possession, Crary does not articulate Moral Internalism or the Platonic theory of concepts explored below.

33. Even with thin concepts, there are limits to divergence. As Foot observed, if someone proclaims as an "ultimate principle [that it is] wrong to run around trees right handed or to look at hedgehogs in the light of the moon", we will doubt that he grasps the concepts of moral right and wrong (Foot 2002: xiv).

34. As even her most perceptive readers have done; see, for instance, Broackes 2012: 14–5: "It is particularly ascriptions of the specialized terms [corresponding to thick moral concepts] that are (in relevant situations) immediately motivating."

> naturally. As the empty choice will not occur the empty word will not be needed. (Murdoch 1970: 40–1)

But the approach is quite misleading. First, what Murdoch emphasizes in the concepts that interest her is not the degree of convergence they exhibit but the potential for "a specialized personal *use* of a concept" in ways that may be private or idiosyncratic (Murdoch 1970: 25). Similarly, Murdoch's claim is not that those who use a given concept non-vicariously, the participants for whom it has life, are bound to be moved by its correct application. It is only "true vision", a perfection we rarely approach, that guarantees right action.

More significantly, Murdoch's focus is not on the concepts with which we describe our options — *just*, *courageous*, *cruel* — and whose application guides action, but on the concepts with which we describe our circumstance and the people with whom we interact. Think back to M and D. What the mother gains is not a deeper apprehension of her own moral character, or of her behaviour, but of her daughter-in-law and what she is really like. Nor does Murdoch confine herself to concepts that carry a specific valence, positive or negative, like the ones on Williams' list. As becomes increasingly clear in the second essay of *Sovereignty*, the knowledge that constitutes virtue is not knowledge of the Good, or even of particular virtues, but of the real existence of other people: "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, the harder it becomes to treat a person as a thing" (Murdoch 1970: 64).

In general, Murdoch's talk of "moral concepts" must be heard in relation to a parenthetic remark: "(That mental concepts enter the sphere of morality is, for my argument, precisely the central point.)" (Murdoch 1970: 23–4) The moral reasons full cognition of which entails choice, according to Moral Internalism, are just descriptions of one's circumstance, as, for instance, that someone is in serious need and one can help them at little cost. McDowell's discussion picks up on this: "as the example of 'shy and sensitive' illustrates, the language used to describe a special reason-constituting conception of a situation need not be explicitly evaluative" (McDowell 1978: 86). The result is a much deeper threat to the distinction between fact and value than the mere expansion of moral concepts envisaged by Williams. On Murdoch's view, the description of a circumstance in mental but not explicitly moral terms can have a property often thought to distinguish ethical concepts from others: that knowledge of its application is essentially motivating. What we expected to fall on the fact side of the fact/value distinction turns out to have the attributes of value. This has transformative implications for moral epistemology and the metaphysics of moral properties.

Take metaphysics first. If anything is common ground amidst the disarray of contemporary meta-ethics, it is the *a priori* supervenience of the ethical: what falls under ethical concept E does so in virtue of falling under non-ethical concepts, N, such that necessarily, what falls under N falls under E. Virtually no-one disputes this claim, which is taken as a guide to the nature of ethics.[35] The claim is plausible not only for thin concepts such as *ought* and *good* but for the concepts of specific virtues. If Murdoch is right, however, the psychology of virtue goes beyond these concepts, to the grounds of their application. The reasons why an act is just or unjust, kind or unkind, right or wrong, will satisfy Moral Internalism. They will motivate those who fully possess them. In this respect, they count as moral facts, even though they do not involve the standard menu of moral concepts, thick or thin. Nor is it *a priori* that such facts — for instance, facts about the needs of others and the cost of helping them — supervene on anything else. Since moral philosophy looks directly to these facts, supervenience is not essential to the concepts with which it works. Nor can we assume that explicitly moral concepts apply on the basis of facts that are not themselves moral, since the facts in question may satisfy Moral

35. Rare exceptions include Griffin 1996: 44–8 and Sturgeon 2009, though Sturgeon's doubts pertain more to the formulation of supervenience than to the truth of something in the vicinity.

Internalism. Without the assumption of supervenience, meta-ethics would be almost unimaginably different.

In epistemology, the consequence is, if anything, more striking. Beliefs about right and wrong, or what there is reason to do, may raise epistemic problems. But perhaps they can be ignored. All we need is knowledge of the facts that constitute reasons, plain facts about our circumstance, which influence the will. Assuming we can know the application of mental concepts, there is no problem about moral knowledge in its most basic form.

No doubt these proposals raise questions of their own. Perhaps the result is to make the epistemology of the mental more mysterious than it seemed. We won't pursue that issue here. Our task is to find in Murdoch a theory of concepts that makes sense of Moral Internalism. She rejects the genetic theory, and her claim is not merely about thick concepts but about the whole range of thoughts with which we articulate our social world. What account does she give? The clue to Murdoch's picture lies in the invocation of "realism", which we encountered in section 1. In a Platonic mode, Murdoch connects the realism of virtue with the "appreciation of beauty in art and nature", which is "a completely adequate entry into (and not just analogy of) the good life, since it *is* the checking of selfishness in the interest of seeing the real" (Murdoch 1970: 63).[36] Great art shares in the exactness and objective attention Murdoch associates with morals. The same is true of *technai* in general: in an intellectual discipline, "I am confronted by an authoritative structure which commands my respect. […] Attention is rewarded by a knowledge of reality" (Murdoch 1970: 87). And in the most general formulation of all: "*The necessity of the good is then an aspect of the kind of necessity involved in any technique for exhibiting fact*" (Murdoch 1970: 64).

It is possible to extract from these increasingly abstract claims a Platonic theory of concepts and concept-possession.[37] This theory applies to *all* concepts, not just to those which are morally relevant: "Are there forms of mud, hair and dirt? If there are then nature is redeemed into the area of truthful vision. (My previous argument assumes of course, in Platonic terms, that there are.)" (Murdoch 1970: 86) In outline, the theory is this: each concept is associated with norms for its proper use, both practical and theoretical; these norms describe when the concept should be applied and what follows from its application, both cognitively and in relation to the will; to grasp a given concept is to approximate, in one's dispositions of thought, a conformity with these norms. Concept-possession thus comes by degree and points to a limit we may never reach: perfect compliance with the norms by which our concepts are defined.

This picture of concepts is related to Davidson's "constitutive ideal of rationality" and to the "normativity of the intentional". A more recent Platonist, Ralph Wedgwood, makes a similar claim:

> [The] doctrine that the intentional is normative can be viewed as a way of cashing out Plato's metaphor that the Form of the Good is to the understanding what the sun is to vision (*Republic*, 507b–509a). We count as sighted because we are appropriately sensitive to light, the ultimate source of which is the sun; in a similar way, we count as thinkers because we are appropriately sensitive to normative requirements, the source of which is a

36. See also Murdoch 1970: 82–5.

37. Perhaps the clearest statement of this view is to be found in *Metaphysics as a Guide to Morals*, but that book is too difficult to be integrated here. The most relevant passage is this:

> Plato assumes the internal relation of value, truth, cognition. Virtue (as compassion, humility, courage) involves a desire for and achievement of truth instead of falsehood, reality instead of appearance. […] Learning anything properly demands (virtuous) attention. Here the idea of *truth* plays a crucial role (as it does also in Kant) and reality emerges as the object of truthful vision, and virtuous action as the product of such vision. This is a picture of the omnipresence of morality and evaluation in human life. On this view it would seem mad to begin philosophy by asserting a complete separation of fact from value, and then attempting to give a satisfactory account of morals. (Murdoch 1992: 39)

> coherent system of eternal and necessary truths about what we ought to think or do or feel. (Wedgwood 2007: 3)

On the Platonic theory, we must respond to the norms of reason, at least by approximation, in using the concepts we do. They are at once a condition of thought and an ideal to which we aspire.[38] The "*necessity involved in any technique for exhibiting fact*" is the necessity of this aspiration and this ideal, a necessity involved in any attempt to depict reality as it is.

Wedgwood's Platonism differs from Murdoch's in several ways. First, Wedgwood concentrates on the concept *ought*, not on thick moral concepts or the concepts with which we specify moral reasons.[39] Second, he associates each concept with a single rule, or a "specific rational disposition" (Wedgwood 2007: 169). Murdoch would adopt a more holistic view. As we deepen our concepts, they become more intricately bound to one another: "reflection rightly tends to unify the moral world" (Murdoch 1970: 56). The norms we approach in possessing a given concept do not consist in an isolated rule; they include every norm into which that concept enters. Finally, Wedgwood makes nothing of the claim that our possession of concepts may be partial, that it comes by degree. These contrasts together exclude from Wedgwood's Platonism what is most distinctive in Murdoch's view: her commitment to Moral Internalism. If full possession of concepts entails conformity not only with local norms but with standards of ideal rationality, full grasp of a decisive reason to ϕ will entail decisive motivation. The truth of Moral Internalism follows from the Platonic theory of concepts, assuming that the knowledge to which it refers involves "true vision" or ideal conceptual grasp and that the norms implicit in our concepts are those of practical and theoretical reason. It is in this sense that "[the] authority of morals is the authority of truth" (Murdoch 1970: 88).

38. Antonaccio 2000: 52, 58–9.

39. He defends this focus explicitly at Wedgwood 2007: 97–8, 105–6.

If this is right, Moral Internalism is controversial but not obviously false. It is vindicated by Murdoch's spin on a contentious, though not implausible, philosophy of mind. Her arguments for this conception are scattered, and I am more concerned to make her view cogent than to make it irresistible. But a central thread runs from the example of M and D to the conclusion that our grasp of concepts can be more or less ideal, where the limit is perfect rationality or responsiveness to reason. Murdoch supports and illustrates this claim with concepts of the natural world, with art, and with *technai*: the idea of perfection works within every field of thought.

This argument does not tell us what there is reason to do. In particular, it does not tell us that the needs of other people provide us with reasons to act. If there are such reasons, however, it follows from the Platonic theory that the corresponding norms are built into our concepts. When there is decisive moral reason to act in a certain way, knowledge of that reason, including ideal grasp of the concepts it involves, entails decisive motivation. Moral Internalism holds. This argument answers the question "Why be moral?" not by showing that there is reason to do what is right, or by persuading the amoralist, but by avoiding the dilemma posed in section 1. Even if the argument for ethical rationalism goes through, we can save the generality of moral reasons without deriving their existence from the nature of the will. On the Platonic theory of concepts, there is another possibility: that rational agents are moved by such considerations because they are reasons — or so we assume against the sceptic — and because rationality belongs to full cognition of the facts.

This reading makes sense of much that is obscure in *The Sovereignty of Good*, from the role of the genetic theory, through Murdoch's realism, to her conception of choice. But it leaves a number of loose ends. One is specific to the context in which I have placed Murdoch's views. I have argued that she avoids the problem of moral reasons as it afflicts the ethical rationalist. But this may be too quick. The truth of rationalism would constrain what constitutes a norm of practical reason. Do the norms involved in our possession of concepts,

according to Murdoch's Platonism, meet this constraint? We can put the problem this way: According to the Platonic theory, concepts are associated with norms to which we must approximately conform, ideal possession of a concept involves full conformity, and this is the standard of ideal rationality. According to ethical rationalism, to be practically rational is to achieve the aim or end of agency, as such. We thus have two potentially conflicting views of practical rationality. But we can reconcile their claims. If ideal rationality is full conformity with the norms inscribed in our concepts, and to be practically rational is to achieve the aim or end of agency, as such, the aim of agency must be to grasp the concepts with which to describe our circumstance. Though she is not explicit about it, this may be Murdoch's view. She conceives attention as "a just and loving gaze directed upon an individual reality" and calls this "the characteristic and proper mark of the active moral agent" (Murdoch 1970: 33). As she goes on to say, this is both "a logical and a normative claim": it has the same dual character as the aim of agency, for the ethical rationalist, which bridges the is-ought gap. On this interpretation, while Murdoch offers something like "the reverse of Hampshire's picture", she agrees, in a way, with

> the only explicit 'ought' in his psychology. We ought to know what we are doing. We should aim at total knowledge of our situation and a clear conceptualization of all our possibilities. (Murdoch 1970: 7)

The difference is in how she conceives such knowledge, not as a matter of public, impersonal fact, but as the object of just perception, and not as expanding our range of possibilities, but as closing them off without precluding freedom, so that right conduct is assured. The "ought" of attention to social reality by which our thoughts are perfected is the "ought" of agency or practical reason *and* the "ought" of meeting the norms by which our concepts are defined.[40]

40. Although this reading is possible, I should stress that the text does not require it. While Murdoch's view can be reconciled with ethical rationalism, it is also consistent with the Platonic theory, and with Moral Internalism, to reject the rationalist argument altogether.

Another loose end is Murdoch's emphasis on perception of the individual as a moral phenomenon. Does this conflict with, or does it follow from, the story of conceptual progress told by the Platonic theory? Is there room for moral development that precedes the acquisition of concepts, in which one comes to perceive a situation, or a person, correctly, in ways one cannot articulate in words? These questions relate to larger issues in the philosophy of mind, and I can only touch upon them here. A tempting view is that shifts in perception that constitute moral growth, even when they are inarticulate, involve conceptual change. One may not have words for the content of one's perception, but one can think about it, if only in demonstrative terms: *this* kind of person, *that* way to act.[41] I am inclined to take this route, but it is not essential to my argument. A more concessive response is that there is no reason why the Platonic theory must explain every aspect of Murdoch's view and no reason to fear that it conflicts with interpretations on which perception has an independent role.

Finally, the Platonic theory casts light on Murdoch's attitude to defective concepts and conceptual change. Suppose that possession of a putative concept involves being disposed to think or react in immoral ways: the concept is one that a virtuous person would not have. Examples might be *chastity* or *self-denial*. It is an implication of Murdoch's view that such putative concepts fail. Since the norms definitive of any concept are norms of reason, genuine concepts cannot be in this way flawed.[42] Those who use the relevant words may appear to be thinking, but they are not. It is consistent with this that our present concepts are limited, that there are facets of reason to which they afford no access, and that there is pressure for us to revise and extend our thoughts.

41. See McDowell 1994: 56–8, against "non-conceptual content".

42. For a similar argument, see Wedgwood 2007: 168–9, and for an opposing view, Williamson 2003.

> Moral tasks are characteristically endless not only because 'within', as it were, a given concept our efforts are imperfect, but also because as we move and as we look our concepts themselves are changing. (Murdoch 1970: 27)

We may need new and better concepts in order to comprehend our reasons, to bring into view the facts for which Moral Internalism holds.

This opens an extraordinary possibility. If new concepts make available new facts, knowledge of which is inextricable from choice, philosophy can make moral progress through intellectual change. In the last two essays of *Sovereignty*, Murdoch insists on the practical nature of her project: "How can we make ourselves better? is a question moral philosophers should try to answer." (Murdoch 1970: 76) On the Platonic conception, they can. As Murdoch wrote in "Vision and Choice in Morality":

> Great philosophers coin new moral concepts and communicate new moral visions and modes of understanding. […] 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: 42, 49)

Given Moral Internalism, the extension of language and thought can constitute moral improvement. Nor does Murdoch simply observe this prospect. In the parts of her book that seem most unorthodox, their rhetoric most high-flown, she tries to enact it, to rehabilitate the concept of the Good, neglect of which is a moral, not just an intellectual, vice: "The image of the Good as a transcendent magnetic centre seems to me the least corruptible and most realistic picture for us to use in our reflections on the moral life." (Murdoch 1970: 73) Against the background of the Platonic theory, we can explain what Murdoch is doing in these passages, why it matters to moral philosophy, and how it constitutes a form of proof.

4. The Sovereignty of Good

The language of my account so far has been conspicuously different from Murdoch's. I have written about the norms of practical and theoretical reason inscribed in our concepts and of the standards of ideal rationality in their use. Murdoch writes instead about perfection and the Good. Rationality is not a central concept in her book. The contrast here is not just terminological. What is at stake in the idea of the Good is the unity and coherence of the norms involved in the perfect grasp of mental concepts. This issue is raised by Murdoch in her essay "On 'God' and 'Good'":

> The notion that value should be in some sense *unitary*, or even that there should be a single supreme value concept, may seem, if one surrenders the idea of God, far from obvious. Why should there not be many different kinds of independent moral values? Why should all be one here? (Murdoch 1970: 55)

To believe in the Good is to believe that, in an evaluative sense, all is one. To illustrate this claim, Murdoch cites the potential unity of the virtues: "if we reflect upon courage and ask why we think it to be a virtue, what kind of courage is the highest, what distinguishes courage from rashness, ferocity, self-assertion, and so on, we are bound, in our explanation, to use the names of other virtues" (Murdoch 1970: 56).[43] But this does not exhaust the belief she has in mind.

> [What] is it for someone who is not a religious believer and not some sort of mystic, to apprehend some separate 'form' of goodness behind the multifarious cases of good behaviour? Should not this idea be reduced to the

43. See also Murdoch 1970: 93; McDowell 1979: 50–3.

> much more intelligible notion of the interrelation of the virtues, plus a purely subjective sense of the certainty of judgements? (Murdoch 1970: 59)

The hoped-for answer is that it should not. We can make sense of a deeper unity and of a single form of Good behind the various kinds of excellence in the application of concepts and responsiveness to reasons.

I think we can understand what Murdoch wants here by noting how much the Platonic theory leaves open. According to this theory, each concept is associated with norms for its proper use. In order to grasp a given concept, one must satisfy these norms, at least to some degree. But there is room for more complete possession of a concept, for one to approach perfection by meeting its norms in full. Nothing in this account rules out the following possibility: that the norms for concept F and concept G are incompatible, that we cannot perfect our grasp of both. Suppose, for instance, that one act falls under F, another under G. Knowledge that an act is F, with full grasp of the concept, entails decisive motivation: the fact that the act is F is a decisive reason to perform it. At the same time, knowledge that an act is G, with full grasp of the concept, entails decisive motivation: the fact that the act is G is a decisive reason to perform it. It follows that one cannot fully grasp both facts, since one cannot meet both norms. The result is a kind of fragmentation in reason.

This description may harbour some hidden incoherence, but it does not conflict with the letter of the Platonic theory. All that is implied is that the standard of ideal rationality in the possession of every concept is unattainable. There are tragedies in which we have decisive reason to do incompatible things. Whether this is true or not is, for Murdoch, a real question: "The notion that 'it all somehow must make sense', or 'there is a best decision here', preserves from despair; the difficulty is how to entertain this consoling notion in a way which is not false" (Murdoch 1970: 55). If the Good exists, the norms involved in our concepts are compatible: there are no tragedies in which, whatever one does, one acts against a decisive reason. There is always a right decision, one that satisfies the norms involved in every concept that applies to one's circumstance. Belief in the Good is expressed in "the idea […] that the lines really do converge" (Murdoch 1970: 97): the lines traced out by the norms of each concept, which converge in the Good. "For all our frailty the command 'be perfect' has sense for us" (Murdoch 1970: 90). Perfect grasp of every concept may be psychologically out of reach, but it is not impossible.

Belief in the Good is a protection against despair because it tells us that every problem can be solved. But the consolation may go further. At times, Murdoch suggests that faith in the Good is a recognition of "the absolute pointlessness of virtue [and] its supreme importance", that "nothing in life is of any value except the attempt to be virtuous" (Murdoch 1970: 84–5). The unity here is not just that of a right decision, one that responds to every fact, but that if one makes this decision, nothing else matters: there is no cause for regret or dismay; all other reasons are "silenced".[44] Murdoch emphasizes, too, that while it may be difficult, "contemplation of the Good [is] a source of uncontaminated energy, a source of *new* and quite undreamt-of virtue" (Murdoch 1970: 99). It is a "psychological fact, and one of importance to moral philosophy, that we can all receive moral help by focusing our attention upon things which are valuable: virtuous people, great art, perhaps […] the idea of goodness itself" (Murdoch 1970: 54–5).[45]

Murdoch's discussion at times conflates the Platonic theory of concepts with belief in the Good. In fact, she is a Platonist twice over: once in relating each concept to the idea of its own perfection and then in the idea of a single perfection in which every concept shares. As my discussion shows, these ideas are distinct. There is the Platonic theory which supports Moral Internalism, and there is the further claim of unity in the Good. Although she is not careful to distinguish them, Murdoch sees the need to argue for the second claim in a way that goes beyond the first. She does not end with *"the necessity of the good [as] an*

44. For the idea of silencing, see McDowell 1978: 90–4, 1979: 53–6.

45. On the difficulty here, and whether the Good can be a direct object of attention, see Murdoch 1970: 67–8, 95, 97, 99.

aspect of the kind of necessity involved in any technique for exhibiting fact", which I have taken as an expression of the Platonic theory (Murdoch 1970: 64). Instead, she worries, repeatedly, that for all its psychological benefits, the idea of "an uncorrupted good" is "the merest consolatory dream" (Murdoch 1970: 59). Attention to the Good may be a source of moral energy; it may save us from despair to think that everything makes sense, that "we must receive a return when good is sincerely desired" (Murdoch 1970: 62). "[The] only difficulty is that none of it is true" (Murdoch 1970: 70). What is Murdoch's argument for the reality of the Good?

Looking back to the end of section 3, let us begin with this: According to Murdoch, we can become morally better by seeing the world in light of the Good. Murdoch offers this concept precisely as a source of "moral help" and "uncontaminated energy". In a way, there is no mystery here. It is not mysterious what we gain from belief in the Good or how it can affect our actions, if Murdoch is right. This belief is a source of moral motivation. The puzzle is what this has to do with the existence of the Good. Why should practical reasons for using a concept or holding a belief show that the concept is not empty or that the belief is true? Murdoch is adamant that hers "is not a sort of pragmatism or a philosophy of 'as if'" (Murdoch 1970: 72–3). We should not simply pretend that the lines converge; we are justified in thinking that they do.

In order to make sense of this, we need to say more about the concept of the Good. To think in terms of this concept — of the perfection of thought, as such — is to interpret the norms involved in other concepts as parts of a coherent whole. Among the norms that define the concept of the Good itself are these: from the fact that the norms of a concept require some response, infer that it is required by the Good; and from the fact that the Good requires some response, infer that no other response can be required. If we reason in this way, we will conclude that the norms of every concept are consistent in practice with the norms of every other. We will be committed to the resolution of conflict. When the demands of disparate concepts appear to come apart, we will regard the tension as merely apparent, revising our use of these concepts until it gives way. The idea of the Good thus operates as a regulative ideal for our changing conceptual grasp.

Now, so far, this is mere description: it is an account of how a concept works, in the terms set out by the Platonic theory. But it supports an "ontological proof".[46] Ask, first, whether it is rational to make the inferences that define the concept of the Good. If it is irrational, "the Good" is not the name of a genuine concept, for reasons given at the end of section 3. On the other hand, if it is rational to think in these ways, it is rational to revise our concepts in the direction of unity, to resolve apparent conflicts between them, and to act in light of the understandings we reach. It follows that the norms involved in our concepts do not essentially conflict. There is a rational way to understand these concepts on which their norms cohere with one another: a way in which we are led to understand them if we reason in terms of the Good. It follows in turn that there is always a right decision, one that is ratified by every norm. There is always an answer to the question, "What should I do?" There are no insoluble tragedies, and we are right to believe in the existence of the Good.

This line of thought is an ontological proof, in that it moves from the nature of a concept to the reality of what it stands for. If thought about the Good is possible, if it is not the mere illusion of thought, then on the Platonic theory of concepts, it must be rational, and if it is rational, the concept of the Good cannot be empty. Of course, there is no proof of the antecedent. Murdoch asks, about our invocations of the Good: "Can we give them any clear meaning or are they just things one 'feels inclined to say'?" (Murdoch 1970: 59) The best assurance of sense is that our putative thoughts about the Good are ones that seem intelligible to us in "ordinary life" (Murdoch 1970: 72). After that, "the

46. Murdoch cites the ontological proof, parenthetically, in relating perfection or absolute good to necessary existence (Murdoch 1970: 59–60). (This is one of the points at which I think she conflates the reality of the Good with the Platonic theory of concepts.) There is a chapter on the ontological proof in *Metaphysics as a Guide to Morals* that is broadly consonant with the reading offered here; an adequate treatment lies beyond our scope.

philosophical 'proof' […] is the same as the moral 'proof'" (Murdoch 1970: 73). The practical argument for the Good — that it is rational to think in such terms — is the proof that Good exists.

As before, I do not mean to endorse Murdoch's reasoning but to explain how it works. Her argument is controversial not just in assuming the intelligibility of the Good but in its characterization of the concept and its reliance on the Platonic theory of concept-possession. Still, despite appearances, Murdoch does have an argument. When she recommends the "image of the Good as a transcendent magnetic centre", she is not engaging in mere psychology or self-help but stating a premise of her ontological proof. Her concern with the metaphors we live by is not a sign that her interests are more literary than philosophical but a consequence of the Platonic theory, on which our concepts may determine what we do and on which we can argue, in moral terms, about how to think. Most surprisingly of all, we can argue for the reality of the Good from its moral value as an object of attention, blinding but illuminating, at the limit of sight.

5. Concluding Thoughts

There is more in *The Sovereignty of Good* than even an extended study could explore. I have argued that "The Idea of Perfection" makes room for Moral Internalism by rejecting the genetic theory of concepts. Through Moral Internalism, Murdoch can avoid the problem of moral reasons. She need not derive from the nature of agency or the will a sensitivity to moral concerns. The rest of the book defends a form of Platonism about concepts that vindicates Moral Internalism, and a second Platonism, about the unity of rational norms and the existence of the Good. I end by touching on two further themes, the first of which can be framed as an objection.

With the example of M and D, Murdoch shows how our grasp of mental concepts can deepen and grow, how it can transcend what was apparent in their acquisition. She goes on to make a more startling claim, which she repeats more than once, that "the movement of understanding is onward into increasing privacy" (Murdoch 1970: 28). Murdoch insists that the language of moral reasons is "unavoidably idiosyncratic and inaccessible" (Murdoch 1970: 33). There is no single end-point to the perfect grasp of mental concepts: there may be several ways to apply a concept, and to respond to its use, that are equally and perfectly ideal. As we evolve in different ways, we may become increasingly unintelligible to one another, without any of us having gone wrong.[47] This feature of Murdoch's view makes her opposition to the genetic theory, with its emphasis on public standards, especially sharp. But it is not essential to it: a Platonist for whom each concept has a single perfect form would still deny the genetic theory and could argue for Moral Internalism in the same way. And it is potentially at odds with Murdoch's second Platonism, about the unity and reality of the Good, which stresses convergence, not idiosyncrasy. Is there a contradiction here? In my view, there is not: the appearance of conflict is superficial. To believe in the Good is to believe that one can perfect one's grasp of every concept. It is not to believe that there is just one way of doing so. Belief in the Good is thus consistent with the privacy of perfect understanding. Still, on my reading, the doctrine of privacy can be severed from the rest of Murdoch's view: it does not follow from her central claims.

The final theme is in a way the first: Murdoch opens her book by listing, among the "facts […] which seem to have been forgotten or 'theorized away'" in contemporary philosophy, "the fact that love is a central concept in morals" (Murdoch 1970: 2). The second essay of *Sovereignty* repeats the charge: "We need a moral philosophy in which the concept of love, so rarely mentioned now by philosophers, can once again be made central" (Murdoch 1970: 45). Apart from its illustrative use, my discussion has rarely mentioned love. It figures incidentally in reference to attention — the "true vision [that] occasions right conduct" — as just and loving. But it has not been explained.

47. Murdoch's claim may be less radical: that moral development can follow different paths to the ideal, which may involve stages of mutual incomprehension. I focus on the stronger reading in part because it seems more problematic.

Murdoch's theory of love is initially puzzling. Why should accurate perception of another, even with perfect grasp of the concepts applied, go along with love, not loathing or contempt? Suppose D really were tiresome and juvenile, or, if those are defective concepts, consider the "clear-eyed contemplation of the misery and evil of the world" (Murdoch 1970: 59). Must the agents of evil be objects of love? Murdoch's thought is that they must. The air of paradox in this claim can be dissolved in part by recalling that the love in question is not selective — it is not fondness or affection or the desire for intimacy — but love as a moral emotion.[48] (The example of M and D is misleading in this respect. It is an accident that the object of M's loving gaze is her daughter-in-law, not a murderer or an acquaintance at work.) The love that interests Murdoch is the love one should have for one's neighbour — that is, for anyone with whom one interacts.[49] If love in this sense is partial, that is only because we are limited. "Love is the extremely difficult realisation that something other than oneself is real" (Murdoch 1959: 215). Since everyone is real, we ought to love them all. It does not follow from this that we should feel affection for everyone or that we should pursue their interests. Love involves "true vision" and leads us to act towards others as we are morally required to act. It might be out of love that we prevent them from doing harm, or berate them, or refuse to associate with them.

The question I leave unanswered here is whether we can treat Murdoch's use of "love" in the moral context as more than quixotic. What does love as true vision have to do with love in any ordinary sense? Are there materials in the Platonic theory of concepts, or in the ontological proof, for a secular interpretation of "love your neighbour as yourself"? To complete the reading of *Sovereignty*, one would have to work through the relationship of love to truth and to the Good. One would have to explain why "'Act lovingly' [will] translate 'Act perfectly', whereas 'Act rationally' will not" (Murdoch 1970: 99). And one would have to square this with the fact that love is "capable of infinite degradation and is the source of our greatest errors" (Murdoch 1970: 100). Though I cannot do these things, I am optimistic that they can be done, that the lines converge, and that with refinement or deepening, we can come to understand the concept *love* as Murdoch does.[50]

48. Compare Velleman, who cites Murdoch in support of his Kantian view: "respect and love [are] the required minimum and optional maximum responses to one and the same value", that of personhood or rational will (Velleman 1999: 366). For Velleman, love disarms our emotional responses and involves "really looking" at the person we love (Velleman 1999: 361). Despite the resonance, this invocation of Murdoch cannot be right. The love that involves "really looking", for Murdoch, is not optional but required, and directed not at our rational natures but at our morally ambiguous mental lives.

49. On love of one's neighbour, see Murdoch 1970: 21, 72, and in connection with Kant, Murdoch 1959: 219–20.

References

Antonaccio, M. 2000. *Picturing the Human: The Moral Thought of Iris Murdoch*. Oxford: Oxford University Press.

Broackes, J. 2012. Introduction. In J. Broackes, ed., *Iris Murdoch, Philosopher*. Oxford: Oxford University Press, 2012: 1–92.

Clark, P. 2001. Velleman's autonomism. *Ethics* 111: 580–593.

Crary, A. 2007. *Beyond Moral Judgment*. Cambridge, MA: Harvard University Press.

Darwall, S. 1983. *Impartial Reason*. Ithaca, NY: Cornell University Press.

Diamond, C. 1995. *The Realistic Spirit: Wittgenstein, Philosophy, and the Mind*. Cambridge, MA: MIT Press.

Dreier, J. 1997. Humean doubts about the practical justification of morality. In G. Cullity and B. Gaut, eds., *Ethics and Practical Reason*. Oxford: Oxford University Press, 1997: 81–100.

Foot, P. 2002. *Virtues and Vices and Other Essays in Moral Philosophy*; second edition. Oxford: Oxford University Press.

Geach, P. 1956. Good and evil. *Analysis* 17: 32–42.

50. For discussion of these topics, I am grateful to John McDowell and Evgenia Mylonaki, to participants in my spring 2012 seminar on love and ethics at the University of Pittsburgh, and for comments on section 1, to Julia Markovits and to an audience at the Pacific APA. Special thanks to the anonymous readers for *Philosophers' Imprint*, whose detailed and constructive feedback was the source of many revisions.

Griffin. J. 1996. *Value Judgement: Improving Our Ethical Beliefs*. Oxford: Oxford University Press.

Hampshire, S. 1959. *Thought and Action*. Notre Dame, IN: Indiana University Press.

Kant, I. 1785. *Groundwork of the Metaphysics of Morals*. Translated by M. Gregor. Cambridge: Cambridge University Press, 1998.

Korsgaard, C. 1996. *The Sources of Normativity*. Cambridge: Cambridge University Press.

Korsgaard, C. 2009. *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press.

McDowell, J. 1978. Are moral requirements hypothetical imperatives? Reprinted in *Mind, Value, and Reality*. Cambridge, MA: Harvard University Press, 1998: 77–94.

McDowell, J. 1979. Virtue and reason. Reprinted in *Mind, Value, and Reality*. Cambridge, MA: Harvard University Press, 1998: 50–73.

McDowell, J. 1994. *Mind and World*. Cambridge, MA: Harvard University Press.

Murdoch, I. 1956. Vision and choice in morality. *Proceedings of the Aristotelian Society, Supplementary Volume* 30: 32–58.

Murdoch, I. 1959. The sublime and the good. Reprinted in *Existentialists and Mystics: Writings on Philosophy and Literature*. London: Chatto & Windus, 1997: 205–220.

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

Murdoch, I. 1992. *Metaphysics as a Guide to Morals*. London: Chatto & Windus.

Nagel, T. 1970. *The Possibility of Altruism*. Oxford: Oxford University Press.

Putnam, H. 2002. *The Collapse of the Fact/Value Dichotomy and Other Essays*. Cambridge, MA: Harvard University Press.

Railton, P. 1997. On the hypothetical and non-hypothetical in reasoning about belief and action. In G. Cullity and B. Gaut, eds., *Ethics and Practical Reason*. Oxford: Oxford University Press, 1997: 53–80.

Raz, J. 1999. *Engaging Reason: On the Theory of Value and Action*. Oxford: Oxford University Press.

Raz, J. 2011. *From Normativity to Responsibility*. Oxford: Oxford University Press.

Setiya, K. 2007. *Reasons without Rationalism*. Princeton, NJ: Princeton University Press.

Setiya, K. 2010. Sympathy for the devil. In S. Tenenbaum, ed., *Desire, Practical Reason, and the Good*. Oxford: Oxford University Press, 2010: 82–110.

Setiya, K. 2012. Internal reasons. In K. Setiya and H. Paakkunainen, eds., *Internal Reasons: Contemporary Readings*. Cambridge, MA: MIT Press, 2012: 1–34.

Setiya, K. forthcoming. Intention, plans, and ethical rationalism. In M. Vargas and G. Yaffe, eds., *Rational and Social Agency: Essays on the Philosophy of Michael Bratman*. Oxford: Oxford University Press.

Smith, M. 1994. *The Moral Problem*. Oxford: Blackwell.

Smith, M. 2009. The explanatory role of being rational. In D. Sobel and S. Wall, eds., *Reasons for Action*. Cambridge: Cambridge University Press, 2009: 58–80.

Smith, M. 2010. Beyond the error theory. In R. Joyce and S. Kirchin, eds., *A World Without Values: Essays on John Mackie's Error Theory*. Dordrecht, Holland: Springer 2010: 119–140.

Sturgeon, N. 2009. Doubts about the supervenience of the ethical. *Oxford Studies in Metaethics* 4: 53–90.

Taylor, C. 1989. *Sources of the Self: The Making of the Modern Identity*. Cambridge, MA: Harvard University Press.

Tenenbaum, S. 2007. *Appearances of the Good: An Essay on the Nature of Practical Reason*. Cambridge: Cambridge University Press.

Velleman, J. D. 1989. *Practical Reflection*. Princeton, NJ: Princeton University Press.

Velleman, J. D. 1992. The guise of the good. *Noûs* 26: 3–26.

Velleman, J. D. 1999. Love as a moral emotion. *Ethics* 109: 338–374.

Velleman, J. D. 2000. *The Possibility of Practical Reason*. Oxford: Oxford University Press.

Vogler, C. 2002. *Reasonably Vicious*. Cambridge, MA: Harvard University Press.

Wallace. R.J. 1990. How to argue about practical reason. *Mind* 99: 355–385.
Wedgwood, R. 2007. *The Nature of Normativity*. Oxford: Oxford University Press.
Williams, B. 1979. Internal and external reasons. Reprinted in *Moral Luck*. Cambridge: Cambridge University Press, 1981: 101–113.
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: 249–293.
Wolf, S. 1990. *Freedom Within Reason*. Oxford: Oxford University Press.