WAITING FOR GOOD

Iris Murdoch on Moral Freedom

By
Darren Steven Rondganger

Submitted to the Department of Philosophy,
Central European University

In partial fulfilment of the requirements for the degree of Master of Arts

Supervisor: Professor Simon Rippon

Vienna, Austria
2024

'Love is the extremely difficult realisation that something other than oneself is real' – Iris Murdoch

Acknowledgements

Much like Murdoch's picture of moral life, I often need help from the 'outside' in order to get anything done. Consequently, there are many people who made the completion of this dissertation possible. Firstly, I would like to thank Simon, whose supervision, guidance, and encouragement immeasurably improved the quality of this thesis. I would also like to thank Cathy for her encouragement, and additionally for providing access to chapters from her forthcoming book, which significantly shaped how I think about Murdoch's moral philosophy. I would also like to thank Stephen, Brylea, and Levan for providing critical feedback on earlier versions of all of my chapters which greatly improved their quality. Finally, I would like to thank my family – without the support of my dad, mom, and sister this thesis would not be possible.

Contents

Abbreviations

The following abbreviations are for works by Iris Murdoch. Page numbers are referenced for each of Murdoch's essays to the versions reprinted in *Existentialists and Mystics* (1999), edited by Peter J. Conradi. Additionally, Murdoch's book *Metaphysics as a Guide to Morals* (1993) is abbreviated as 'MGM'.

DPR - 'The Darkness of Practical Reason'

IP - 'The Idea of Perfection'

M&E - 'Metaphysics and Ethics'

NP - 'Nostalgia for the Particular'

OGG - 'On "God" and "Good"'

S&G - 'The Sublime and the Good'

SGC - 'The Sovereignty of Good Over Other Concepts'

VCM - 'Vision and Choice in Morality'

P.H. Nidditch's revised edition (1978) of David Hume's *A Treatise of Human Nature* (1740) is abbreviated as 'T' (cited by book, section, and page number).

Mary Gregor's translation (1996) of the *Groundwork of the Metaphysics of Morals* (1785) which appears in the Cambridge edition of the works of Immanuel Kant is abbreviated as 'G'. References are made to the pagination of the standard German Edition of Kant's works, *Kants Gesammelte Schriften*, edited by the Royal Prussian Academy of Sciences.

G.M.A Grube's translation, revised by C.D.C Reeve, of Plato's *Republic* which appears in *Plato: Complete works* (1997) edited by John M. Cooper is referenced as 'R'. References are made to the standard Stephanus numbers of Plato's work.

Introduction

Iris Murdoch's moral philosophy offers a unique and compelling way to make sense of our moral lives. For Murdoch, our moral life and our life as a whole are indistinguishable. Crucial to this picture of moral life is a substantial picture of the moral agent. On her account, moral freedom consists in the suppression of the 'self'. However, this more compelling account of moral life is threatened by the worry that the individual disappears at the moment of freedom. It appears that the substantial individual, which Murdoch wants to give an account of, disappears if a key constitutive element of them disappears when they are free.

Our ordinary conception of freedom holds that freedom consist in freedom from something in order to do something else. In this way, our conception of freedom is relational. A necessary component of this conception of freedom is that what one is free from cannot be an essential part of what makes you who you are. Ana Barandalla proposes to solve this problem by reading Murdoch's individual as the same sort of individual that emerges from Korsgaard's reading of Kant's moral philosophy.

This solution, whilst seemingly promising, is unable to do justice to Murdoch's wider philosophical commitments. In particular, this account cannot make sense of Murdoch's suggestion that an individual exercises their moral agency through paying loving attention to another individual's reality in order to gain knowledge of them. I will argue that we ought to conceive of Murdoch's conception of moral freedom as consisting of two constitutive elements: (1) freedom from Freudian forces and (2) knowledge of the reality of other individuals. If we take up this twofold conception of moral freedom, then we can not only make sense of Murdoch's account of moral agency but also dissolve the apparent problem of freedom.

Chapter One: The Problem of Moral Freedom for Iris Murdoch

'But can we conceive of a being that isn't capable of Will at all, but only of Idea (of seeing for example)? In some sense this seems impossible. But if it were possible then there could also be a world without ethics' - (Wittgenstein, 1961, p. 77e)

In this chapter, I will introduce the picture of moral agency that is the target of Murdoch's critique of moral philosophy. Following her critique of this picture, I will introduce her alternative account of moral agency which centres around a substantial individual. This alternative, however, faces a problem in accounting for what it means to be morally free. In the last part of this chapter, I will present how Murdoch's account of moral freedom threatens to undermine her conception of moral agency centring around a substantial individual.

1.1 The 'Man' of Modern Moral Philosophy

Over the course of our lives, we are often brought to reflect upon what can be called 'moral experiences'. The exact content of these experiences may vary from case to case – from lying to a friend to donating to a charity. What unifies our reflection upon our moral experiences is that we often question how we ought to have *acted*. When we reflect upon our moral lives, we are, more often than not, thinking about whether I ought to have *lied* to such-and-such or whether it would have been good of me to have *helped* so-and-so. Our reflection upon our moral lives is almost always characterized in terms of what we chose to do or what we failed to choose to do. This choice-centric account of moral life inevitably gives us a certain picture of our moral lives: it is but the sum of our good and bad *choices*.

However natural this picture of moral life appears to us, it has a particular socio-historical origin which arises out of several contingent social, political, and religious revolutions.[1] Moral

[1] See (MacIntyre, 2007) for an opinionated account of how the broadly Enlightenment conception of moral philosophy arose.

philosophers in the early 20th century, steeped in the aftermath of these revolutions, built a certain picture of what a moral agent is: rational, autonomous, and acting in a world of objective facts. Iris Murdoch identifies this post-Kantian 'man' of modern philosophy as arising out of two broadly shared philosophical assumptions (VCM p. 77):

1. Empiricism.
2. 'Willing' is at the heart of moral activity.

These two philosophical assumptions make the choice-centric view of moral life very appealing. The assumption of Humean empiricism suggests that we live in a world which is independent from our values (VCM p. 78). For Hume, our knowledge of the world is ultimately grounded in experience (T p. 4; Guyer, 2017, p. 338). Following this epistemic assumption, the nature of value is to be uncovered through observation. For Hume, this ultimately leads to the conclusion that 'Vice and virtue…may be compar'd to sounds, colours, heat and cold, which are not qualities in objects, but perceptions in the mind' (T III i p.1-2, quoted in Mackie, 1980, p. 64). Hume's empiricism thus leads him, and many of those who endorse his empiricism, to a view of a world where facts can be separated from our value-laden perception of the world. The 'man' of modern philosophy is taken to live in this world where facts and our values are fundamentally cleaved apart.

The second philosophical assumption, that 'willing' is at the heart of moral philosophy, follows from Kant's account of the origin of normativity. In the *Groundwork for the Metaphysics of Morals* (1785), Kant argues that the sole ground of value is the good will, and that value is given to action in virtue of the action being an expression of the good will of the agent (G 4:393; Schapiro, 2020, p. 160). For Kant, the will is thus the heart of moral activity because the will is to be identified with practical reason:

> Everything in nature works in accordance with laws. Only a rational being has the capacity to act in accordance with principles, or has a will. Since reason is required, for the derivation of actions from laws, *the will is nothing other than practical reason* [emphasis added]. If reason infallibly determines the will, the actions of such a being that are cognized as objectively necessary are also subjectively necessary, that is, the will is a capacity to choose only that which reason independently of inclination cognizes as practically necessary, that is, as good (G 4: 413).

Following this view of the nature of practical reason, we can see how Kant's account of morality gives us a choice-centric view of moral philosophy. If the will is a faculty for choosing and these choices are necessarily represented as good, then it appears to follow that value is imbued in the world through our choices. Whilst most of the philosophers who endorse the choice-centric view of ethics do not endorse the implicit metaethical claims about value in the aforementioned quotation, they still take from Kant the idea that choice is connected with how we determine the rightness or wrongness of actions.

These two philosophical assumptions are highlighted in Murdoch's suggestion that modern moral philosophers make moral life like a visit to the shop:

> On this view, one might say morality is assimilated to a visit to a shop. I enter the shop in a condition of totally responsible freedom, I objectively estimate the features of the goods, and I choose. The greater my objectivity and discrimination the larger the number of products from which I can select (IP p. 305).

When one visits the shop and selects a product, one's actions are public and thus, at least in principle, objectively specifiable to everyone who may observe your shopping. If this is what our moral lives are like, then it seems that it implies a particular account of moral agency. In Murdoch's two key essays on moral agency *Vision and Choice in Morality* (1956) and *The Idea of Perfection* (1970), she identifies three central features of a choice-centred theory of moral agency, which can all be seen in the shop analogy:

1.) The site of moral action is the will (IP p. 301) .

2.) What happens between our overt acts of willing is morally irrelevant (IP p. 304).

3.) Moral concepts and moral judgments concern an objective and public activity with an associated attitude directed towards the action (VCM p. 77, IP p. 305).

What one is responsible for is what one *chooses* (1) to buy, what happens between one's purchases is irrelevant (2), and the only concepts (3) readily available to determine what one chooses are the ones given by the objective features of the goods. All three features can be seen to be expressions of the two philosophical assumptions of empiricism and 'willing' being at heart of moral activity.

For Murdoch this choice-centric account of moral agency is unable to do justice to the reality of our moral lives on at least three accounts: (1) empirical inadequacy, (2) philosophically impoverished, and (3) morally deleterious (IP p. 306). I believe it is best to understand her objection on these three accounts by looking at the alternative picture of moral life which she puts forward.

1.2 Murdoch on Moral Agency and Moral Freedom

Murdoch's chief complaint against the choice-centric account of moral agency is the kind of features of moral agents it asks us to endorse. The aforementioned philosophical assumptions, paint a particular picture of moral life that Murdoch finds objectionable. The problem with this picture of moral agency is that, as Murdoch famously quips: 'the agent is thin as a needle, appears in the quick flash of the choosing will' (OGG p. 343). Moral agents appear to only arise at moments of overt willing before disappearing between their public acts. The choice-centred moral agency is unable to account for what Murdoch takes to be two obvious facts about our moral lives (IP p. 299):

1.) The unexamined life can be virtuous.

2.) Love is a central moral concept.

In order to see why a proper account of our moral lives must account for these facts, it would be helpful to look at Murdoch's famous tale about moral progress:

> A mother, whom I shall call M, feels hostility to her daughter-in-law, whom I shall call D. M finds D quite a good-hearted girl, but while not exactly common yet certainly unpolished and lacking in dignity and refinement. D is inclined to be pert and familiar, insufficiently ceremonious, brusque, sometimes positively rude, always tiresomely juvenile. M does not like D's accent or the way D dresses. M feels that her son has married beneath him. Let us assume for purposes of the example that the mother, who is a very 'correct' person, behaves beautifully to the girl throughout, not allowing her real opinion to appear in any way. [...] However, the M of the example is an intelligent and well-intentioned person, capable of self-criticism, capable of giving careful and just attention to an object which confronts her. M tells herself: 'I am old-fashioned and conventional. I may be prejudiced and narrow-minded. I may be snobbish. I am certainly jealous. Let me look again.' Here I assume that M observes D or at least 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. And as I say, ex hypothesi, M's outward behaviour, beautiful from the start, in no way alters (IP pp. 312–313).

How are we to describe her moral progress? The choice paradigm does not give us many resources. There is no action or choice that M commits that seems to make sense of her moral transformation. Furthermore, the choice-centric paradigm might not even allow us to say that D has improved morally – especially if what happens between our actions does not matter. That is, we are inclined to say that she has improved morally *despite* the fact that it has nothing to do with how M treats D *ex hypothesi*. If this is so, then it appears that the site of moral action, the place where one exerts moral agency, cannot be reduced to 'movements' of the will in an objective and public world. If one can exert moral agency by coming to change how one sees someone, then it appears that the site of moral action must, at the very least, also include changes in vision.[2]

[2] What a moral vision ultimately amounts to, will concern much of chapter 3 and chapter 4. All that is necessary, for present purposes, is that it is prima facie plausible that moral vision is not reducible to the movement of the will.

The possibility of a change in vision, as constituting moral improvement, allows us to see Murdoch's alternative account of moral agency. It follows *a fortiori* that if a change in vision constitutes an exertion of moral agency, then what happens between acts of overt choice is morally significant. Furthermore, M's moral improvement has not occurred because M has reflected upon principles of justice or the nature of virtue, but rather that she improved morally in virtue of having come to *see* D as she really is. For Murdoch, what aids us in our change of vision is love (IP p. 329).

When M comes to see D clearly, M attends to the features of D with a just and loving gaze and thus sees her for who she really is. M's moral transformation is a transformation by means of love. What Murdoch means by 'love' is up for much debate, but what is clear is that the concept of love is closely related to what Murdoch takes to be the central mode of moral activity: loving attention (IP p. 327). It is through a loving attention that we come to see the world clearly and act justly. We can thus see how love becomes central to Murdoch's account of moral agency in virtue of its epistemic powers to reveal the reality of others clearly.

Another central aspect of Murdoch's account of moral agency, which emerges from the tale of M and D, is what we can call the privacy of moral concepts. The story of M and D reminds us that the concepts which M uses to characterise D are definitively M's *own* (IP p. 317). That is, the concepts employed in M's changing judgments about D are concepts to which M has the greatest grasp. Murdoch, thus ultimately presents a picture of moral agency in which primacy is given to the inner life of substantive individuals, for whom their moral progress involves their whole personality.

On Murdoch's account, limiting moral action to overt expressions of the self through the will is both empirically implausible in giving an account of the way humans act morally, and furthermore it also fails to capture the two facts of moral life: that the unexamined life can be

virtuous and that love is a central moral concept. Against this picture of moral life being constituted by free choices made by a rational agent, Murdoch proposes that our moral lives ought to be seen as a struggle to come to see other individuals in a just and loving manner. This alternative account of moral life insists that our freedom lies not in our ability to make free choices, but rather in our ability to see the world differently (M&E p. 73).

Murdoch's picture of moral agency challenges the orthodox conception of freedom, as Murdoch wishes to construe M's change in how she sees D as a kind of activity (IP p. 314). Murdoch proposes an account of freedom that allows for inner activity, changing one's vision, to be considered as moral activity. One's inner life is thus a component of moral activity. For Murdoch, moral freedom ultimately amounts to a reflective state in which one is able to see the world clearly and accurately (VCM p. 95).

What exactly is involved in having an accurate vision of the reality of individuals? For Murdoch, one necessarily suppresses the 'self' (OGG p. 354). Why does Murdoch take a suppression of the self to be necessary to achieve moral freedom? The answer lies in her somewhat dark view of human nature. Although critical of Freud's overall account of human nature, Murdoch is sympathetic to Freud's account of human motivation (OGG p. 341). Freud's suggests the human psyche is not only egocentric, but also that it is determined by its own peculiar history, and that it has 'natural attachments' which are not under control of the individual (OGG p. 341). The tendency of the human ego towards egocentricity is what obscures the reality of others from us (OGG p. 342). The human ego, and the fantasies that it engenders, are the enemy of seeing others as they really are.

If moral freedom consists of an accurate and loving vision, then the suppression of the egocentric self is necessary to see other individuals as they really are. When M see's D in a just

and loving manner there is, as Murdoch puts it, a 'suppression of self' which allows M to see D in a just and loving manner (OGG pp. 353–354).

However, if the 'suppression of self' is a necessary component of freedom for Iris Murdoch, then it appears that M 'disappears' when she sees D justly. The difference between M when she sees D as 'juvenile' and 'undignified' and when she sees D as 'youthful' and 'spontaneous' appears to be a difference resulting from M's loss of 'herself' (IP p. 313). If we cannot make sense of M's moral freedom without maintaining M's substantial self, then it appears that Murdoch's account fails to do what it set out to do: give an account of moral life that honours the substantial nature of individuals.

1.3 The Logic of Freedom and Murdoch's Individual

For Murdoch, individuals are – unlike the 'lonely will' of the Kantian moral agent – connected to the whole history of their lives (IP p. 328). When evaluating the moral character of individuals, it is necessary to have a grasp of the moral concepts which *particular* agents act with and through. These concepts are determined by the 'background of our lives' and ultimately determine our actions (OGG pp. 343–344).

The constitutive features of Murdoch's individual enable us to give an account of the moral motivation that is connected to each individual's particular life. Murdoch points out that the moral agents exemplified by the post-Kantian moral agents lack an adequate account of moral motivation (OGG p. 343). Since we are not told what 'prepares' the individual for their moral choice (OGG p. 343). In contrast, as we have seen Murdoch's individuals are *essentially* historically and particularly situated moral agents who are influenced by forces which guide their actions.

However, keeping these historically situated features and forces as *essential* components of moral agents produces a problem in accounting for the moral freedom of these individuals. When we ordinarily conceive of moral agents as free, we conceive of them as free just in the case that they are *free from* certain constraints impacting their agency. In order for Murdoch's conception of the free individual to remain coherent, it must be the case that the constraints impacting the freedom of such individuals are not constitutive elements of the substantive individual.

Ana Barandalla elaborates on this problem by pointing out that Murdoch's account of freedom appears to be undermined by the ordinary structure of the concept of freedom (Barandalla, 2023). Barandalla presents the ordinary structure of the concept of freedom in this manner:

> X is free from Y to Φ just in the case that Y is not a part of X qua Φer & Y is not expressed in X's Φing (Barandalla, 2023, p. 206).

In order to see why this concept of freedom presents a problem to Murdoch's account of moral freedom, we ought to further analyse what this conception of freedom suggests about the nature of free individuals. This conception construes the concept of freedom as fundamentally relational (Barandalla, 2023, p. 206). Individuals thus bear the property of freedom in *relation* to some set of constraints which an individual is free from. Conversely, when one is unfree, it is the case that some set of constraints renders one unfree. It is an important aspect of this conception of freedom that the relata (X and Y) are not constitutive of each other.

Barandalla asks us to consider the example of the relationship between a runner and the hair colour of the runner (Barandalla, 2023, p. 206). The runner is free from their hair colour to run because the hair colour is *not* a constitutive element of the runner qua runner, nor is the hair colour of the runner an expression of the activity of running. The freedom of individuals is thus

intimately tied to what we take individuals to be, what they can be free from, and what they are free to do. Moral freedom takes this more general structure of the concept of freedom with respect to individuals as moral agents. We think individuals are blameworthy or praiseworthy, as moral agents if their actions are expressions of their freedom from constraints on their agency (Barandalla, 2023, p. 206).

However, it is in this aforementioned conception of moral freedom that the free substantive individual appears to disappear in according to Murdoch's account. If the Freudian forces and 'self' are that which the free substantive individual is *free from*, then Freudian forces and 'self' cannot be that which constitute the substantive individual. Otherwise, we are left with the question of *who* exactly sees the world in a just and loving manner. if seeing the world in a just and loving manner requires the 'suppression of the ego', which is supposed to be a constitutive element of the substantive individual, then it cannot be a part of the free substantive individual.

Much of Murdoch's motivation for developing her alternative account of moral agency is built on the idea that the man of modern philosophy is too 'thin'. Murdoch's alternative and far more compelling picture gives us a substantive individual with a personal history. However, if these Freudian forces – which are part of the substantial self – disappear at the moment of clear vision, then it appears that these Freudian forces cannot be an *essential* element of the substantial individual. If one wanted to save Murdoch's conception of freedom by suggesting that the Freudian forces are not an essential element of the individual, then one loses the substantial and realistic conception of moral agents that Murdoch wanted to develop (Barandalla, 2023, p. 212).

If one thinks that there is something compelling about Murdoch's substantial conception of moral agents, then one ought to find a way to dissolve this problem of freedom. In the next chapter of my thesis, I will explicate and evaluate Barandalla's own attempt to dissolve the problem of freedom.

Chapter Two: Barandalla's Constitutivist Solution

"Do not merely show us by argument that justice is superior to injustice, but make clear to us what each in and of itself does to its possessor, whereby the one is evil and the other is good"
- Plato (R 367b).

Following the problem of freedom identified in the previous chapter, an account of Murdoch's substantial individual ought to be to be made congruent with the ordinary conception of freedom. Ana Barandalla's own solution to this problem relies on the Kantian constitutivist conception of agency. In this chapter, I first unpack the core ideas of Kantian constitutivism and then see how Barandalla applies this idea to Murdoch's account of agency and the problem of freedom.

2.1 Korsgaard's Constitutivism

Constitutivism has its origin in Christine Korsgaard's reading and defence of Kant's moral philosophy.[3] Korsgaard identifies the reflective structure of human consciousness as the ground of normativity (Korsgaard, 1996, pp. 47- 48). For Korsgaard, in attempting to give an account of the authority of moral claims, we must give a justificatory account of the origin of normativity (Korsgaard, 1996, pp. 9-10). In giving a justificatory account, we are able to answer the 'normative question': What justifies morality? (Korsgaard, 1996, p. 10). Korsgaard's answer to this question lies in our ability to stand back and reflectively endorse the principles upon which we act (Korsgaard, 1996, p. 129).

[3] Why not Kantian Constructivism as opposed to Kantian Constitutivism? I stick with Barandalla's terminology for the sake of terminological consistency, however there is no substantial difference between the constitutivism and constructivism when it comes to giving an account of either normativity or the nature of practical agents. For a critical introduction to constitutivism as an account of the origin of normativity, see (Enoch, 2021).

When we are faced with any principle that we could act upon, we can always ask: *Ought I act upon this principle?* For Korsgaard, this ability is not only key to giving an account of the origin of normativity, it is also what gives us our identity as practical agents (Korsgaard, 1996, p. 129). Following this characterization of reflective endorsement as key to our practical identity, Korsgaard develops an account of action where something can count as an action attributable to someone just in the case that it issues from the person's *whole* constitution (Korsgaard, 2009, 1999). That is, we can only attribute an action to someone if we can attribute this action to their whole person. On this account, we end up with the following conception of an action (Katsafanas, 2018, p. 372):

1. An agent's F-ing is an action iff F-ing is attributable to the agent as a unified whole.

Korsgaard's account of action, thus depends on what it is for an individual to be a unified agent. Building on her reading of Plato, Aristotle, and Kant she develops a constitutional model of agents (Korsgaard, 2009). In developing her view, she begins with a rejection of a certain conception of agents found in Hume's account of practical reason (Korsgaard, 1999, p. 2). Hume rejects what Korsgaard calls a 'combat' model of the soul (T II.3.3 413–418 ; Korsgaard, 1999, p. 2). According to such a view, reason and passion are the two constitutive elements of the agent that battle for the determination of human action. An agent's actions are thus determined by either reason or passion, whichever carries the stronger motivational force in the agent at the moment.

Hume, of course, goes on to deny that there is in fact any real 'combat'. For Hume, reason is nothing but a 'slave of the passions', as reason merely procedurally operates upon our competing desires which are our ultimate reasons for action (T II.3.3 415; Mackie, 1980, pp. 1–2, 44–45). Only our passions can provide us with reasons to act (Mackie, 1980, pp. 44–45).

Before evaluating Hume's model of the soul, Korsgaard asks us to take another look at the assumptions of the Combat model (Korsgaard, 1999, p. 2). When we assume that the constitutive elements of an agent are (1) reason and (2) passion and that agents have to choose between reason and passion, then it seems that the agent is choosing between constitutive elements of themselves.

If agents are constituted by reasons and passions, *who* is it that decides between reasons and passion? The dilemma is as follows: Either the agent genuinely chooses between reason and passion, and is thus neither identified with both, or the agent is identified with reason and passion, and thus the agent has nothing to choose from which is independent of herself. Korsgaard, rightly, notes that this combat model, which Hume assumes to be coherent, does not tell us *who* it is that chooses (Korsgaard, 1999, p. 2).

Korsgaard points to another philosopher who gives us a much clearer picture of the constitution of the agent who acts: Plato (Korsgaard, 1999, p. 3). Following Plato's suggestion in *the Republic* that the human soul is structured like the constitution of the *Polis*, Korsgaard calls this view the constitutional model of the soul (R 440e; Korsgaard, 1999, p. 3). On the constitutional account, something is considered an action in virtue of the fact that it is attributable to the whole person (Korsgaard, 1999, p. 3). On this account, actions are made attributable to the whole person because such actions are in accordance with one's constitution (Korsgaard, 1999, p. 3).

What's important about this view for Korsgaard is that it provides an internal standard for good and bad action. Given that something is only identifiable as an action insofar as it stems from your whole constitution, we can derive a standard for actions dependent upon the degree to which an action stems from one's whole constitution (Korsgaard, 1999, p. 3). On such an account, the actions which unify an agent are the ones that flow from the agent's constitution and thus are more attributable to the agent as the author of their action (Korsgaard, 1999, p. 3).

We can also conceive of bad actions as those actions which are not authored by whole person, but rather as those actions which stem from, as Korsgaard puts it, 'something at work *in* or *on* the person' (Korsgaard, 1999, p. 3).

One might wonder why it should be the case that the more attributable an action is to an agent the *better* the action is. Korsgaard ultimately argues that the normativity of action stems from the nature of what it is for something to be an action in the first place – its constitutive standards (Korsgaard, 2009, p. 160). If actions are ways in which agents express what constitutes them, then getting clear on the constitutive elements will tell us how actions may be classed as good or bad.

In Plato's account of the constitutive elements of the human soul, there are three parts: Reason, Appetite, and Spirit (R 439e-441c). Thus, in order for something to be classified as an action for a human being, it must unify all three (Korsgaard, 1999, p. 6). According to Plato's account, Appetite proposes an action; Reason considers whether one ought to act upon the action, and Spirit acts upon this deliberation (Korsgaard, 1999, p. 7). For Korsgaard, this can be thought in analogy to the deliberative procedure of a city deciding to adopt a new law (Korsgaard, 1999, p. 7). The deliberative nature of the procedure confers normativity upon the outcome of the procedure.

Laws, for example, may be deemed to be constitutional *if and only if* they are passed by the right legislature. For example, a law which is passed by the Constitutional Court in South Africa is constitutional law just in the case it is passed by the Constitutional Court. The procedures of the Constitutional Court give the law its status as constitutional law in virtue of the procedures of the court. Similarly, on Plato's account, an action is given its status as an action just in case it is endorsed by the deliberative procedures of the human soul (Korsgaard, 1999, pp. 9–10). In both the legal deliberative procedure and the deliberative procedure of the human soul, the

procedure is able to confer normativity on the outcome *because* the procedure constitutes internal standards for the outcome. Without the legal procedure, there is no law and without the deliberative procedure, there is no human action (Korsgaard, 1999, p. 10).

Korsgaard sees Kant's account of morality in the same manner. According to her view, Kant's account of what it is to be a rational being tells us what a good or bad action is (Korsgaard, 1999, p. 11). For Kant, to be a rational being is to act under the idea of freedom (G 4:448). That is, insofar as one takes oneself to be a rational being, one should also take oneself to have a free will. However, the free will must still be determined by *something* – a law. However, such a law cannot be an external cause, if one is free (G 4:448). Kant, thus insists that the law that the free will is determined by must be a law given by the will itself – self-legislation (G 4:448; Korsgaard, 1999, p. 11). For Kant then, the ultimate principle of right action just *is* the law of the free will. He calls this internal constitutive standard of action the categorical imperative (G 4:414).

Kant, similarly to Plato, has a constitutional model of action, where actions are determined by the constitution of the agent. According to Kant's philosophical psychology, inclination presents you with a maxim to be acted upon, and then reason considers whether to act in accordance with inclination, and an action is thus legislated (Korsgaard, 1999, p. 12). We can consider the case of deciding to eat an ice cream cone. Upon the pleasant sight of an ice cream cone, an inclination arises within one to take the means to eat the ice cream cone. That is, the pleasant sight of the ice cream gives one the maxim: take means M to eat the ice cream.

Given that one has a free will, it is open to one whether or not to act upon this maxim. Thus, the decision to take that maxim as your law is to see whether it passes the categorical imperative. On this view, the action of taking the means to eat an ice cream is *only* attributable to you insofar as you freely endorse the aforementioned maxim as a law for yourself. We can thus see

how an action can be seen as *your* action insofar as the maxim which produces the action stems from one's autonomy (Korsgaard, 1999, p. 14).

How then does this view give us a notion of good or bad actions? Both for Plato and Kant, actions are good actions and actions are bad actions insofar as they fail at being actions *simpliciter*. Justice for Plato and universalizability for Kant, are internal to the idea of action, and thus the normative status of an action is internal to whether or not it meets the standard of Justice or Universalizability (Korsgaard, 1999, p. 15). Good actions and bad actions are then not different in kind, but rather in degree. The greater the extent to which something is an action, the better it is as an action *qua* action and the greater the extent to which something fails to meet the internal standards of an action, the worse it is as an action *qua* action. The question of an external source of normative standards makes no sense according to constitutivism. Bad actions and good actions are thus the same sort of activity, bad actions are just defective actions (Korsgaard, 1999, p. 15).

Additionally, we can see how the constitutional model also gives us a standard for determining whether or not one is a good or bad agent. If universalizability is internal to the idea of an action, as per Kant, then any agent that fails to act with the principle of universalizability fails to be an agent and thus is a defective agent (Korsgaard, 1999, p. 15). Korsgaard explains this with a helpful example based on the character Harriet from Jane Austen's novel *Emma* (1815):

> Imagine a person I'll call Harriet, who is, in almost any formal sense you like, an autonomous person. She has a human mind, she is self-conscious, with the normal allotment of the powers of reflection. She is not a slave or an indentured servant, and we will place her — unlike the original after whom I am modeling her—in a well-ordered modern constitutional democracy, with the full rights of free citizenship and all of her human rights legally guaranteed to her. In every formal legal and psychological sense we can think of, what Harriet does is *up to her*. Yet whenever she has to make any of the important decisions and choices of her life, the way that Harriet does that is to try to figure out what Emma thinks she should do, and then that's what she does […]This is autonomous action and yet it is *defective* as autonomous action. Harriet is self-

> governed and yet she is not, for she allows herself to be governed by Emma (Korsgaard, 2009, p. 162).

This example shows us, that according to constitutivist model, agents are better or worse *qua* agents if the principles upon which they act reflect the *internal* standards of the deliberation process of an agent. Harriet is a better agent to the extent that her deliberation process reflects the standards of universalizability, and worse as an agent to the extent that her deliberation process does not reflect this.

We can also think of the deliberation process, on the constitutivist's account, as an ordering of values which in turn determines the principles upon which one could act (Barandalla, 2023, p. 212). Deliberation is thus, also a creation of a value system that you take to be determinative of who you are (Korsgaard, 2009, p. 25). If any action is to be attributable to you, then it must flow from your constitution. When it flows from your constitution, then it expresses your value system. One is unified as an agent insofar as the values expressed in action are the values which reflect the nature of one's constitution. As Barandalla puts it, "the space between your commitment and whether you realise it, is, on constitutivism, the normative realm" (Barandalla, 2023, p. 213).

2.2 Murdochian Freedom as Constitutivist Freedom

For Barandalla, a constitutivist construal of the substantial individual enables us to resolve the problem of freedom in Murdoch's account of moral agency. In order to see how it does so, a short restatement of the problem of moral freedom is necessary (Barandalla, 2023, p. 213):

1.) The Freudian forces are part of the individual.

2.) The individual is free from the Freudian forces.

Barandalla thinks that the constitutivist account of freedom enables us to overcome the apparent contradiction between 1 and 2. That is, constitutivism allows us to claim both that the Freudian

forces are elements of an individual *and* that an individual may be free from them (Barandalla, 2023, p. 214). How does it do so? The constitutivist framework allows for this by construing the Freudian forces and the attraction to the Good as *values* (Barandalla, 2023, p. 214). If we construe them as values, then we can see them as both "making claims" on the will (Barandalla, 2023, p. 214). That is, Freudian forces and the attraction to the Good can be seen as possible elements of our constitutive value system. If our attraction to the Good is higher than the Freudian forces in our value system, then this is because we have ordered our value system in such a way. For the constitutivist one is not just identified with one's values, but also with how one *orders* the values in your value system (Barandalla, 2023, p. 214).

To further make sense of this idea, Barandalla asks us to consider the nuclear family as an explanatory example of a legislative unit (Barandalla, 2023, p. 214). The nuclear family is composed of many different parts, usually a pair of adults and a couple of children. The adults are ordinarily considered to be the legislators of the nuclear family. That is, the adults make decisions for the family and not the children. However, the children are still constitutive elements of the nuclear family. There is no nuclear family without the children and thus no family decisions. So, the children are still constitutive elements of the decisions of familial legislation. That is, familial legislation, as an activity for a nuclear family, is only *possible* with the existence of children. *Mutatis Mutandis*, for the Freudian forces in the free substantial individual.

The free substantial individual contains both the attraction to the Good and Freudian forces. Whenever the free substantial individual acts, both the attraction to the Good *and* the Freudian forces are there, but the attraction to the Good is ranked higher than the Freudian forces in their value system, and thus the action taken by the free substantial individual reflects this. The constitutivist can make sense of this precisely because the constitutivist claims that freedom consists in nothing but the expression of oneself (Barandalla, 2023, p. 214). Thus, when one

'suppresses' one's Freudian forces they do not disappear, but rather they are, as Barandalla puts it, *subjugated* to one's attraction to the Good in one's activity (Barandalla, 2023, p. 214).

This account of the relationship between the Freudian forces and the free substantial individual meets the conditions of the concept of freedom. Previously, it was suggested that the logic of freedom is relational. That is, one is free from some constraint in order to do something. The free substantial individual, according to the constitutivist reading, is free from their Freudian forces in the expression of themselves in action and in epistemic activity (what Murdoch would characterise as part of inner life). One is free from them to express oneself *because* they do not play a role in one's actions. The substantial individual by way of instilling the value of the attraction to the Good over and above the Freudian forces, becomes free from the influence of the Freudian forces in the expression of themselves. Thus, the Freudian forces are still part of you when you exercise your freedom, *but* you are free from their influence (Barandalla, 2023, p. 215). The apparent problem of freedom dissipates.

In defending this reading, Barandalla highlights several features of Murdoch's morality and shows how they are amenable to the constitutivist account of moral agency. Barandalla lists the following features of Murdoch's moral philosophy as being amenable to the constitutivist reading of Murdoch (Barandalla, 2023, p. 215):

1.) Substantive account of the individual.
2.) Expression of self in both epistemic activity and overt activity.
3.) Historical and contextual nature of the individual: 'fabric of being'.
4.) Close relationship between our inner life and our overt actions.
5.) The expansive conception of the moral realm.

Barandalla rightfully recognises that key aspects of Murdoch's substantial picture of moral life can be captured by the constitutivist account of moral action (Barandalla, 2023, p. 215). The first feature, that of the substantial individual, can be captured by the fact that the constitutivist account allows one to suggest that the Freudian forces are a necessary part of the substantive individual. Thus, the constitutivist also agrees with the somewhat pessimistic picture of human nature that Murdoch endorses (OGG pp. 343–345).

The second feature, the expression of ourselves in epistemic activity, is also on the constitutivist account regarded as no less essential to who we are as individuals than our overt actions.[4] The constitutivist identifies our moral activity with whatever is involved with ordering and ratifying our values. Thus, all the epistemic activity that is implicitly involved with ordering and ratifying our values is moral activity. On the constitutivist account, we are responsible for the values which make up our value system, and thus much like Murdoch's substantial individuals are responsible for the 'slow delicate processes of imagination and will which have put those values there' (DPR p. 200).

The third feature, the 'fabric of being'(our personal history of imagining and willing) is much more difficult to capture on the constitutivist account (Barandalla, 2023, p. 215). Barandalla notes that there is a *prima facie* tension between Murdoch's historical notion of our 'fabric of being' and constitutivism (Barandalla, 2023, p. 215). According to the consititutivist, one is *identified* as a practical agent with the expression of one's values in each decision. Thus, it appears that one cannot have a historical self, if one is only identical to the latest expression of one's values (Barandalla, 2023, p. 215).

[4] Barandalla seems to use the term epistemic activity to refer to all the cognitive and evaluative work that an individual may do in their minds, it appears that she wants to contrast this to overt activity – publicly observable activity. I have mentioned earlier that Murdoch uses the term 'inner life' to refer to much of the activity that Barandalla would use the term epistemic activity to refer to. Nothing of note turns on this.
(Barandalla, 2023, p. 208; IP p. 311).

Barandalla suggests that such a worry is misplaced, as the constitutivist account of the self is very much dependent on the values, and the activity of ordering those values, which came before the current order of values in the value system (Barandalla, 2023, pp. 215–216). The historical ordering of values, and the values inherent them, have a claim on one's will insofar as they play a role in shaping the epistemic and overt actions that one does. One's history of values, and the decisions that reflect those values, are always binding on you as long as one has not rejected them (Barandalla, 2023, pp. 215–216; Korsgaard, 2009, p. 23). As Barandalla suggests, the continual self-constitution through the ordering of values seems to resemble Murdoch's claim that the nature of moral progression is akin to creating 'pictures' of oneself and then coming to resemble those 'pictures' (Barandalla, 2023, p. 217; M&E , p. 75).

The fourth feature, about the importance of the inner life for determining overt action, also seems to be captured by constitutivism. The constitutivist holds that we are constituted through our self-conceptions (the value system with which we identify), which in turn determines which actions we take to be available to us (Barandalla, 2023, p. 216). Thus, the ways in which we conceive of ourselves determines which overt actions we take. For the constitutivist, the moral agent is constantly ordering and ratifying their values which requires our moral imagination. The connection between inner life and outer life is thus extremely intimate according to the constitutivist view.

Finally, the pervasiveness of the ethical element to all our activities is also captured on the constitutivist account because morality is a subset of normativity, and since the normative realm is involved in *all* aspects of epistemic and overt activity, then it follows that the constitutivist can account for the expansive conception of moral life that Murdoch holds onto.

The constitutivist picture of the moral agent seems to appear much like Murdoch's: an individual struggling against forces within themselves in the expression of their values

(Barandalla, 2023, p. 217). This constitutivist moral agent is not at all like the 'man of modern philosophy' who is the target of Murdoch's critique of much of moral philosophy. For Barandalla, there is a philosophical irony here, Murdoch thought Kant was the enemy of an adequate account of moral life, but returning to his moral philosophy is in fact what is necessary to make sense of the substantial individual's freedom from Freudian forces (Barandalla, 2023, p. 218; SGC p. 365).

There is a lot to be said for this account. I think that *if* we construe Murdoch's concept of freedom the way Barandalla does, and *if* we read Murdoch's account of the individual in a certain way, then we may dissolve the problem of freedom in the manner that she suggests. However, Murdoch's picture of moral life is ultimately incompatible with constitutivist account of moral agency. In the next chapter, I will introduce some of the key features of Murdoch's broader commitments in moral philosophy and show how they do not allow us to construe Murdoch's account of moral agency in a constitutivist manner.

Chapter Three: Constitutivism, Agency, and Love

'We have to try to cure our faults by attention and not by will.' – (Weil, 2003, p. 116).

At the end of the previous chapter, I suggested that the constitutivist reading of Murdoch can both solve the problem of freedom in Murdoch's account of moral agency and also capture many of the key features of Murdoch's moral philosophy. However, I believe it only does so by giving an account of moral freedom that is irreconcilable with her broader philosophical commitments. In this chapter, I will first outline Murdoch's views on moral agency in the context of her broader philosophical views, then I will show that a constitutivist reading of Murdoch's account of moral freedom is irreconcilable with her views about moral agency.

3.1 The Context of Murdoch's Account of Moral Agency

Murdoch's broader philosophical commitments influence her account of moral agency, and more than that, they give us a particular picture of what moral freedom amounts to. This picture may share many of the same features of the constitutivist account but has a very different frame. In Anil Gomes' reading of Murdoch's moral philosophy, he identifies three core claims (Gomes, 2022, p. 143):

1.) Realism: There are objective moral truths and properties which are both practical and theoretical in nature (IP pp. 329–330).
2.) Knowability: We have epistemic access to moral truths and properties which are both practical and theoretical in nature (IP p. 332, OGG p. 347).
3.) Concept Involvement: Our moral truths and moral properties are mediated by moral concepts (IP pp. 324–325).

Murdoch's commitment to realism, is the first major difference between the constitutivist account of moral agency and Murdoch's. It was noted in the previous chapter that constitutivism not only gives an account of what moral agents are up to but also gives an account of normativity as such. In contrast to this, Murdoch's account of moral activity does not attempt to do this. This difference, as I will later spell out, has radical consequences for the different manner in which they account for the features of moral life. But first, it is necessary to see how we ought to construe Murdoch's account of moral realism.

Murdoch's moral realism is not like many ordinary conceptions of moral realism. Standard conceptions of moral realism typically hold that moral truths are objective *and* mind-independent of individuals (Sayre-McCord, 2023). Murdoch's account of moral reality is different in this respect. Whilst there are many competing accounts of Murdoch's moral realism[5], Murdoch does seem to hold that moral reality is in some sense transcendent from the viewpoint of the moral agent (OGG p. 347, IP p. 332). However, this commitment does not entail that moral reality ought to be construed as mind-independent.

We can make better sense of Murdoch's account of the relationship between moral reality and minds, by looking at how she construes the objectivity of moral claims. For Murdoch, there are objective moral truths and properties *in virtue* of the relationship between the individual and the reality which they perceive (Mason, 2023, p. 659; OGG p. 353).[6] On this account, the objectivity of moral claims and properties arises out of a relationship between the individual and the reality which they perceive. Whilst this latter claim may be construed along standard

[5] For some different readings of Murdoch's realism, see (Antonaccio, 2012; Hopwood, 2018; Mason, 2023).
[6] McDowell, inspired by Murdoch, develops an account of value that resembles this idea (McDowell, 1998). See (Broackes, 2012, pp. 8–12) for a discussion of the relationship between McDowell's work and Iris Murdoch's. Additionally, Murdoch's commitment to a moral reality which emerges out of virtuous perception have led some philosophers to read her as a response-dependent theorist (Jordan, 2014). See (van Roojen, 2015, p. 127) for a critical introduction to Response-Dependence theories in metaethics. However, I am resistant to reading Murdoch along these lines, and also to reducing her metaethical views to those of McDowell's, primarily because of Murdoch's distinctive views on the nature of truth. For an argument against reducing Murdoch's moral realism to accounts like McDowell's see (Mason, forthcoming, pp. 17–20).

realist readings, Murdoch's conception of reality *simpliciter* as a fundamentally *normative* concept makes her conception of moral realism distinctive. There is a 'reversal' of the relationship between metaphysics and ethics in Murdoch's view (Mason, 2023, p. 661). The standard moral realist holds that metaphysics is prior ethics, whereas for Murdoch metaphysics is fundamentally an ethical endeavor.

The second philosophical commitment to knowability, suggests that moral reality is epistemically accessible to us and that through loving attention we gain both practical and theoretical knowledge of moral reality. This commitment must be seen in relation to Murdoch's first commitment. If morality reality is knowable, then the fact that moral reality is fundamentally relational between the perceiver and object of perception helps explain this fact. Furthermore, Murdoch connects theoretical moral knowledge with practical moral knowledge (Gomes, 2022, p. 143; IP p. 333). Murdoch has a moral psychology in which one acts in accordance with one's vision (IP p. 329). Thus, theoretical moral knowledge entails practical moral knowledge.

The last commitment to conceptual involvement reflects Murdoch's views on the thoroughgoing relationship between reality and moral language. Murdoch, at various places in her work, suggests that moral language is expressive of our representations of reality (VCM pp. 82–83, MGM pp. 25–26). But this in fact understates the radical nature of the relationship between our moral concepts and reality that Murdoch endorses. For Murdoch, all the concepts with which we represent reality are *fundamentally* evaluative (MGM pp. 25–26). On this view, moral language is not only expressive of our representations of reality, but our representations of reality are irreducibly value-laden. Such a view has radical implications for our conception of our moral agency. This can be seen as part of Murdoch's commitment to what can be called a 'Platonic Theory of Concepts' (Setiya, 2013, p. 8).

According to such a view of moral concepts, we gradually obtain and develop our knowledge of moral concepts by beginning with a purely public concept and move towards a fuller 'possession' of the concept that is private (Setiya, 2013, p. 9). That is, we begin with a basic grasp of a moral concept and slowly begin to develop a greater degree of competence with respect to the concept which goes over and above its public use. We can imagine a child who watches *The Wizard of Oz* (1939) and begins to learn about the concept of courage by attending to the character development of the Cowardly Lion throughout the film. Despite the merits of the film, the concept of courage, is clearly impoverished in several respects, and if it were one's only reference it would give one a merely basic competence of its use. Throughout one's life, one would develop this competence and gain a deeper and richer conception of courage and, as Murdoch suggests, see the connection between the virtue of courage and other virtues (OGG p. 346).

On Murdoch's account then, as we develop competence with respect to our moral concepts, they also become concrete and more private (Bagnoli, 2012, pp. 222–223; IP p. 322). Our particular experiences of the display of courage, both in others and in ourselves, deepens in a manner that is uniquely tied to the historical nature of each individual (IP p. 322). Murdoch thus has a view of moral concepts that start out as public abstract universals that then through one's experiences become more concrete, private, and uniquely tied to the historicized nature of one's life (Hopwood, 2017, p. 263). This view of concepts informs her view of moral reasons. Murdoch argues against the predominant view that moral reasons, by their very nature, ought to be considered agent-neutral (VCM; Hopwood, 2017, p. 248). If moral concepts are tied to our own unique historical individuality, then the moral reasons that become available to us will correspondingly also be uniquely tied to us. On Murdoch's account, there may be moral reasons available only to us and to no one else in our situation (VCM p. 86).

The above views on realism, knowability, and conceptual involvement are the key features of Murdoch's views on philosophy with which one must interpret loving attention as the 'proper mark of an active moral agent' (IP p. 327). They jointly provide a framework from which to make sense of Murdoch's account of moral agency as these three commitments are 'unified' in loving attention (Gomes, 2022, p. 143). A proper understanding of what is involved in loving attention can thus help explain how it is that these three features distinguish Murdoch's moral agency from constitutivist moral agency.

3.2 Loving Attention

Murdoch's appeal to the metaphor of vision is no mere analogy. For Murdoch, the metaphor of vision cannot be translated without a loss of substance (SGC p. 363). The need to describe our moral agency in terms of the metaphor of vision reflects her unique form of moral realism (Gomes, 2022, p. 146). When Murdoch suggests that we cannot do without moral theorizing in relation to the metaphor of perception, it seems natural to interpret loving attention as involving the perception of a moral reality that is outside of oneself. After all, perception implies an object of perception. However, the metaphor of vision also suggests that the nature of the perceiver and the nature of the object of perception ought to be considered.

Thus, when thinking of how loving attention is the mark of our moral agency, it is important to conceive of 'agency' in the light of the metaphor of vision. Loving attention is a kind of "passive activity" (S. Panizza, 2022, p. 165; Weil, 1973, p. 194). It appears to involve both a receptivity to an external reality and a sustained moral effort to perceive reality as it really is. The importance of passivity is revealed by the fact that Murdoch suggests that moral effort is not always enough to have a true vision of reality (SGC pp. 366–369).

What Murdoch has to say about the nature of conscious experience during attention is instructive here. She gives the following example:

> I am looking out of my window in an anxious and resentful state of mind, oblivious of my surroundings, brooding perhaps on some damage done to my prestige. Then suddenly I observe a hovering kestrel. In a moment everything is altered. The brooding self with its hurt vanity has disappeared. There is nothing now but kestrel. And when I return to thinking of the other matter it seems less important. And of course this is something which we may also do deliberately: give attention to nature in order to clear our minds of selfish care (SGC p.369).

There is no act of will here. The kestrel captures one's attention and one's mind is drawn away from egoistical self-consoling. Our receptive nature is thus essential to an account of our moral agency. Of course, it would be puzzling to construe this receptivity as an aspect of our agency. It is difficult to think of receptivity as a kind of activity. What must be recognized is that our receptivity is not merely an enabling condition for our moral agency, it also gives us a grasp of moral reality that no moral effort could have delivered. For Murdoch, anything which changes our quality of consciousness away from our natural selfishness towards an objective view of reality is to be regarded as virtuous (SGC p. 369).

Thus according to Murdoch's account, conscious experience is deeply significant for us (NP p. 51). In conscious experience, new values are discovered and other values are put into context. 'Experiences announce themselves' (NP p. 52). In her view, the conscious awareness of anything outside oneself is the beginning of our moral progress (SGC p. 370). The receptive and sensitive nature of our consciousness is thus paramount to the possibility of moral progress.

This account of moral development through loving attention entails a very different picture of moral agency from the constitutivist account, even though they may share many of the same features. To see how the constitutivist view and Murdoch's view differ we should look again at the five features that Barandalla suggests the constitutivist can account for, and see how they might differ in light of Murdoch's wider philosophical views.

3.3 Constitutivist Agency vs. Murdochian Agency

According to Barandalla the following features can be captured on a constitutivist reading of Murdoch (Barandalla, 2023, p. 215):

1.) Substantive account of the individual.

2.) Expression of self in both epistemic activity and overt activity.

3.) Historical and contextual nature of the individual: 'fabric of being'.

4.) Close relationship between our inner life and our overt actions.

5.) The expansive conception of the moral realm.

In order to tease out the important differences between the constitutivist account of moral agency and the Murdochian account of moral agency, it may be helpful to return to the case of M and D in order to make sense of the differences in accounting for the above features. Barandalla's reading of Murdoch's observation about the M and D case especially brings out some of the core differences. According to Barandalla, Murdoch's observation that M has been morally active in coming to change her views about D, does entail that our moral acts are not reducible to overt actions, but it does *not* follow that there is more to moral activity than willing (Barandalla, 2023, p. 208).

For Barandalla, we can think of the epistemic activity (inner mental activity) that M engages in as a kind of willing (Barandalla, 2023, p. 208). The scope of what can be the object of our will is enlarged according to her account. We can not only will overt actions, but we can also will to think differently – to pay attention. However, Murdoch's notion of the 'fabric of being', which she identifies as another constitutive element of individuals, also needs to be accommodated in this picture somehow (IP p. 316).

For Barandalla, this can be accommodated by thinking of the 'fabric of being' as a derivative of our history of willing (Barandalla, 2023, p. 209). That is, the fabric of being can be seen as the values and conceptions of the world that emerge out of the history of our actions. Each of our thoughts, actions, and value judgments form a continuingly constructed cognitive perspective which Murdoch calls our 'total vision of life' (VCM p. 80). This constructed conceptual scheme is based upon our history of 'willing' – in the broader sense – and it is also the context from which we think, act, and make further judgements about the world (Barandalla, 2023, p. 209).

This enlarged conception of willing allows Barandalla to correctly highlight that, for Murdoch, our epistemic activity is fundamentally evaluative (Barandalla, 2023, p. 209; DPR pp. 199–201). The value judgments implicit in our epistemic engagement with the world not only inform the overt actions which we take, but also form the context in which we determine which actions are possible for us. The 'structures of value' which we build up with our epistemic engagement with the world thus determines how it is that we act (Barandalla, 2023, p. 209; IP p. 329).

According to Barandalla's account, the substantive picture of the individual, the expression of the self in epistemic and overt activity, as well as the 'fabric of being', can all be reduced to the will and the value system which determines its actions. Thus, when accounting for the moral agency of M we get a different picture than the one that Murdoch provides. M engages in epistemic activity and comes to change the order of values in her value system such that the attraction to the Good is higher than the Freudian forces on M's value system. M, thus now no longer views D as 'bumptious', but rather comes to view her as 'gay' (IP p. 317).

However, whilst the constitutivist can argue that the moral transformation of M occurs in virtue of epistemic activity, this transformation is not necessarily in relation to a change in one's conceptual scheme. Strictly, all that the constitutivist is committed to with respect to moral

transformation is a change in the order of values in one's value system. That is, according to the constitutivist moral transformation can occur without a change in one's concepts – all one needs is a different order of values. However, according to Murdoch, M's moral transformation necessarily involves a change in her conceptual scheme. That is, in her cognitive relationship to reality.

The constitutivist might suggest that in order to change one's value system, one needs a change in one's conceptual scheme. The suggestion would be that the epistemic activity involved in M coming to view D justly, *necessarily* involves a change in M's conceptual scheme. On this view, the struggle of coming to order and unify one's values, where the attraction to Good is higher than one's Freudian forces, necessarily involves a change in one's conceptual scheme. However, this reply fails to capture what is distinctive about Murdoch's account of moral progress. For Murdoch, moral progress consists in a greater knowledge of an individual's reality in the light of the Good (Hopwood, 2018, p. 486; IP p. 323). The moral progress of M, is on Murdoch's view, a movement towards knowledge of an *individual's* reality.

The moral progress of M ought not be seen as a change in her value system, but rather as coming to see D as she *really* is. We should, I think, take Murdoch at her more radical suggestions that moral freedom consists in 'exist[ing] without fear and perceiving what is real' (DPR 201). This suggestion points to Murdochian moral freedom as consisting of two components: (1) freedom from the influence of Freudian forces, and (2) knowledge of the reality of individuals.

The constitutivist account is unable to satisfactorily account for both components of Murdochian moral freedom. According to the constitutivist, M is morally free, with respect to D, insofar as M is free from the influence of Freudian forces in her epistemic relationship to D. However, this does not adequately capture the second aspect of Murdochian freedom which is the knowledge of the reality of individuals. The constitutivist may retort that having the value

of the attraction to the Good higher in one's value system than the value of the Freudian forces just *is* what it means to have knowledge of the reality of individuals. This retort fails on a couple of grounds, the first is that, as Murdoch has noted, moral effort is not always enough for accurate vision. It may be the case that one's attraction to the Good is ranked higher in one's value system than Freudian forces, but one may still not obtain knowledge of the reality of individuals. In this case, having the right constitution (value system) is not enough for moral freedom. Suppressing one's Freudian forces is only half the job, one must also have knowledge of the reality of others.

The other and perhaps more concerning worry for the constitutivist, is that Murdoch suggests that the accurate perception of others is a state of consciousness – an experience (OGG p. 354, IP pp. 313, 317). In particular, it is an experience of the individual reality of *another*. The constitutivist account renders freedom as consisting not so much in the individual's relationship to reality, but rather to do with the relationship between the forces and values within the individual. On the constitutivist view, morality is primarily about a struggle for integrity *within* an individual (Korsgaard, 2009, p. 7). The constitutivist thinks of moral progress as primarily about having the right values within one's value system, not about having the right conceptual framework (relationship to reality).

For Murdoch, standing in the correct epistemic relationship to the reality of another individual is what is necessary for moral freedom (Samuel, 2021, p. 358). For Murdoch having the right values and the activity of placing these values in the right order is necessary for free action, but knowledge of the reality of other individuals is also necessary. Murdoch appears to have two concepts of freedom at work in her moral philosophy. On the one hand, she takes freedom to be a kind of ability to 'look again' – loving attention unencumbered by Freudian forces – and on the other hand, she takes freedom to be a state of consciousness (IP p. 313, OGG p. 354).

The two aspects of moral freedom for Murdoch: (1) freedom from Freudian forces and (2) knowledge of an individual's reality, are also reflected in her view of loving attention as central to our moral agency. Whilst the constitutivist reading of Murdoch may suggest that loving attention is central to reordering one's value system, it also makes knowledge of the reality of others secondary to achieving integrity amongst one's value system.

The centrality of loving attention to our moral agency reflects the importance of knowledge in her account of our moral progress. The core difference between the constitutivist reading of our moral agency and Murdoch's is that moral progress for the constitutivist fundamentally concerns the relationship between values *within* the individual, whilst, for Murdoch, moral progress concerns the individual's relationship to the reality of other individuals.

Barandalla reads Murdoch as suggesting that freedom consists in freedom from Freudian forces in order to engage the world with loving attention (Barandalla, 2023, p. 211). But this reading fails to capture Murdoch's suggestion that freedom is also an *achievement* of our loving attention (OGG p. 354). Our ability to engage with the world by loving attention is a *means* to achieve knowledge of the reality of individuals – not the end of our freedom.

Although many of the features of Murdoch's moral philosophy may be captured on a constitutivist reading, it fails to capture how these features fit into Murdoch's broader philosophical commitments which fundamentally concern an individual's relationship to reality. In the final chapter, I will suggest that both (1) freedom from Freudian forces and (2) knowledge of the reality of individuals are necessary to make sense of what moral freedom means for Murdoch. Additionally, an adequate understanding of how these two components relate to each other will enable a dissolution of the problem of freedom for Murdoch.

Chapter Four: Murdochian Moral Freedom

'What gives truth to the things known and the power to know to the knower is the Form of the Good. And though it is the cause of knowledge and truth, it is also an object of knowledge.' – Plato (R 508e)

In the last chapter, I suggested that Murdoch's conception of freedom can be split into two constitutive elements: (1) freedom from Freudian forces and (2) knowledge of the reality of other individuals. Whilst Murdoch appears to not keep these two elements of freedom distinct from each other, this distinction is in fact crucial to understanding what moral freedom amounts to on her account, and furthermore it is also crucial to dissolving the problem of freedom in Murdoch's account of our moral agency. In this chapter, I will suggest that we ought to construe Murdoch's free individual as having a 'purified' consciousness. The process of purifying one's consciousness consists in the transformation of the substantial elements of the individual so that the individual perceives others in a just and loving manner. Ultimately, moral freedom for Murdoch does not consist in freedom from some particular elements of an individual, but rather this freedom consists in a transformation of the individual.

4.1 Murdoch on Plato's Allegory of the Cave

At several points throughout *Metaphysics as a Guide to Morals (*1993) Murdoch presents Plato's allegory of the cave as a way to make sense of our moral progress. Murdoch reads the allegory of the cave as a progression in our states of consciousness (MGM p. 177). In *The Republic*, Plato uses the analogy of prisoners breaking free from a cave as an analogy to make sense of the process of education and the effect it has on an individual (Meinwald, 2016, p. 236; R 514a). In the analogy, prisoners are chained with their heads facing towards the cave wall where shadowy images are projected (R 514a). The shadowy images are the product of puppet figures that are illuminated by a fire (R 514b). When a prisoner is freed from their chains,

they make their way out of the cave and into the 'real' world illuminated by the blinding light of the sun (R 515d).

Murdoch suggests that we should make sense of our moral progress analogously to the process of the freed prisoner moving from the fire-lit shadowy caves to the world illuminated by the sun (MGM p. 183). In Murdoch's reading of the analogy, the fire which produces the false images is the egocentric 'self' (SGC p. 382). When the prisoners are freed and begin to make their way out, they see the 'self' as the source of their false view of reality (SGC p. 383). But crucially for Murdoch, our moral progress does not stop here (SGC p. 383). For her, our moral progression occurs through different levels of our conscious experience and awareness of the world and ends with the contemplation of the Good (SGC p. 383; MGM p. 183).

Thus, this process of purification is not merely the suppression of Freudian forces which distort our vision, but also coming to see the reality of individuals in the light of love, truth, and justice (OGG p. 354). Murdoch's conception of moral freedom is thus not just a 'mode of reflection' in which we perceive others just and lovingly, but also a state of consciousness in which we have knowledge of others – the *aim* of our moral progress (VCM p. 95). This state of consciousness, in which one is 'unselfed' (where Freudian forces are suppressed) is not so much a loss of the individual personality, but rather a transformation of the individual. This transformation should *not* primarily be characterized as a change in the value system of an individual – although it is also that. This transformation should rather be construed as a change in our consciousness. Murdoch says the following about it:

> Platonic philosophy and some religious positions take ordinary egoistic consciousness to be a veil which separates us from the order and true multiplicity of the real world: reflectively understood this is obviously true. Moral progress (freedom, justice, love, truth) leads us to a new state of being. This higher state does not involve the ending but rather the transformation of the 'ordinary' person and world. There is a false unity and multiplicity and a true unity and multiplicity. There is the selfish ego surrounded by dark and menacing chaos, and the more enlightened soul perceiving the diversity of creation in the light of truth (MGM p.165).

Murdoch thus represents moral progress not as involving mere suppression of the selfish ego, but the transformation of forces which give rise to the egocentric ego. I think a return to considering what is involved in the nature of loving attention is needed to make sense of both the nature of moral transformation and Freudian forces.

4.2 The Metaphysics of Attention and Freudian Forces

Murdoch says quite a bit about the nature of 'unselfing' (suppression of Freudian forces) through loving attention, but also different interpretations of Murdoch have suggested many different ways to make sense of this 'unselfing'. In her reading of Murdoch's views on attention, Panizza suggests two possible interpretations of 'unselfing': (1) the 'Tame View' (hereafter TV) according to which the substantial self remains in loving attention, and the (2) 'Radical View' (hereafter RV) which suggests that loving attention consists in a complete loss of the substantial self (S. C. Panizza, 2022a, p. 64). As we have seen in Chapter One, Barandalla seems to think that something like RV is incompatible with the traditional conception of freedom precisely because it would mean that the substantial self could not be the same self which is free (Barandalla, 2023, pp. 211–212). If one grants that RV is incompatible with the ordinary conception of freedom, then an account of TV must be the only way of making sense of our moral freedom.[7]

If we are to make sense of moral freedom according to TV, how ought we do so? There are many ways to make sense of what happens to the substantial self in loving attention. One such suggestion is that we can draw a distinction between the 'ego' and the 'self' (Meszaros, 2016; S. Panizza, 2022, p. 163). According to such a suggestion, the ego is to be identified with the fantasy which obscures our vision of reality, but the substantial 'self' is still present when we

[7] I do not necessarily think it is implausible to make sense of moral freedom without a substantial self. Though, I do think it is implausible to read Murdoch in this manner. See (S. C. Panizza, 2022b) for a defense of the Radical View reading of Murdoch.

are freed from Freudian forces. That is, our 'fabric of being' (conceptual scheme and historical self) is still part of the free individual, but what is lost are the egocentric forces.

Translating this solution to Barandalla's terminology, we can now think of the Freudian forces as a *non-essential* element of the unfree substantial individual. That is, the Freudian forces are still elements of the substantial individual, but one may maintain one's substantial identity without them. However, Barandalla thinks that such a solution would be implausible as it would suggest that an individual who is free from Freudian forces, would be left only with their good constitutive parts, such as their attraction to the Good and their will, and thus it would allow for a 'too rosy picture of human beings' (Barandalla, 2023, p. 212).

I think this worry arises from Barandalla's strange, perhaps theoretically driven, interpretation of Freudian forces and the attraction to the Good as values (Barandalla, 2023, p. 214). If we hold that Freudian forces are not values, but rather aspects of the substantial self that normally motivate and dispose us to engage with the world in an egocentric manner, then we can think of the transformation of these forces as part of our moral transformation. This is what the 'purification' of consciousness amounts to. The Freudian forces that once motivated and disposed us to engage with the world in an egocentric manner are transformed into a disposition to see the world in a loving and truthful manner.

To make sense of this moral transformation, let's return to the example of M and D. When M chooses to 'look again', to pay loving attention to D, she exercises her freedom to suppress the Freudian forces within her and to see D in a just and loving manner. When M comes to see D as no longer 'noisy' but rather as 'gay' this can be read as a transformation of the quality of consciousness of M (IP p. 313). 'Noisy', in this sense, is an egocentric description of D. D appears as an annoyance to M precisely because she views her under the guise of self-concern. When M looks at her again and sees D as 'gay', she does so through loving attention. What

changes? There is a transformation of the usually egocentric Freudian forces through loving attention such that M can see D both accurately and lovingly. M is no longer motivated and disposed to see D as 'tiresomely juvenile' but rather she is motivated and disposed to see her as 'delightfully youthful' (IP p. 313).

Crucially M's moral freedom does not consist in the loss of her personal history and context, but rather in a transformation of the Freudian forces within her, such that she comes to see D accurately. This thought is in line with Murdoch's commitment to conceptual involvement and development in our moral progress. If our moral concepts are always informed by our particular personal experiences, then our achievement of moral freedom – a just and loving vision of others – will also similarly be unique to us. Murdoch suggests that different virtuous individuals may have different views of moral reality that are *equally* just (Mason, 2023, p. 667; OGG p. 347).

A commitment to the aforementioned claim is only possible if free individuals can have different, but equally adequate, conceptual schemes with which they have knowledge of the reality of other individuals. Murdoch is thus committed to the substantial view of the free individual because each free individual would be free in a manner that is unique to them. 'M's activity is peculiarly her own' (IP p. 317). The only question then is, how ought we construe such an individual? Barandalla thinks that in order to capture this substantial individual, we ought to think of Freudian forces as an essential and constitutive element of even the free individual. However, we need not construe the substantial individual this way.

I have suggested that we can think of the substantial individual as genuinely free from the Freudian forces if we construe Freudian forces as a non-essential element of the substantial self. When the reality of other individuals is obscured, the usually self-concerning motivational forces and dispositions can be characterized as Freudian forces. If we think of the Freudian

Forces in such a manner, then we can 'lose' the Freudian forces without any threat to our substantial nature. That is, once we 'purify' our consciousness, and change the orientation of the motivational and dispositional forces from ourselves to other individuals, we are free to come to know other individuals in a just and loving manner.

4.3 The Logic of Freedom, Briefly Revisited

I have suggested that we think of Murdoch's conception of freedom as involving both (1) freedom from Freudian forces and (1) knowledge of the reality of other individuals. This can be further explicated by returning to Barandalla's construal of the ordinary conception of freedom:

> X is free from Y to Φ just in the case that Y is not a part of X qua Φer & Y is not expressed in X's Φing (Barandalla, 2023, p. 206).

If we adopt the two-fold reading of Murdoch's freedom, where it is both activity and a state of consciousness, then each constitutive element can be seen to be compatible with this relational conception of freedom. The first element, (1) freedom from Freudian forces, does this through loving attention. One is free from the Freudian forces insofar as they do not play a role in one's attention to individuals when one suppresses the Freudian forces. This is achieved by the substantial individual with their own particular 'fabric of being' (conceptual scheme and historical self). The second element, (2) knowledge of the reality of other individuals, is also compatible with this notion of freedom. One is free from Freudian forces in virtue of one's knowledge of the individual reality of others.

This twofold conception of moral freedom, allows one to capture the connection between moral freedom and knowledge which Murdoch insists upon: 'Freedom, itself a moral concept and *not just*[emphasis added] a prerequisite for morality, cannot be separated from the idea of

knowledge' (IP p. 330). For Murdoch, our knowledge of the reality of others and our ability to engage in just and loving attention are both important and can also be logically teased apart. For the Kantian, moral freedom and moral activity are identical. For Murdoch, moral activity is a *means* to achieve moral freedom. We exercise one kind of freedom, the freedom of moral imagination through loving attention, in order to achieve another kind of moral freedom – knowledge of the reality of others. Freedom is not just activity but consists in knowledge as well.

Conclusion

"For now we see through a glass, darkly" - 1 Corinthians 13:12, King James Version

Iris Murdoch's account of moral progress is both compelling and yet also tragic. The account is compelling because it restores a certain kind of dignity to our inner life. According to Murdoch's account, what we spend most our lives doing (thinking, imagining, doubting) is no longer to be thought of as mere idle mental chatter, but rather as carrying supreme moral significance. However, the account is also tragic because our inner lives are, as we well know, all too often the place of refuge for our narcissistic, anxious, and self-aggrandizing thoughts.

Murdoch gives a realistic account of the moral lives of substantial individuals who are submerged in a moral reality that, despite their best efforts, they will never come to completely know. We are only truly free when we have purified the self-centred nature of our consciousness, and have attained knowledge of the reality of others under the light of the Good.

Barandalla is right to think that the Freudian forces are an element of the substantial individual, but wrong to think that they are an essential element. According to Barandalla, the free Murdochian individual cannot be free from Freudian forces without losing their substantial identity. The loss of a substantial picture of the individual is at the heart of Murdoch's worries about the picture of moral life presented to us by most other moral philosophers. Barandalla is correct to maintain that Murdoch's view of moral freedom must accommodate the substantial individual.

However, Barandalla's attempt to accommodate this substantial individual by suggesting that Murdoch's individual is identical to the individual given to us by constitutivist accounts of agents is misguided. Moral progress for the Murdochian individual and moral progress for the

constitutivist individual is fundamentally different. The constitutivist individual becomes a better moral agent by unifying their values. In contrast, the Murdochian individual becomes a better moral agent by the greater knowledge they have of the reality of other individuals. Barandalla is right to point out the importance of epistemic activity for both accounts, but this is not enough to suggest that the Murdochian individual and the constitutivist individual are identical.

The solution to the apparent problem of moral freedom thus needs to reflect that moral freedom consists, in part, of knowledge. I have suggested that moral freedom, on her account, consists of (1) freedom from Freudian forces and (2) knowledge of the reality of other individuals. Freedom of the first kind enables but does *not* guarantee freedom of the second kind. Murdoch, thus presents us a picture of human freedom that is pessimistic. We only have freedom of a limited kind. The freedom to 'look again'. However, by the moral effort of love, we may just yet discover the reality of others.

Bibliography

Antonaccio, M., 2012. A Philosophy to Live By: Engaging Iris Murdoch. Oxford University Press-USA.

Bagnoli, C., 2012. The Exploration of Moral Life, in: Broackes, J. (Ed.), Iris Murdoch, Philosopher. Oxford University Press, pp. 197–225.

Barandalla, A., 2023. Murdoch and freedom, in: Saunders, J. (Ed.), Freedom after Kant: From German Idealism to Ethics and the Self. Bloomsbury Publishing, London, pp. 205–219.

Broackes, J., 2012. Introduction, in: Broackes, J. (Ed.), Iris Murdoch, Philosopher. Oxford University Press, pp. 1–93.

Enoch, D., 2021. Constitutivism: on rabbits, hats, and holy grails, in: Chang, R., Sylvan, K. (Eds.), The Routledge Handbook of Practical Reason. Routledge, pp. 336–348.

Gomes, A., 2022. Iris Murdoch: Moral Vision, in: Hopwood, M., Panizza, S. (Eds.), The Murdochian Mind. Routledge, pp. 142–155.

Guyer, P., 2017. Hume, in: The Cambridge History of Moral Philosophy. Cambridge University Press, pp. 338–351.

Hepburn, R.W., Murdoch, I., 1956. Symposium: Vision and Choice in Morality. Aristotelian Society Supplementary Volume 30, 14–58.

Hopwood, M., 2018. 'The Extremely Difficult Realization That Something Other Than Oneself Is Real': Iris Murdoch on Love and Moral Agency. European Journal of Philosophy 26, 477–501.

Hopwood, M., 2017. Murdoch, Moral Concepts, and the Universalizability of Moral Reasons. Philosophical Papers 46, 245–271.

Hume, D., 1978. A Treatise of Human Nature, Second. ed. Oxford University Press.

Jordan, J.E.G., 2014. Reconsidering Iris Murdoch's Moral Realism. The Journal of Value Inquiry 48, 371–385. https://doi.org/10.1007/s10790-014-9416-2

Kant, I., 1996. Groundwork for the Metaphysics of Morals, in: Practical Philosophy, The Cambridge Edition Of The Works Of Immanuel Kant. Cambridge University Press, pp. 37–108.

Katsafanas, P., 2018. Constitutivism about Practical Reasons, in: Star, D. (Ed.), The Oxford Handbook of Reasons and Normativity. Oxford University Press, p. 0. https://doi.org/10.1093/oxfordhb/9780199657889.013.0017

Korsgaard, C.M., 2009. Self-constitution: agency, identity, and integrity. Oxford University Press.

Korsgaard, C.M., 1999. Self-constitution in the ethics of Plato and Kant. The Journal of Ethics 3, 1–29.

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

MacIntyre, A., 2007. After Virtue: A Study in Moral Theory, Third Edition. University of Notre Dame Press.

Mackie, J.L., 1980. Hume's Moral Theory. Routledge & Kegan Paul Ltd.

Mason, C., 2023. Reconceiving Murdochian Realism. Ergo: An Open Access Journal of Philosophy 10, 649–672.

Mason, C., 2021. Iris Murdoch and the Epistemic Significance of Love, in: Cushing, S. (Ed.), New Philosophical Essays on Love and Loving. Palgrave-Macmillan, pp. 39–62.

Mason, C., forthcoming. Murdochian Realism.

McDowell, J.H., 1998. Values and Secondary Qualities, in: Mind, Value, and Reality. Harvard University Press, pp. 131–150.

Meinwald, C., 2016. Why sensibles are not fully real, in: Plato, Routledge Philosophers. Routledge, pp. 211–244.

Murdoch, I., 1999a. Metaphysics and Ethics, in: Conradi, P. (Ed.), Existentialists and Mystics: Writings on Philosophy and Literature. Penguin Books, pp. 59–75.

Murdoch, I., 1999b. Nostalgia for the Particular, in: Conradi, P. (Ed.), Existentialists and Mystics: Writings on Philosophy and Literature. Penguin Books, pp. 43–58.

Murdoch, I., 1999c. On "God" and "Good," in: Conradi, P. (Ed.), Existentialists and Mystics: Writings on Philosophy and Literature. Allen Lane/the Penguin Press, pp. 337–362.

Murdoch, I., 1999d. The Darkness of Practical Reason, in: Conradi, P. (Ed.), Existentialists and Mystics: Writings on Philosophy and Literature. Penguin Books, pp. 193–202.

Murdoch, I., 1999e. The Idea of Perfection, in: Existentialists and Mystics: Writings on Philosophy and Literature. Allen Lane/the Penguin Press, pp. 299–336.

Murdoch, I., 1999f. The Sovereignity of Good Over Other Concepts, in: Conradi, P. (Ed.), Existentialists and Mystics: Writings on Philosophy and Literature. Allen Lane/the Penguin Press, pp. 363–385.

Murdoch, I., 1999g. Vision and Choice in Morality, in: Conradi, P. (Ed.), Existentialists and Mystics: Writings on Philosophy and Literature. Allen Lane/the Penguin Press, pp. 76–98.

Murdoch, I., 1998. Existentialists and Mystics: Writings on Philosophy and Literature. Allen Lane/the Penguin Press.

Murdoch, I., 1993. Metaphysics as a guide to morals. Allen Lane, Penguin Press.

Murdoch, I., 1970. The Sovereignty of Good. Routledge.

Panizza, S., 2022. Attention, in: Hopwood, M., Panizza, S. (Eds.), The Murdochian Mind. Routledge, pp. 156–168.

Panizza, S.C., 2022a. Attention without self, in: Ethics of Attention: Engaging the Real with Iris Murdoch and Simone Weil. Routledge, pp. 87–109.

Panizza, S.C., 2022b. Attention without self-concern, in: Ethics of Attention: Engaging the Real with Iris Murdoch and Simone Weil. Routledge, pp. 63–87.

Plato, 1997. Republic, in: Cooper, J.M., Hutchinson, D.S. (Eds.), Grube, G.M.A., Reeve, rev. C.D.C. (Trans.), Plato: Complete Works. Hackett Publishing Company, Indianapolis.

Samuel, J., 2021. Thin as a Needle, Quick as a Flash: Murdoch on Agency and Moral Progress. Review of Metaphysics 75, 345–373.

Sayre-McCord, G., 2023. Moral Realism. Stanford Encyclopedia of Philosophy.

Schapiro, T., 2020. Kant's Approach to the Theory of Human Agency, in: The Routledge Handbook of Practical Reason. pp. 160–171.

Setiya, K., 2013. Murdoch on the Sovereignty of Good. Philosophers' Imprint 13.

Stern, R., 2022. I—The Presidential Address: 'How Is Human Freedom Compatible with the Authority of the Good?' Murdoch on Moral Agency, Freedom, and Imagination. Proceedings of the Aristotelian Society 122, 1–26.

Weil, S., 1973. Waiting for God. Harper & Row, Publishers, Incorporated.

Wittgenstein, L., 1961. Notebooks 1914-1916. Basil Blackwell Publisher, Oxford.